Part IV

ign and lysis of gle-Factor dies

Chapter

Introduction to the Design of Experimental and Observational Studies

In Parts I–III, we focused on the use of linear and nonlinear statistical models for the analysis of experimental and observational data. There, an observed response vector Y and associated design matrix X were used to model the relationship between response and the predictors and to develop appropriate statistical inferences. We will now emphasize the *statistical design* of scientific studies.

Our basic goal will be to design studies in such a way that they lead to a simple, effective statistical analysis. Since nearly all scientific studies are analyzed using linear statistical models, the ability to design studies properly depends critically on an understanding of the materials covered in Parts I–III. For example, in Section 4.7, we discussed the range, spacing, and number of X levels when the objective of the study was to estimate a simple linear relation between a response Y and a single predictor X. We observed there that the range and spacing of the Xs have a direct effect on the precision with which we estimate key parameters, such as the slope. We showed that the variance of the estimated slope is minimized when the Xs are split evenly at minimum and maximum levels for the scope of the experiment. Minimization of this variance leads to a more precise parameter estimate and improved statistical power.

In this chapter and those that follow, we consider the *design* of scientific studies and the specialized linear models—called *analysis of variance (ANOVA) models*—employed in their analysis. We emphasize that the proper design of a scientific study is far more important than the specific techniques used in the analysis. As we shall see, a well-designed study is usually simple to analyze. On the other hand, a poorly designed study or a botched experiment often cannot be salvaged, even with the most sophisticated analysis.

We begin in the current chapter with an overview of the design of scientific studies. Generally, a scientific study can be categorized as either an experimental study or an observational study. The distinction is important because experimental studies provide a much firmer basis for the establishment of cause-and-effect relationships between one or more explanatory factors and a response variable than do observational studies. With the latter, one can establish association between the explanatory factors and the response variable,

but not causation. We continue with an overview of the basic concepts and planning approaches used in the design of experimental and observational studies. Finally, we present a case study to illustrate both the design and analysis of an experimental study based on a matched pairs design.

15.1 Experimental Studies, Observational Studies, and Causation

Experimental Studies

For many persons, the first exposure to the concept of an *experiment* was in a high school or elementary school science class. For example, a high school science teacher might demonstrate the influence of atmospheric pressure on boiling temperature by showing that water will boil at room temperature in a near vacuum. We note that this example was not an experiment, but was simply a demonstration. Designed experiments are conducted to demonstrate a cause-and-effect relation between one or more explanatory factors (or predictors) and a response variable. The demonstration of a cause-and-effect relationship is accomplished, in simple terms, by altering the levels of the explanatory factors (i.e., the Xs) and observing the effect of the changes on the response variable Y. Furthermore, designed experiments are frequently *comparative* in nature.

For example, a famous experiment on the effects of vitamin C on the prevention of colds in 868 children was conducted in 1976. Of the 868 children studied, half were randomly selected for the *experimental group*. Children in this group received a 1,000-mg tablet of vitamin C daily for the test period. The remaining children, who made up the *control group*, received a placebo—an identical tablet containing no vitamin C—also on a daily basis. The results showed that the average number of colds per child was .38 for children receiving vitamin C, while the average for children receiving the placebo was .37. The difference between the two groups (.01 colds per child) was not statistically significant.

The explanatory factor in the vitamin C example is a qualitative predictor X having two levels: $X = 1$ if child received vitamin C; $X = 0$ if child did not receive vitamin C. The different levels of the explanatory factors in an experimental study are frequently referred to as *treatments*. Just as there are two levels of the explanatory factor in the vitamin C experiment, there are two treatments: vitamin C and placebo. The objects or entities to which treatments are applied are generally referred to as *experimental units*. Here the experimental units are the children who received either of the two treatments.

Assignment of the treatments (factor levels) to the experimental units was performed using a process called *randomization*. We shall discuss randomization in detail in Section 15.2, but we note for now that the purpose of randomization here was to balance the characteristics of the children in each of the treatment groups, so that differences in the response variable can be attributed to treatment differences, and not to differences between the two groups of children. For example, one could imagine a poorly designed version of this study, in which the 868 children attended two elementary schools. For convenience, the investigator might use children from one school as the experimental group, and children from the second school as the control group. In such a plan, it would be impossible to distinguish the effects of being in a particular school—which could be severe if a particularly contagious cold virus broke out in one of the schools—from the presence or absence of vitamin C. In contrast, with randomization, we are guaranteed that about half of the children from each

school would receive the vitamin C regimen. Therefore any differences in the incidence of colds in the two groups will likely not be attributable to or confounded with the schools.

Thus a characteristic feature of an experimental study is that the investigator exercises control over the assignment of treatments to the experimental units through the process of randomization. If important differences in the responses result between the treatment groups, we can attribute them to the treatments. We give the following definition of a comparative experimental study.

> In a *comparative experimental study*, randomization is employed to assign a set of treatments to the experimental units, and the observed outcomes among the treatment groups are compared to assess treatment effects. The treatments are defined by the levels of one or more explanatory factors, referred to as *experimental factors*. Cause-and-effect relationships between the experimental factors and the outcome or response variable can be established in an experimental study. **(15.1)**

We now present another example of an experimental study.

Example 1

Experimental Study of Quick Bread Volume. A simple comparative experiment was conducted to study the effect of baking temperature on the volume of a quick bread prepared from a package mix. Four oven temperatures—low, medium, high, and very high—were tested by randomly assigning each of the four levels of temperature to five package mixes. This is an experimental study because the levels of the explanatory factor (baking temperature) are randomly assigned to the experimental units. The experimental units here are the 20 packages of mix. The experimental design used is called a *completely randomized design*, with each of the 20 packages of mix having an equal chance to be assigned to each of the four cooking temperatures. Note that the design used in the vitamin C example was also a completely randomized design.

Comment

The vitamin C experimental study is an example of a clinical trial. A *clinical trial* is defined as a prospective intervention study, where one is interested in comparing the effects of different treatment interventions starting at one point in time on the outcome at a later point in time. ■

Observational Studies

An observational study differs from an experimental study in that randomization of the treatments to experimental units does not occur. For example, a study of the effects of education and type of work experience of sales people on their sales volumes was made by selecting a random sample of sales people currently employed by a company and obtaining information on highest degree obtained, type of experience, and sales volume for each of the selected employees. This is an observational study because it is not possible to randomly assign the levels of the predictor variables of interest (education and type of experience) to the employees.

We focus here on "comparative" observational studies, where two or more groups (populations, subpopulations, processes, etc.) are compared. The sales example just mentioned is a comparative observational study, because sales volumes for different groups of sales people were to be compared for different levels of education, experience, and so forth. This

is in contrast to simple descriptive studies that do not involve statistical comparisons of groups. We give the following definition of a comparative observational study:

> In a *comparative observational study,* random samples are obtained from two or more populations (or subpopulations) and the observed outcomes are compared across populations (or subpopulations). The populations or subpopulations are defined by the levels of one or more explanatory factors, referred to as *observational factors*. A cause-and-effect relationship between the explanatory factors and the outcome or response variable is difficult to establish in an observational study. Usually, evidence external to the observational study would be required to rule out possible alternative explanations for cause and effect. **(15.2)**

At times, investigators use nonrandom convenience or quota samples. These samples are sometimes referred to as pseudo-random samples or representative samples and treated as if they were truly random. It must be cautioned here that random selection or assignment greatly enhances the generalizability of the study results and avoids potential biases that otherwise may occur when nonrandom selection is used.

The following is an example of an observational study.

Example 2

Observational Study of Teaching Effectiveness. Recently, the administration of a college of business offered its faculty the opportunity to participate in a summer workshop on case teaching methods. Faculty were not required to attend the workshop, but were asked to sign up on a first-come, first-served basis. Of the 110 faculty in the business school, 63 faculty elected to attend the seminar.

At the end of the following academic year, the administration compared the recent teaching performances of faculty who attended the seminar to those who did not attend. Students evaluated faculty on a 7-point scale, where 1 indicates poor performance and 7 is outstanding. Average teaching ratings for all faculty members during the year following the seminar were obtained. The aligned dot plots in Figure 15.1 compare the performances of faculty who attended the seminar with faculty who chose not to attend. These plots suggest that faculty who attended the seminar were generally rated more highly by students than faculty who did not attend, and this is confirmed by the sample averages. The average rating for faculty who attended was 5.76; the average for those who did not attend was 5.26. On the basis of two-sample t-test (A.67), administrators concluded that the observed difference (.50) was statistically significant. The P-value of the test was 0+.

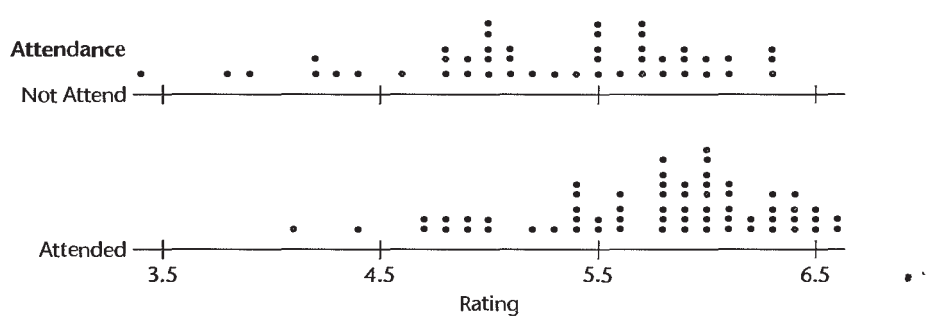


FIGURE 15.1 Teaching Performance Comparison—Teaching Effectiveness Example.

It is tempting to conclude, on the basis of this analysis, that the seminar was effective in improving the quality of teaching. However, this is clearly an observational study, because a random assignment of the treatments (attend workshop, do not attend workshop) to experimental units (instructors) did not occur. Thus cause-and-effect between the explanatory factor (workshop attendance) and the response (teaching effectiveness) cannot be directly inferred. It is possible that the workshop improved teaching quality, but a number of alternative explanations for the observed difference are also plausible. For example, it may have been that better, or more highly motivated, teachers volunteered for the workshop. In this case, the workshop attendees would be rated more highly on average even if the workshop had no beneficial effect.

This investigation would have been an experimental study if the administration had chosen a subset of the faculty at random for participation in the workshop. If the results led to a difference in teaching quality, such as that shown in Figure 15.1, the administration would be justified in concluding that the seminar had a beneficial effect on teaching effectiveness. The reason that a cause-and-effect conclusion would be justified here is that the randomization would tend to balance out the differences in other factors, such as pre-workshop teaching ability or motivation, leaving the observed differences attributable to the experimental treatment.

Comment

Ordinal level data are frequently assumed to approximate equally spaced interval data and as such are appropriately analyzed using statistical techniques designed for continuous, equal interval level measurements. We have done so with the teaching effectiveness scores but caution the reader that at times this assumption may not be supported, in which case specialized techniques for the analysis of ordinal level data, such as those discussed in Chapter 14, should be employed. ■

Mixed Experimental and Observational Studies

A third type of study, which involves aspects of both experimental and observational studies is also possible. We illustrate this third case with an example.

Example 3

Mixed Experimental and Observational Study of Mechanics' Training. An appliance manufacturer operates three regional training centers in the United States for training mechanics to service the company's products. At each regional center, two different training programs were studied, with the trainees from the region assigned at random to one of the two training programs. One may view this as a two-factor study, the factors being training program (experimental factor) and training center (observational factor). If the same training program is superior to the other in all three centers, the evidence is quite clear as to the comparative effects of the training programs since at each center the trainees from the region were assigned at random to the two programs.

Note that the training center was not randomly assigned to subjects; each trainee was assigned to the center for the region in which the trainee is located. Therefore a cause-and-effect relationship between training centers and quality of training cannot be demonstrated rigorously. One center may excel for any number of reasons, such as because its staff is doing a better training job, because it has better facilities, or because trainees assigned to it come from a geographic region in which better education is provided. Evidence external to this study would be required as to whether or not the education of trainees at the three

centers is the same, whether or not the facilities are equal, and the like, before a clear understanding of the reasons for differences between training centers could be obtained.

As we will see, this is an example of a blocked experimental study, where the blocks refer to the training centers (observational factor) and the training program is the treatment (experimental factor).

Experimental Studies: Basic Concepts

The *design of an experiment* refers to the structure of the experiment, with particular reference to:

- The set of explanatory factors included in the study.
- The set of treatments included in the study.
- The set of experimental units included in the study.
- The rules and procedures by which the treatments are randomly assigned to the experimental units (or vice versa).
- The outcome measurements that are made on the experimental units.

In this section we discuss each of these topics in turn.

actors

A *factor* is an explanatory variable to be studied in an investigation. For instance, in an investigation of the effect of price on sales of a luxury item, the factor being studied is price. Similarly, in a study comparing the appeal of four different television programs, the factor under investigation is television program. In the quick bread volume example, the factor under investigation is baking temperature. In a regression context, factors are typically referred to as predictors or independent variables.

A factor may be categorized as to whether it is an experimental factor or an observational factor. An *experimental factor* is one where the level of the factor is assigned at random to the experimental unit. An illustration is the factor baking temperature in the bread volume example. In any investigation based on observational data, the factors under study are observational factors. An *observational factor* pertains to the characteristic of the units under study and is not under the control of the investigator. Observational factors can be found in experimental studies, and therefore it is important to recognize them as such, since cause-and-effect inferences cannot be made for these factors. As we noted earlier in the mechanics' training example, the training program was an experimental factor, while the training center was an observational factor.

Just as in regression, where both qualitative and quantitative predictors can be employed, experimental factors can be either quantitative or qualitative. A *qualitative factor* is one where the levels differ by some qualitative attribute. Examples are type of advertisement, brand of rust inhibitor, or television program. In Chapter 8 we described the use of $r - 1$ indicator variables to model a qualitative predictor having r levels. A *quantitative factor* is one where each level is described by a numerical quantity on an equal-interval scale. Examples are temperature in degrees Celsius, age in years, or price in dollars.

A *factor level* is a particular form of that factor. In the bread volume example, four baking temperatures were used, namely, 320°F (low), 340°F (medium), 360°F (high), and 380°F

(very high). Each of these temperatures is a level of the factor under study, and we say that the temperature factor has four levels in this study. As another example, in a study of the effect of color of the paper used in a mail questionnaire on response rate, color of paper is the factor under study, and each different color used is a level of that factor.

Crossed and Nested Factors

Investigations differ as to the number of factors studied. Some are *single-factor studies*, where only one factor is of concern. For instance, the study of the effect of four different baking temperatures on quick bread volume mentioned earlier is an example of a single-factor study. In *multifactor studies*, two or more factors are investigated simultaneously. An example of a multifactor investigation is a study of the effects of three levels of temperature and two levels of concentration of solvent on the yield of a chemical process. Here, two factors—temperature and concentration—are studied simultaneously to obtain information about their effects on the yield. The three levels of temperature and two levels of solvent concentration lead to $3 \times 2 = 6$ factor-level combinations:

Factor Combination	Temperature	Solvent Concentration
1	Low	Low
2	Low	High
3	Medium	Low
4	Medium	High
5	High	Low
6	High	High

These factor combinations can be represented by the two-way table in Figure 15.2a. We say that the two factors are *crossed* when all combinations of the levels of the two factors are included in the study. The sales volume study is another example of a study in which the factors, education and type of experience, are crossed.

FIGURE 15.2 Crossed Factors and Nested Factors—Chemical Yield and Production Yield Experiments.

(a) Crossed Factors—Chemical Yield Experiment

Solvent Conc.	Temperature		
	Low	Medium	High
Low	X	X	X
High	X	X	X

(b) Nested Factors—Production Yield Experiment

Plant	Operator								
	1	2	3	4	5	6	7	8	9
1	X	X	X						
2				X	X	X			
3							X	X	X

In some studies the levels of one or more of the factors are unique to a particular level of another factor. For instance, in a study of the effects of operators on production yield in three manufacturing plants, three operators were selected in each of the three plants, and their production yields were recorded for five batches of product. A diagram of this experiment is given in Figure 15.2b. Note that the first three operators are employed only in plant 1, the next three are employed only in plant 2, and the last three are employed uniquely in plant 3. Here, operators are said to be *nested* within manufacturing plants.

Treatments

The set of treatments to be included is determined by the set of factors and the levels of each factor. In single-factor studies, a treatment corresponds to a factor level. Thus, in a study of five advertisements, each advertisement is a treatment. In multifactor studies, a treatment corresponds to a combination of factor levels. For instance, in a study of the effects on sales volume of price ($.25, $.29) and package color (red, blue), each price-color combination, such as $.25 price—red package color, is a treatment. When a treatment is indicated by a combination of two or more factor levels, the combination of levels is sometimes referred to as a *treatment combination*. This particular study contains four treatments or treatment combinations since there are four price-color combinations.

The definition of a treatment can at times be a difficult problem. Consider an experiment to study whether C or JAVA is a better programming language to teach in an introductory computing course. Some teachers will prefer C, others JAVA. Should the treatments then be defined as the programming language taught by instructors who prefer that language? If so, differences in findings may be due to differences between the two groups of instructors. Should the definition of a treatment not include the instructor, and instructors be randomized, with some being forced to teach a language they do not prefer? Or should instructor preference be a second factor, with each instructor teaching both languages? Problems of this kind need careful resolution so that the results of the study will be useful.

Choice of Treatments

Generally, the investigator must decide upon the number of factors to be included, the number of levels of each factor, the range of levels within each factor (for quantitative factors), and the need for a control treatment. We shall discuss each of these aspects in turn.

Number of Factors. In the initial stages of an investigation or when little theory is available, there is frequently a desire to include many more factors than can possibly be studied in a single experiment. For example, the quick bread volume experiment discussed above was adapted from a much larger optimization study of quick bread production. When the study was initiated, process engineers and food scientists conducted a brainstorming session to identify factors that could potentially affect quick bread volume. Cause-and-effect diagrams (also known as Ishakawa or fish-bone diagrams), such as that shown in Figure 15.3, are often used to guide such sessions and to summarize results. This particular session identified over 15 potential causal factors—far too many to include in the experiment. From this number, four factors—oven temperature, proof time, yeast type, and flour protein level—were included, each at two levels. This led to $2^4 = 16$ treatment combinations.

Number of Levels of Each Factor. For qualitative factors, the number of levels may be dictated by the nature of the factor. For example, in the incentive system example discussed

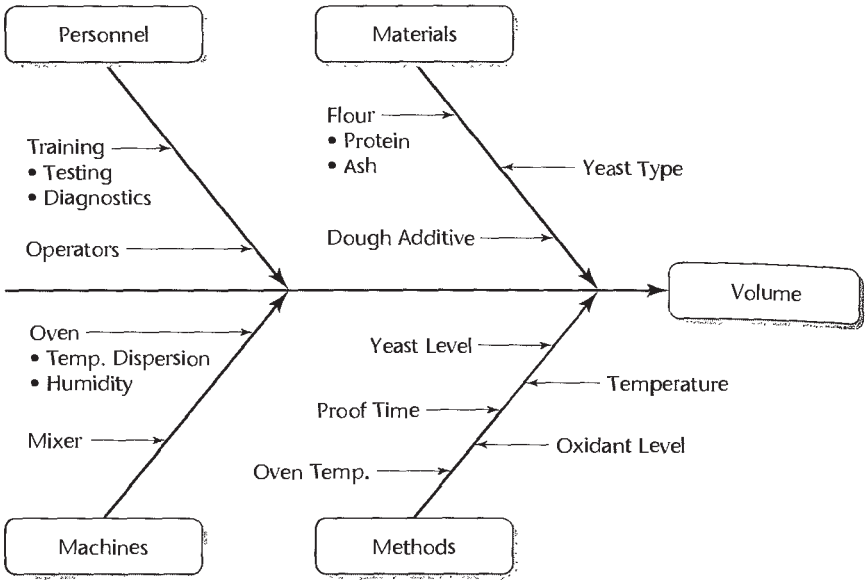


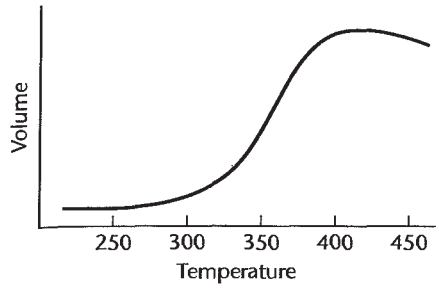
FIGURE 15.3 Cause-and-Effect Diagram—Quick Bread Optimization Example.

earlier, three alternative incentive systems were under consideration. One involved increases to hourly wages, another involved the use of bonuses and financial awards, and another involved recognition and the awarding of additional vacation time. Thus the company felt that all three levels of the incentive system factor should be included in the experiment. In other instances, it might be necessary to drop one or more of the levels of a qualitative factor in order to reduce the cost of the experiment. For example, in an experiment to investigate the effect of color of paper (blue, green, orange, and yellow) on the response rates for questionnaires, it might be concluded that a least-promising color should simply be eliminated in order to reduce the cost or complexity of the experiment.

For quantitative factors, the number of levels chosen should reflect the type of trend expected by the experimenter. If the experimenter believes that the change in the response will be roughly linear in the range chosen for the factor, two levels—the minimum and the maximum of the specified range—may be sufficient. Three levels are useful if the experimenter believes that the response will follow a quadratic trend in the chosen range, or if a linear trend is expected, but a test for lack of fit is desired. Use of four or more levels is justified if a highly detailed examination of the shape of the response curve is desired, or if the response curve is increasing or decreasing to an asymptotic value. Often, three equally spaced levels are sufficient.

Range of Levels for Quantitative Factors. Choosing the range of a quantitative factor to be explored is one of the most important design decisions. If the range is too small, the effect of a change from the smallest level to the largest level of the factor may be too small to detect. If the range is too large, important changes in the mean response may be missed. For example, suppose that the true regression function in the quick bread volume example is given by the curve in Figure 15.4. The response increases in roughly linear fashion for baking temperatures between 300 F and 400°F, and levels off for baking temperatures outside this range. If the range is too small and we are in an area where the change in the mean response is small or moderate, for example 250°F–300°F, we will conclude that temperature has

15.4



little effect on volume. If the range is too large, for example 250°F–450°F, and only these two levels are used as treatments, important features of the curve—such as the maximum (near 400°F) may be missed. We see that an effective choice of range for a quantitative factor frequently requires a good prior knowledge of the nature of the relationship between the mean response and the factor(s) under study.

Control Treatment. A control treatment is needed in some experiments, but not in all. A control treatment consists of applying the identical procedures to experimental units that are used with the other treatments, except that none of the treatments are applied. In a study of food additives, for instance, a treatment may consist of a portion of a vegetable containing a particular additive that is served to a consumer in a particular experimental setting in the laboratory. A control treatment here would consist of a portion of the same vegetable served to a consumer in the identical experimental setting except that no food additive has been used.

A control treatment is required when the general effectiveness of the treatments under study is not known, or when the general effectiveness of the treatments is known but is not consistent under all conditions. In the food additives example, suppose it is known that food additive A is highly effective in enhancing the tastiness of vegetables and it is desired to see if additives B and C are equally effective or possibly even more effective. In that case, a standard of comparison is available and no control treatment is required. On the other hand, suppose there is no knowledge about the general effectiveness of the three additives, and the following results are obtained (ratings can range between 0 and 60):

Additive	Mean Rating
A	39
B	37
C	41

Assume that the sample sizes are large so that the mean ratings are very precise. In the absence of a standard of comparison, one would not know here whether each of the three additives is effective or whether none of the additives is effective.

It is crucial that the control treatment be conducted in the identical experimental setting as the other treatments. In the food additives example, for instance, a survey of consumers at home, in which persons are asked to rate the general tastiness of the vegetable (without any additive) on the same scale as in the experiment, would not qualify as a control treatment. Such a survey might yield a mean rating of 22, suggesting that the three additives substantially increase the tastiness of the vegetable. This conclusion, however, could be grossly misleading. If the control treatment actually were incorporated into the experiment

so that consumers are given portions of the vegetable with no additive in the laboratory setting, the mean rating for the control treatment might be 40. This result would imply that none of the three additives is effective in enhancing the tastiness of the vegetable. The reason for the higher mean rating in the laboratory setting could be a "halo" effect connected with the experimental procedures. Possibly, foods served in the experimental setting taste better than at home, or perhaps consumers try to oblige by giving higher ratings when they participate in an experimental study. Thus, only a control treatment incorporated into the experiment can serve as the proper standard of comparison.

Experimental Units

As we noted earlier, the experimental units are the objects or entities to which the treatments are applied in an experimental study. There are times when confusion may arise as to the precise nature of the experimental unit. The following definition makes clear that experimental units are determined by the method of randomization employed.

> An experimental unit is the smallest unit of experimental material to which a treatment can be assigned; the experimental unit is thus *determined by the method of randomization.* **(15.3)**

For example, consider again the experimental study of two incentive pay systems. We asked above if the basic study unit should be an individual employee, a shift, or a plant. As noted, it may be impossible to assign different incentive pay systems to individual employees or to individual shifts, but a random assignment of different incentive systems to different plants would be feasible. Here, the smallest unit of experimental material to which a treatment (incentive system) can be assigned is the plant, and so it follows that the plant is the experimental unit.

Representativeness of the experimental units is another important consideration in the design of experimental studies. Consider a study of management behavior with different communications networks. A university investigator may be tempted to use students as subjects because of their ready availability. If, however, information is desired about the behavior of business people, the students may not be representative experimental units. It hardly needs to be stated that an investigator should make every effort to obtain representative experimental units. Conversely, one should be cautious in extending results of an investigation to groups for which the study units are not representative. Thus, if the communications network study cited above *did* use students, one should not automatically assume that the findings are relevant to business people.

A different aspect of defining the basic unit of study occurs in investigations of sales and similar phenomena. Suppose that we wish to measure the effectiveness of five different television commercials in terms of sales during a period of time subsequent to their showing. Should the length of time be one week, two weeks, one month, or some other time period? Clearly, the purposes of the study will need to govern the length of time that makes up the basic study unit here.

Sample Size and Replication

Sample size is usually determined by statistical considerations, by resource or budget considerations, or both. Generally, the larger the sample size, the greater will be our ability to detect any differences in responses due to the treatments. Thus a key step in any experimental

design is to assess the *power* of the statistical tests to be used in the analysis, or the precision of the estimates to be produced by the analysis, as a function of sample size. Ultimately, a trade-off must be made between the increase in power and precision resulting from higher sample sizes, and the added cost or time required to field the experiment. Statistical procedures used for determining power and precision depend on the particular experimental design used. We shall discuss these methods throughout the remainder of the text as new experimental designs are introduced.

We note that in many designed experiments, the sample size is an integer multiple of the number of treatments. For example, in the bread volume experiment, there were eight experimental units (packages of bread mix) and four treatments. Thus each treatment was repeated twice. We say that there were two *complete replicates* of the experiment. Frequently, the total sample size is simply determined by the number of complete replicates chosen in the experimental design. Replication makes it possible to estimate the experimental error variance, which is required for testing the presence of treatment effects or for establishing confidence interval estimates of these effects. When a treatment is repeated, any difference in the response from prior responses for the same treatment (under similar experimental conditions) is due to experimental error, and it therefore provides one additional piece of information (i.e., one degree of freedom) about the pure error variance. If this experimental error variance is small, the response is sometimes said to be highly *reproducible*. If the error variance is high, the response has low reproducibility.

Randomization

Randomization in experiments is a relatively recent idea, first introduced by the famous British statistician Sir R. A. Fisher during the early part of the twentieth century. In the past, treatments had been assigned to experimental units either on a systematic or on a subjective basis. We noted in the teaching effectiveness example how biases can arise when self-selection is employed to assign experimental units to the treatments. The same dangers exist with systematic and subjective selection. For instance, consider an experiment using 10 employees and two treatments, where the first five employees on the payroll listing are assigned treatment 1 and the next five treatment 2. Suppose that the payroll listing is by seniority, and that experience is related to productivity, the phenomenon under study. A comparison of treatments 1 and 2 then will reflect not only differences between the two treatments but also differences in the amount of experience between the two groups of employees. This potential bias may be so transparent that no good experimenter would use the type of systematic assignment just described. Nevertheless, there may be many other sources of bias in systematic selection that are not so apparent.

Subjective assignments of treatments to experimental units can also lead to selection bias, as when an experimenter subconsciously tends to assign one treatment to highly extrovert subjects and the other treatment to less extrovert subjects.

With randomization, the treatments are assigned to experimental units at random. Randomization tends to average out between the treatments whatever systematic effects may be present, apparent or hidden, so that comparisons between treatments measure only the pure treatment effects. Thus, randomization tends to eliminate the influence of extraneous factors not under the direct control of the experimenter and thereby precludes the presence of selection bias. Cochran and Cox (Ref. 15.1, p. 8) have likened randomization to an insurance policy in that it is a precaution against biases that may or may not occur.

Randomization is appropriate not only for the assignment of treatments to experimental units but also for any other phases of the experiment where systematic effects not under the control of the experimenter may be present. For instance, consider an experiment in which five treatments (alternative methods of measuring subjective probability) are studied and 20 subjects are used. Only one subject can be run per day; thus, four weeks are required to complete the experiment. In this type of situation, it usually is highly desirable to determine the order of the treatments randomly since a variety of systematic time effects could be present. The experimenter may with time improve the explanation of the methods of measuring subjective probability, there may be a streak of extremely hot weather during a week, and the like. With these possible time effects, a systematic assignment of one treatment per week could lead to seriously biased results. Randomization, on the other hand, will tend to average out whatever systematic effects are present, whether anticipated or not.

How to Randomize. Randomization requires that a series of experimental units (or treatments) be placed in a random order. To illustrate this in simple fashion, we consider again the quick bread volume example with two replicates. Here four treatments (T_1—low, T_2—medium, T_3—high, T_4—very high) are considered and $2 \times 4 = 8$ package mixes, labeled 1 through 8, are to be used as experimental units. The situation is:

Treatments	T_1	T_2	T_3	T_4
Sample Sizes	2	2	2	2

and the eight treatments to be assigned to package mixes are listed as (the order is arbitrary):

T_1	T_1	T_2	T_2	T_3	T_3	T_4	T_4

To randomly assign the treatments to the experimental units, we obtain a random ordering of these treatments. To do so we generate eight random numbers from any continuous probability distribution (or obtain eight random numbers from a table of random digits) and associate each number obtained in sequence with the above list of treatments. The eight random numbers below were obtained from a standard normal random number generator:

T_1	T_1	T_2	T_2	T_3	T_3	T_4	T_4
−0.37	0.01	1.40	−1.65	0.16	−0.25	−1.10	0.77

We now rearrange the pairs above in ascending sequence for the random numbers and associate them with the package mixes, which we have arbitrarily labeled "1" through "8." Thus we obtain the following randomized assignment of treatments to experimental units:

Treatment:	T_2	T_4	T_1	T_3	T_1	T_3	T_4	T_2
Random number:	−1.65	−1.10	−0.37	−0.25	0.01	0.16	0.77	1.40
Package mix:	1	2	3	4	5	6	7	8

As a result of the randomization, treatment T_1 (low temperature) is to be assigned to package mixes 3 and 5; treatment T_2 (medium temperature) is to be assigned to package mixes 1 and 8; and so on. The experimental trials should be conducted in a random order.

Some statistical packages provide facilities for randomly permuting the treatments (or experimental units) directly, which can simplify the process considerably.

Comments

1. Randomization also can provide the basis for making inferences without requiring assumptions about the distribution of the error terms. We shall discuss this use of randomization in Section 16.9.

2. The implications of randomization may be viewed in a somewhat different fashion than that presented so far. The random errors of experimental units that are adjacent in time or space are often correlated, not independent, as a result of various systematic effects over time or space. Randomization does not eliminate this correlation pattern but, by making it equally likely that any two treatments are adjacent, tends to eliminate the correlations between treatments with increasing replications. Thus, randomization makes it reasonable to analyze the data as though the model random error terms are independent, an assumption that has been made in almost all models discussed so far.

3. Occasionally, randomization may provide a pattern that makes the experimenter uneasy. For instance, randomization of the time sequence in which four experimental units were assigned to treatment 1 and four assigned to treatment 2 may result in a randomized sequence where the four experimental units for treatment 1 are exposed first and then the four experimental units for treatment 2 are exposed. This is not a likely occurrence, but one that can take place. Some solutions have been suggested for this problem, but none provides a final answer. In practice, the experimenter typically will discard a randomization sequence that has apparent dangers of systematic effects for the particular experiment and select another randomization. ■

Constrained Randomization: Blocking

Blocking is a technique that can be used to increase precision in any experiment. To provide some context and to motivate the concept, we shall again consider the vitamin C experiment discussed earlier.

Recall that half of the children in the vitamin C example were randomly assigned to the control group, and half were assigned to the experimental group. At the end of the test period, the number of colds Y contracted by each child was recorded. A linear statistical model for the ith child's response is:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i \tag{15.4}$$

where:

$$X_i = \begin{cases} 1 & \text{if } i\text{th child receives vitamin C} \\ 0 & \text{if } i\text{th child receives placebo} \end{cases}$$

With X_i defined in this fashion, β_0 is the population mean response for children in the control group (i.e., those receiving the placebo), and $\beta_0 + \beta_1$ is the population mean response for children in the experimental group (i.e., those receiving vitamin C). The treatment effect parameter, β_1, represents the increase or decrease in the average number of colds per child due to the vitamin C regimen. Finally, the experimental error ε_i is the deviation of the number of colds for the ith child from the true mean of the child's treatment group—sometimes called the specific effect associated with the ith experimental unit. The variance of the experimental error is $\sigma^2 = \sigma^2\{\varepsilon_i\}$.

We shall assume that the goal of the study is precise estimation of (or inference about) the treatment effect, β_1. Then a key quantity of interest is the variance of the least squares estimator, b_1, of this effect. From (2.3b), we have:

$$\sigma^2\{b_1\} = \frac{\sigma^2}{\sum(X_i - \bar{X})^2} \tag{15.5a}$$

It is easy to show, when the number of children in the two treatment groups are the same, that the variance of b_1 is:

$$\sigma^2\{b_1\} = \frac{4\sigma^2}{n} \tag{15.5b}$$

Thus for a given sample size (here $n = 868$), increased precision can only come about through reductions in the experimental error variance, σ^2.

One way to reduce $\sigma^2\{\varepsilon_i\}$ is to identify and control factors that contribute to variation in the ε_i. In the vitamin C example, some factors (other than vitamin C) that might affect the numbers of colds contracted by the ith child might include: the gender of the child, the age of the child, the general health status of the child, the nutritional habits of the child, and so on. These factors, which affect the response but are not of primary interest to the investigator, are referred to as *nuisance* or *confounding* factors. For simplicity, we will assume that there is just one nuisance factor in the experiment other than the treatment effect, namely, gender. This source of variation could be removed from experimental error by using only males or only females.

For example, if only females are used as subjects, the model for our response is now:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \varepsilon_i^F \tag{15.6}$$

where ε_i^F is the experimental error when subjects are exclusively female. If females tend to have fewer (or more) colds than males, then the female experimental units are more homogeneous and the experimental error variance will be reduced.

Of course there are disadvantages to limiting the experiment to one gender. First the sample size n is reduced, which increases the variance of our estimated treatment effect in (15.4b), and second, we would not be able to generalize the results of the experiment to the gender that was omitted. These disadvantages are overcome by a technique known as blocking.

In a *blocked* experiment, the heterogenous experimental units are divided into homogeneous subgroups called *blocks*, and separate experiments are conducted in each block. For example, blocking on gender in the vitamin C example would be accomplished by conducting separate experiments on males and females. Because gender does not vary within blocks, the effect of vitamin C is more efficiently estimated within each block. The overall effect of the experimental factor is obtained by combining the estimated effects from each of the blocks.

Note that because blocking requires that separate experiments be conducted in each block, it follows that separate randomizations of treatments to experimental units (or vice versa) must be carried out within each block. The within-block randomization is sometimes referred to as a *restricted randomization* because assignments of treatments can only be made to experimental units within the given block.

15.5 Randomized Complete Block Design—Vitamin C Example.

	Restricted Randomizations						
Male	1	2	3	4	...	433	434
Treatment:	Vitamin C	Vitamin C	Placebo	Vitamin C	...	Vitamin C	Placebo
Female:	1	2	3	4	...	433	434
Treatment:	Placebo	Vitamin C	Placebo	Vitamin C	...	Placebo	Vitamin C

An example of a blocked layout for the vitamin C example is given in Figure 15.5. Notice that each block consists of 434 subjects (assuming half of the 868 subjects are male and half are female), and that the control and experimental treatments are each assigned to half of the subjects in each block. This is accomplished with two restricted randomizations.

The advantages of a blocked experiment over a completely randomized design should be evident in this example. Randomization alone cannot guarantee that the same number of males and females will receive each treatment. Thus if one gender tends to have fewer colds, differences in the treatment groups may be observed even when the experimental treatment has no effect. Another benefit of blocking is that it can increase the range of validity for the conclusions from the experiment. Blocking of experimental units according to their characteristics (e.g., by age) can be employed to provide sufficient variability between groups of experimental units in different blocks for a wide range of generalizability and yet achieve high precision because of small experimental errors within blocks.

As a general principle, an experimenter should always try to remove any known or potential sources of variability, either by holding the nuisance factors constant throughout the experiment or by blocking. Randomization within blocks provides additional protection against any unknown sources of variability that may be present.

Comments

1. The amount of variance reduction achieved by blocking can be seen from a regression context. Suppose in the vitamin C example that the model for the response of the jth subject having gender i ($i = 1$ if female; $i = 0$ if male) is:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \varepsilon_{ij} \tag{15.7}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if } ij\text{th child receives vitamin C} \\ 0 & \text{if } ij\text{th child receives placebo} \end{cases}$$

$$X_{ij2} = \begin{cases} 1 & \text{if } i\text{th child is female} \\ 0 & \text{if } i\text{th child is male} \end{cases}$$

Here β_1 can again be interpreted as the change in mean response due to receiving vitamin C (relative to receiving the placebo) and β_2 is the change in mean response for females (relative to males). We will consider this new model which takes into account the potential effects of gender to be the "full" model. If gender is ignored in the design of the study, the appropriate "reduced" model is (15.4). Let $SSE(F)$ denote the sum of squares for the full model—corresponding to the blocked design,

and let $SSE(R)$ denote the sum of squares for the reduced model—corresponding to the completely randomized design. Then we have:

$$SSE(F) = SSTO - SSR(X_1, X_2) = SSTO - [SSR(X_1) + SSR(X_2|X_1)] \tag{15.8}$$

If the number of observations in each block is the same, it can be shown that X_1 and X_2 are uncorrelated (i.e., orthogonal), hence $SSR(X_2|X_1) = SSR(X_2)$. Thus:

$$SSE(F) = SSTO - [(SSR(X_1) + SSR(X_2)] \tag{15.9}$$

From reduced model (15.4),

$$SSE(R) = SSTO - SSR(X_1) \tag{15.10}$$

and it follows from (15.9) and (15.10) that $SSE(F) = SSE(R) - SSE(X_2)$. Therefore $SSR(X_2)$ represents the reduction in the error sum of squares achieved with blocking.

2. When blocking on a nuisance factor is not possible at the design stage, variance reductions can sometimes be achieved at the analysis stage by including the nuisance factor as an additional predictor in the linear model for the response. Returning to the vitamin C example, suppose that blocking by prior gender was not possible. Nevertheless, model (15.7), which considers gender effects, could be employed at the analysis stage if the gender of each subject is recorded. By adding gender (X_2) as an additional predictor to model (15.4), we may realize variance reductions similar to those described in Comment 1 for blocking. This approach, called the *analysis of covariance*, is discussed in Chapter 22. ■

Measurements

The measurement process is another important element of experimental designs. Ideally, the measurement process should produce measurements that are unbiased and precise. *Measurement bias* can cause serious difficulties in the analysis of a study. An important source of measurement bias is due to unrecognized differences in the evaluation process. For example, a group of plants randomly assigned to a new fungicide treatment might unintentionally be evaluated by the investigators to be responding better to the treatment than actually is the case because of a desire to show the new treatment to be effective. When the experimental unit is a person, knowledge of the treatment by the person may also influence the measurement obtained. For instance, a person who knows that the food additive is salt may respond differently in the evaluation of the tastiness of a vegetable than if the additive were unknown. This source of measurement bias can be minimized by concealing the treatment assignment to both the experimental subject and the evaluator. A study using this kind of concealment is called a *double-blind study*. When knowledge of the assignment is withheld only from the experimental subject or the evaluator, the study is called a *single-blind study*.

15.3 An Overview of Standard Experimental Designs

In this section, we give an overview of the best-known and most frequently used experimental designs. In addition, we provide linear statistical models associated with the most basic of these designs. Each of the designs introduced here will be treated in greater detail in the chapters that follow.

pletely Randomized Design

The simplest form of designed experiment is the *completely randomized design*. With this design, treatments are randomly assigned to the experimental units. This design is most useful when the experimental units are relatively homogeneous. Completely randomized designs are quite flexible; they can be used with any number of treatments and permit different sample sizes for different treatments.

The quick bread experiment is an example of a completely randomized design. This design was based on four treatments (low, medium, high, and very high temperatures) and eight experimental units (package mixes) leading to two replicates of each treatment. The results of the experiment have been summarized using a scatter plot in Figure 15.6. This scatter plot suggests that temperature does affect bread volume and that the largest volume is obtained by baking the bread at the high oven temperature.

A linear statistical model for the response is:

$$Y = \begin{bmatrix} \text{Overall} \\ \text{Constant} \end{bmatrix} + \begin{bmatrix} \text{Treatment} \\ \text{Effect} \end{bmatrix} + \begin{bmatrix} \text{Experimental} \\ \text{Error} \end{bmatrix} \tag{15.11}$$

We shall model the treatment effect as a qualitative factor having four levels. Thus, as described in Section 8.3, we can employ three indicator variables:

$$X_1 = \begin{cases} 1 & \text{if treatment 1} \\ 0 & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if treatment 2} \\ 0 & \text{otherwise} \end{cases}$$

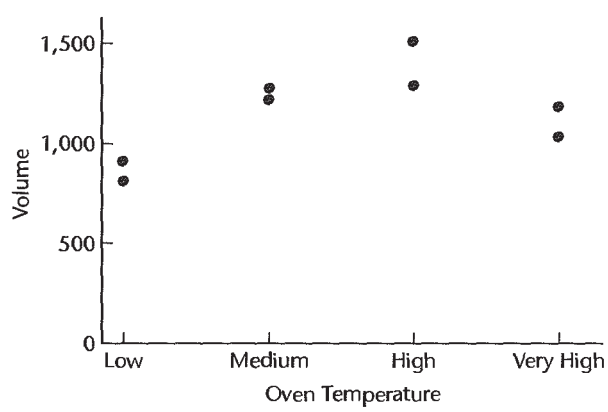
$$X_3 = \begin{cases} 1 & \text{if treatment 3} \\ 0 & \text{otherwise} \end{cases}$$

and we obtain for the jth replicate of treatment i:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3} + \varepsilon_{ij} \tag{15.12}$$

Notice that all of the predictors are indicator variables. For this reason, as we shall see in Chapter 16, the model in (15.12) is sometimes referred to as an *analysis of variance* model.

FIGURE 15.6 Summary Plot—Quick Bread Volume Example.



Assuming that the errors are independent $N(0, \sigma^2)$, testing for the presence of treatment effects is accomplished using the overall F^* test statistic (6.39b) for the presence of a regression relation:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$
$$H_a: \text{ not all } \beta_k \ (k = 1, 2, 3) \text{ equal zero} \tag{15.13}$$

If H_0 is rejected, the investigator may want to determine which levels of temperature lead to different volumes, which lead to similar volumes, and, perhaps, which temperature maximizes bread volume. These and other issues concerning the analysis of completely randomized designs are taken up in Chapters 16–18.

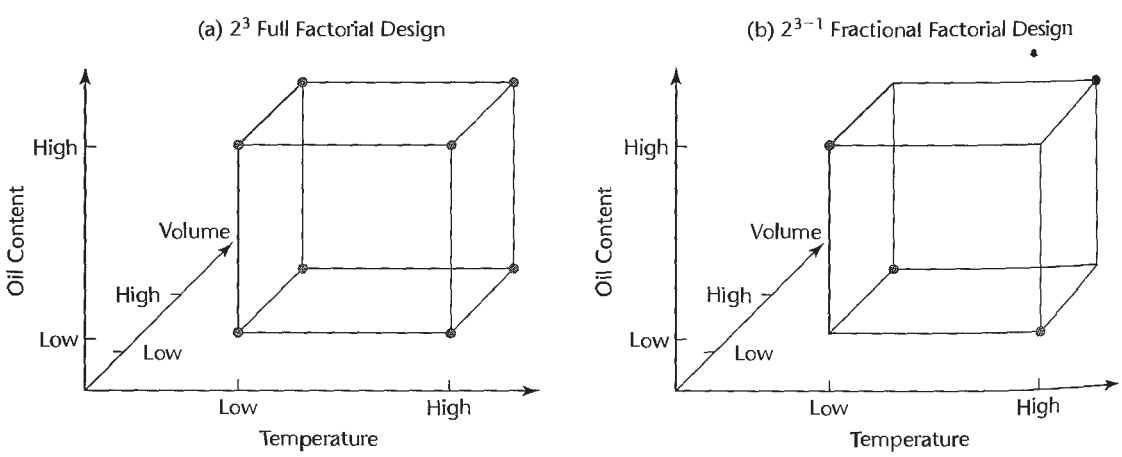
Factorial Experiments

Completely randomized designs can be used in single-factor studies or crossed, multifactor studies. Recall that in a crossed multifactor study the treatments correspond to the set of all possible combinations of the factor levels. Such designs are also referred to as *completely randomized factorial designs*.

The chemical yield experiment—whose treatment combinations are displayed in the two-way table in Figure 15.2a—is an example of a 2×3 factorial design. In another example, a sheet-aluminum manufacturer was interested in characterizing the effects of three coolant factors on the quality of the finish of the aluminum produced. During the manufacturing process, a molten aluminum strip is cooled using a mixture of water and oil at three different points during production. Factors and associated levels of interest were: coolant temperature (low, high), coolant oil percentage (low, high), and coolant volume (low, high). The $2^3 = 8$ treatment combinations are displayed in the cube plot in Figure 15.7a. This design is sometimes referred to as a $2 \times 2 \times 2$ or a completely randomized 2^3 factorial design.

Analysis of completely randomized factorial designs again involves the use of model (15.11) for completely randomized designs. However, when the treatments have factorial structure, it is often of interest to determine whether or not there are interaction effects among the individual factors. A linear statistical model that incorporates the factorial treatment

FIGURE 15.7 Full Factorial and Fractional Factorial Designs—Aluminum Rolling Mill Example.



structure has the following general form:

$$Y = \begin{bmatrix} \text{Overall} \\ \text{Constant} \end{bmatrix} + \begin{bmatrix} \text{First-Order} \\ \text{Treatment Effects} \end{bmatrix} + \begin{bmatrix} \text{Interaction} \\ \text{Treatment Effects} \end{bmatrix} + \begin{bmatrix} \text{Experimental} \\ \text{Error} \end{bmatrix} \tag{15.14}$$

In the sheet aluminum example, let X_1, X_2, and X_3 be the indicator variables that denote the presence ($X_i = 1$) or absence ($X_i = 0$) of each treatment. These are the predictors that correspond to "first-order" treatment effects in (15.14). The interactive treatment effects will be modeled using cross products, just as we did in regression. Here there are four cross product terms to be considered, X_1X_2, X_1X_3, X_2X_3, and $X_1X_2X_3$. Models for factorial experiments are taken up in detail in Chapters 19 and 24.

domized Complete Block Designs

As discussed in Section 15.2, in a blocked design, heterogenous experimental units are divided into homogeneous blocks, and then separate randomizations of treatments to experimental units are carried out within each block. These designs can increase the precision of the inferences concerning treatment effects. An example of a blocked experiment was displayed in Figure 15.5 for the vitamin C example. As a second example, we shall again consider the quick bread volume experiment. Suppose now, however, that the company owns two manufacturing plants—plant A and plant B—and of the eight package mixes available, four were produced in plant A and four were produced in plant B. Investigators expressed concern that the bread volumes might be affected by different processes and raw materials used at the two plants. However, it was felt that the four package mixes produced by each plant would be relatively homogeneous. For this reason, the investigators decided to run the experiment in two blocks of size four. The layout for this randomized complete-block design is shown in Figure 15.8.

Randomized complete block designs are often summarized graphically by producing a simple scatter plot of the results (as in Figure 15.6 for a completely randomized design), where the four responses within each block are connected by lines. Data for the blocked quick bread volume example are displayed in this fashion in Figure 15.9. Notice that there does appear to be a possible block effect: package mixes from plant B lead to consistently higher volumes than those from plant A.

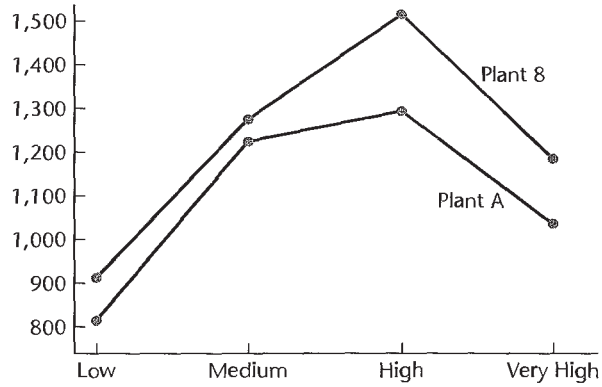
A linear statistical model for the response must reflect both the treatment (oven temperature) effect and the block (manufacturing plant) effect. The response model is:

$$Y = \begin{bmatrix} \text{Overall} \\ \text{Constant} \end{bmatrix} + \begin{bmatrix} \text{Treatment} \\ \text{Effect} \end{bmatrix} + \begin{bmatrix} \text{Block} \\ \text{Effect} \end{bmatrix} + \begin{bmatrix} \text{Experimental} \\ \text{Error} \end{bmatrix} \tag{15.15}$$

FIGURE 15.8 Randomized Complete Block Design—Quick Bread Volume Optimization Example.

Plant (Block)	Experimental Unit (Package Mix)			
	1	2	3	4
A	High	Low	Very High	Medium
B	Medium	High	Very High	Low

FIGURE 15.9 Summary Plot—Blocked Quick Bread Volume Optimization Example.



In the quick bread volume experiment, the four treatment levels are captured using three indicator variables, X_1, X_2, and X_3 as described above for a completely randomized design, and the two block levels can be modeled using a single indicator variable X_4:

$$X_4 = \begin{cases} 1 & \text{if package mix is from block 1} \\ 0 & \text{if package mix is from block 2} \end{cases}$$

and we obtain:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \beta_3 X_{ij3} + \beta_4 X_{ij4} + \varepsilon_{ij} \tag{15.16}$$

where β_1, β_2, and β_3 are the treatment effects, and β_4 is the block effect. Assuming that the errors are independent $N(0, \sigma^2)$, testing for presence of treatment effects is accomplished using F^* test statistic (2.70) for the alternatives:

$$\begin{aligned} H_0&: \beta_1 = \beta_2 = \beta_3 = 0 \\ H_a&: \text{not all } \beta_i = 0 \end{aligned} \tag{15.17}$$

The block effect, β_4, can be tested in similar fashion. The design and analysis of randomized complete block designs are discussed in greater detail in Chapter 21.

Nested Designs

Experiments involving purely nested factors are called *nested designs*. We discussed in Section 15.1 the use of nested factors in a study of the effects of operators on production yield in three manufacturing plants. Recall that three operators were selected in each of the three plants, and their production yields were recorded for five batches of product. The diagram of this experiment, shown in Figure 15.2b, indicates the nesting of operators within production plants: the first three operators are employed only in plant 1, the next three are employed only in plant 2, and the last three are employed uniquely in plant 3.

Multifactor experiments can involve both crossed and nested factors. In the production yield example, suppose that management was considering the use of control charts for monitoring of the production line. Then a new factor, statistical process control (SPC), is to be incorporated having two levels (SPC, No SPC). This factor can easily be crossed with manufacturing plant and operator, as shown in Figure 15.10. This is an example of a *crossed-nested design*.

E 15.10
ed-Nested
riment—
uction
ple.

SPC	Plant	Operator								
		1	2	3	4	5	6	7	8	9
SPC	1	X	X	X						
	2				X	X	X			
	3							X	X	X
No SPC	1	X	X	X						
	2				X	X	X			
	3							X	X	X

The design and analysis of experiments involving nested factors are discussed in Chapter 26.

Repeated Measures Designs

In one type of repeated measures design, the same subject (person, store, plant, etc.) receives all of the treatment combinations under study. For example, a repeated measures design was used to evaluate the effectiveness of a set of anti-inflammatory drugs, where the same patient was treated with each of the alternative drugs. Repeated measures designs are frequently used in product rating experiments, where the same consumer evaluates a set of products. We now consider one such example in detail.

Consider a taste-testing experiment to be conducted by a food manufacturer in which consumer acceptance of three breakfast cereal formulations is to be assessed. The three cereal formulations are identical except for the three levels (low, medium, and high) of sweetener to be used in the formulation. Each formulation is to be rated on a 10-point hedonic (likability) scale, and 12 consumers are available to rate the products.

With 12 consumers, a completely randomized experiment could be used, allowing for four complete replicates, as shown in Figure 15.11a. However, consumers differ considerably in their sensory perception of food products (e.g., children prefer higher levels of sweetness, adults prefer lower levels of sweetness) and so our experimental units would not be particularly homogeneous. One could consider blocking on age, but an even more effective approach is to have each consumer rate all three products. With this setup, each consumer becomes a block, and the experimental units are the separate evaluations conducted by each consumer. The layout for this repeated measures design in given in Figure 15.11b. This study involves repeated measures, because multiple responses are obtained from the same subject.

Suppose now that management is also interested in determining if the perceived level of wholesomeness has any effect on the ratings of the product by the consumers. Two levels of the perceived wholesomeness factor are to be employed, and half of the subjects are to be assigned to each level. Consumers in the control group are told only that the product they are about to test is a new breakfast cereal product. Consumers in the experimental group are told that the product is a new health cereal, manufactured from organic whole grains. As before, each consumer then tastes and evaluates three versions of the cereal based on low, medium, and high levels of sweetener. The layout for this experiment is shown in Figure 15.11c.

FIGURE 15.11 Alternative Designs—Food Product Taste-Testing Example.

(a) Completely Randomized Design

Consumer	Formulation
1	F_2
2	F_1
3	F_1
4	F_2
5	F_3
6	F_2
7	F_3
8	F_3
9	F_1
10	F_3
11	F_2
12	F_1

(b) Repeated Measures Design

Consumer	Formulations		
1	F_2	F_1	F_3
2	F_1	F_2	F_3
3	F_1	F_2	F_3
4	F_2	F_1	F_3
5	F_3	F_2	F_1
6	F_3	F_1	F_2
7	F_3	F_2	F_1
8	F_3	F_2	F_1
9	F_1	F_3	F_2
10	F_3	F_1	F_2
11	F_2	F_3	F_1
12	F_1	F_3	F_2

(c) Split-Plot Repeated Measures Design

Perceived Wholesomeness	Consumer	Formulations		
Wholesome	1	F_2	F_1	F_3
	2	F_1	F_2	F_3
	3	F_1	F_2	F_3
	4	F_2	F_1	F_3
	5	F_3	F_2	F_1
	6	F_3	F_1	F_2
Not Wholesome	7	F_3	F_2	F_1
	8	F_3	F_2	F_1
	9	F_1	F_3	F_2
	10	F_3	F_1	F_2
	11	F_2	F_3	F_1
	12	F_1	F_3	F_2

This is an example of a second type of repeated measures design, in which randomizations at two distinct levels are being conducted. The three levels of sweetener are randomly applied to the three individual tastings by a given consumer; thus, for comparisons involving levels of sweetness in the product formulation, the individual tastings are the experimental units. Similarly, perceived wholesomeness is applied directly to consumers. Thus, for comparisons involving the levels of perceived wholesomeness, consumers are the experimental units. When the subject serves as an experimental unit for another treatment, the repeated measures design is sometimes referred to as a *split-plot design*.

The design and analysis of repeated measures and split-plot designs are taken up in Chapter 27.

Incomplete Block Designs

Until now, we have only discussed the use of blocking where each block contains one or more replicates of the treatment combinations. Can blocking be used when block sizes are smaller than the number of treatments? The answer to this question is "yes," although such designs are slightly more difficult to analyze.

Consider again the breakfast cereal formulation example, only now we shall assume that five alternative product formulations, instead of just three, are to be evaluated by consumers. It is well known that a consumer's ability to discriminate among similar products in taste-testing diminishes rapidly with the number of samples tested. Generally, no more than three taste evaluations are permitted. With this restriction, we see that it will not be possible for any given consumer to evaluate all five product formulations in a single session. Since only three of the five alternatives can be rated, each consumer represents a single, incomplete block.

An effective experimental arrangement can still be achieved, however, through the use of a *balanced incomplete block design*, or *BIBD*. In a balanced incomplete block design, every

RE 15.12
ced
plete
gn—Food
uct
Testing
ple.

Consumer (Block)	Product Formulation				
	1	2	3	4	5
1	X	X	X		
2	X	X		X	
3	X	X			X
4	X		X	X	
5	X		X		X
6	X			X	X
7		X	X	X	
8		X	X		X
9		X		X	X
10			X	X	X

treatment appears with every other treatment in the same block the same number of times. In this way, comparisons between pairs of treatments can be carried out on a within-block basis, thus eliminating block-to-block heterogeneity.

A BIBD with five treatments and block size three is shown in Figure 15.12. Note that every treatment occurs together with every other treatment exactly three times. For example, formulations 1 and 2 appear together in blocks 1, 2, and 3. Formulations 1 and 3 appear together in blocks 1, 4, and 5—and so on. Note also that this BIBD requires 10 blocks or subjects. In the breakfast cereal formulation example, 12 subjects were available; however, no BIBD exists for five treatments in 12 blocks of size three. Thus, in order to use this particular BIBD for the breakfast cereal formulation example, only 10 subjects would need to be available.

Another form of incomplete block design, with block size equal to one, is called a latin square design. We take up the construction and analysis of BIBDs and latin square designs in Chapter 28.

Two-Level Factorial and Fractional Factorial Experiments

Factorial designs are effective tools for characterizing the joint effects of multiple factors. However, the number of treatments, which is a product of the numbers of factor levels for each factor, grows rapidly with the number of factors. For example, a crossed three-factor experiment, where each factor has three levels, will involve $3^3 = 27$ treatment combinations. One way of economizing will be to limit each factor to two levels, which reduces the number of treatment combinations to $2^3 = 8$ treatment combinations. The sheet aluminum production example discussed earlier and displayed in Figure 15.7a is an example of a 2^3 factorial design. Two-level designs are extremely useful in exploratory or screening studies where the objective is to identify the most important factors from a larger set of potential factors. When the factors are quantitative, screening experiments are usually followed up with a more exacting experiment, such as a response surface experiment, discussed on the next page.

If there are a large number of factors to be screened, it may be impractical to run a single complete replicate. For example, a complete replicate of a six-factor, two-level experiment requires $2^6 = 64$ treatment combinations. In such cases, a subset of the treatment

combinations can be chosen so that little or no information is lost concerning important main effects and low-order interactions. This chosen subset of treatment combinations is referred to as a *fractional factorial design*.

Consider again the aluminum production example. An alternative fractional factorial design is shown in Figure 15.7b. The half-fraction displayed is based on four (carefully chosen) treatment combinations from the full factorial in Figure 15.7a that will permit estimation of the three factor effects, but with no information about the interactive effects of the factors.

Two-level factorial and fractional factorial designs are discussed in Chapter 29.

Response Surface Experiments

When all factors are quantitative, two-level experiments often provide good information on linear trends in each factor. If there is concern that the response will be substantially convex (bowl-shaped) or concave (mound-shaped), or if the objective of the experiment is to determine precisely the factor levels that lead to an optimum response, use of just two levels will not be adequate. *Response surface designs* were developed for use in these situations. These designs are applicable when all experimental factors are quantitative, and the true response function can be well approximated by a second-order polynomial. Once the second-order response model has been estimated, a detailed mapping of the regression surface can be obtained using three-dimensional response surface plots, contour plots, and conditional effects plots, such as those shown in Figures 8.8 and 8.9 on pages 310–311.

Methods for design and analysis of response surface experiments are taken up in Chapter 30.

15.4 Design of Observational Studies

Observational studies are distinct from experimental studies in that random assignments of factor levels to the experimental units do not occur. Therefore, designed observational studies do not directly demonstrate cause-and-effect relationships between the explanatory factors and the response. They can establish association between explanatory factors and a response, and provide the basis for further study of potential cause-and-effect relationships. To infer causality, potential confounding variables would need to be identified, and subgroup analysis performed to try to rule out possible alternative causal factors. Some observational studies are conducted for descriptive purposes only, such as when various characteristics of a group are summarized. These studies, which are sometimes referred to as analytical surveys or case studies, will not be considered further.

Observational studies have been classified in many ways, but we will consider three commonly used categories, namely, cross-sectional studies, prospective studies, and retrospective studies. Prospective and retrospective observational studies are often designed to study potential causal relationships, and are closer in spirit to experimental studies. We turn now to a discussion of cross-sectional observational studies.

Cross-Sectional Studies

A cross-sectional observational study involves measurements taken from one or more populations or subpopulations at a single point in time or a single time interval. Exposure to a potential causal factor and the response are determined simultaneously. Cross-sectional

studies are sometimes said to provide a "snapshot" of the factors and outcome variable. For example, a cross-sectional study of household incomes by geographic location in a major metropolitan area was conducted by a marketing research department of a luxury SUV manufacturer. The subpopulations consisted of the postal zip-code areas within the city. The response variable was household income, and the explanatory factor was geographical area. Random samples of households were selected within each geographic zip-code area. The objective of the study was to carry out comparisons of household income among subpopulations.

The Minnesota Department of Transportation road use study, discussed in Chapter 11, page 464, is another example of a cross-sectional observational study. Here, data on the average annual daily traffic for a variety of road sections were obtained for a single time interval along with various characteristics of the road sections. Multiple regression techniques were then used to identify important predictors of the outcome variable, namely, the average annual daily traffic for the various road sections.

Cross-sectional studies may be prestratified or poststratified to form subpopulations. In a prestratified cross-sectional study, potential explanatory factors are used to stratify the population into subpopulations, and random samples are obtained within each of the subpopulations. Alternatively, cross-sectional study data can be poststratified by the explanatory factors. Comparisons of outcome measurements among the poststratified subpopulations are then obtained.

Prospective Studies

In a *prospective observational study,* one or more groups are formed in a nonrandom manner according to the levels of a hypothesized causal factor, and then these groups are observed over time with respect to an outcome variable of interest. Prospective studies answer the question: "What is going to happen?" The teaching effectiveness example, discussed in Section 15.1, is an example of a prospective observational study. Faculty either attended or did not attend a teaching workshop on a voluntary basis. Here the groups were self-selected. At the end of the following academic year, teaching effectiveness scores were obtained for all faculty, and it was found that the average effectiveness of faculty who attended the seminar was greater than that for the group of faculty who elected not to attend the seminar. The fact that the "treatment" preceded the response in time is suggestive of a potential cause-and-effect relationship, but, as noted earlier, an experiment is required for "proof." Prospective studies are also known as *cohort studies* and can often be analyzed using regression models or analysis of variance techniques.

Prospective observational studies may be conducted utilizing historical records. For example, from the medical histories obtained from a health maintenance organization, researchers were able to identify women who received estrogen supplements over long periods of time, and women who did not. A prospective study was then carried out to explore potential links between estrogen therapy and heart disease.

Retrospective Studies

In a *retrospective observational study,* groups are defined on the basis of an observed outcome, and the differences among the groups at an earlier point in time are identified as potential causal effects. Retrospective studies answer the question: "What has happened?" A famous retrospective study carried out in the 1950s compared the lifestyles of individuals

with lung cancer to those of individuals who did not have lung cancer. These studies led to hypotheses about the causal effects of cigarette smoking. Notice that in comparison with a prospective study, the roles of the response and explanatory variables are reversed. In a prospective study, the response is the effect (e.g., increased teaching effectiveness) and the explanatory factor is the hypothesized cause (e.g., workshop attendance). In a retrospective study, the response variable is the hypothesized cause (e.g., smoking), and the predictor or explanatory factor is the potential effect (e.g., presence or absence of lung cancer).

Retrospective studies are sometimes used in manufacturing process monitoring. For example, a manufacturer may suddenly receive reports of a cluster of failures of a particular product part while in use in the field. From records, it may be possible to obtain characteristics of the manufacturing process at the times that the failed parts were produced, and to compare these characteristics to those corresponding to other parts that have not failed. This may suggest manufacturing operating conditions that led to the production of the defective parts.

The surgical unit example discussed in Chapter 9 on page 350 is a retrospective observational study. Patients who had a particular type of liver operation and died were selected for study. Preoperative factors were then used to try to predict survival times following the operation using multiple regression techniques.

Retrospective studies have an advantage over comparable prospective studies in terms of efficiency when an outcome of interest occurs infrequently. Epidemiologists frequently use retrospective designs to study rare-event diseases. For example, a prospective study of the effects of a diet on the incidence of stomach cancer may well require a lengthy period of time and many more subjects than would be required by a retrospective study. The retrospective study would identify persons who have stomach cancer (referred to as cases) and persons who do not have stomach cancer (referred to as controls) and look back in time to assess differences in eating habits. Retrospective studies that require subjects or investigators to construct case histories from memory are susceptible to recall bias, and should be used with caution. The process-monitoring study just discussed is an example of an *archival* retrospective study, where the necessary historical data exists. Archival studies do not suffer the same susceptibility to recall bias.

Retropective studies are also known as *case-control* and *ex post facto* studies.

Matching

In our discussion of designed experiments, we noted that if the experimental units were heterogeneous, the experimental error can be reduced and the precision of the comparisons among treatments can be improved through the use of blocking techniques. In an observational study, treatments are not assigned at random to experimental units, so blocking is not technically possible. However, *matching,* a procedure that is analogous to blocking, can be employed to achieve similar reductions in variance.

Returning to the observational study of teaching effectiveness, recall that the treatments (attend workshop, do not attend workshop) were not randomly assigned to the faculty members. Rather, about half of the faculty volunteered to attend the workshop. As teachers, faculty in business schools are relatively heterogeneous. They vary in terms of such factors as age, gender, field or department, quantitative orientation, prior teaching effectiveness, and so on. In a *matched study,* each faculty member who attended the workshop is matched, on the basis of nuisance factors such as those just noted, to another faculty member who did not attend the workshop. Faculty who are not matched are not included in the study. In

effect, each match leads to a "block" size of two. Any observed differences in the teaching effectiveness between the matched faculty members is due either to the treatment factor—here workshop attendance—or to other unidentified or uncontrolled nuisance factors.

There are a number of approaches used for identifying matches. If the nuisance factor is categorical, taking on just a few distinct values (e.g., male, female), a match occurs if two cases fall into the same category or class. This is called *within-class matching*. If more than one categorical nuisance factor is present, for example, grade and gender, a match occurs if two cases fall into the same category for both of the confounding factors. When the confounding factor is discrete or continuous, for example, pretest score on a 0–100 basis, it is common to change the factor into a categorical factor—for example by creating three pretest categories—and then again declaring a match if two cases fall into the same category.

A more precise method of matching discrete or continuous confounding factors is called *caliper matching* or *interval matching*. In caliper matching, two values of a confounding factor are considered to have matched if their absolute difference is less than some prespecified value. For example, two faculty may be considered a match on the age dimension if the absolute difference in their ages is not greater than five years. A disadvantage of caliper matching is that if the specified maximum difference is too small, it may be difficult to find a sufficient number of matches to perform the study.

Other methods of matching continuous confounding factors include *mean matching* or *balancing,* and *nearest available matching*. Reference 15.2 gives a complete discussion of matching methods.

Comment

An alternative to matching at the design stage is the use of covariance analysis. A brief introduction to this approach was given in a comment in Section 15.2. The same adjustment techniques can be used in the analysis of observational studies for known confounding factors that are not held constant. Again, these techniques are discussed in Chapter 22. ■

15.5 Case Study: Paired-Comparison Experiment

In this section we consider the design and analysis of the *paired-comparison* or *matched-pairs* design. This is the most basic form of a randomized complete block design, involving just two treatments arranged in blocks of size two. Because the example uses subjects as blocks, the experimental layout also represents the simplest instance of a repeated measures design. The example will also serve to illustrate the analysis techniques used in a matched observational study.

The objective of a product-improvement project at a major pharmaceutical company was to reduce the sensitivity of skin to the injection of an allergen. A new experimental allergen was developed and dermatologists were interested in comparing the new formulation to the existing product. Reactions to allergen injections vary greatly from person to person, and it was decided that all comparisons of the new treatment and standard control treatment should be conducted on a within-subject basis. Thus a randomized complete block experiment was utilized, where blocks correspond to subjects, and each subject was injected with both the experimental and control allergens, once in each arm. Here, the experimental units are

the subjects' arms, and each block consists of two experimental units. Randomization is accomplished by randomly assigning the treatments to the right or left arms for each subject.

Twenty subjects were randomly chosen from a pool of available subjects for testing. The experimental layout, randomization, and results of the 40 tests are shown in Table 15.1. The response, skin sensitivity, is obtained by measuring the diameter of the red area surrounding the injection in centimeters. The results are plotted, with plot symbols from the same block connected, in Figure 15.13. The preponderance of negative slopes in the plot suggests that the experimental formulation leads to reduced skin sensitivity.

From (15.15) a linear statistical model for the experiment is:

$$Y_{ij} = \beta_0 + \beta_1 X_{i1} + \sum_{j=2}^{20} \beta_j X_{ij} + \varepsilon_{ij} \qquad i = 1, 2 \tag{15.18}$$

where:

$$X_{i1} = \begin{cases} 1 & \text{if experimental treatment} \\ 0 & \text{if control treatment} \end{cases}$$

$$X_{ij} = \begin{cases} 1 & \text{if response is from subject } j-1, \text{ for } j = 2, \ldots, 20 \\ 0 & \text{otherwise} \end{cases}$$

TABLE 15.1 Data and Descriptive Statistics—Skin Sensitivity Experiment.

Subject	Control Treatment	Experimental Treatment	Within-Subject Difference
1	0.59	0.43	−0.16
2	0.69	0.53	−0.16
3	0.82	0.58	−0.24
...	...	...	...
18	0.85	0.60	−0.25
19	0.85	0.65	−0.20
20	0.74	0.58	−0.16
Sample Mean:	.7315	.5400	−.1915
Sample Std Dev:	.0758	.0807	.0501

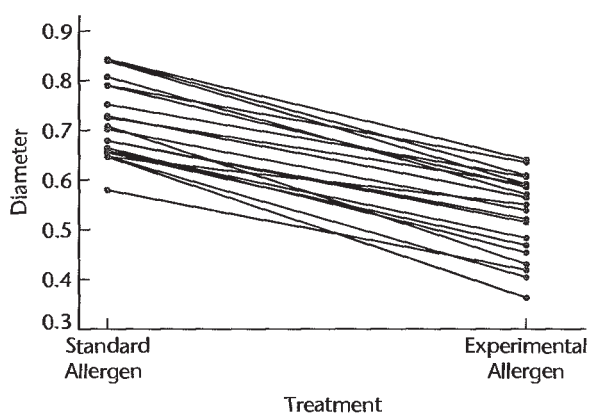


FIGURE 15.13 Summary Plot—Allergen Sensitivity Example.

The dermatologists were primarily interested in determining whether the experimental allergen formulation led to reduced skin sensitivity, but they allowed for the possibility that it might increase skin sensitivity. They thus tested the alternatives:

$$H_0\colon\ \beta_1 = 0 \qquad H_a\colon\ \beta_1 \neq 0 \tag{15.19}$$

MINITAB regression results for this model are shown in Figure 15.14. We see that the estimated treatment effect is $b_1 = -.1915$, and the 19 estimated block effects are $b_2 = -.1500$, $b_3 = -.0500$, and so on. The test statistic corresponding to the estimated treatment effect is $t^* = -17.10$. To carry out the test indicated in (15.19) at the $\alpha = .05$ level, we require $t(.975; 19) = 2.093$. Since $|t^*| = 17.10 > 2.093$, we conclude H_a, that $\beta_1 \neq 0$. Since b_1 was negative, the dermatologists concluded that the new formulation significantly reduces skin irritation.

Note that the investigators were not primarily interested in determining whether or not subject (block) effects were present. Blocking was used here to increase the precision of the comparisons between the experimental and control treatments and it was fully expected that significant subject-to-subject differences would be present. Nevertheless, a test for the

FIGURE 15.14 MINITAB Regression Results—Allergen Skin Sensitivity Example.

Predictor	Coef	SE Coef	T	P
Constant	0.75575	0.02566	29.45	0.000
X1	−0.19150	0.01120	−17.10	0.000
X2	−0.15000	0.03541	−4.24	0.000
X3	−0.05000	0.03541	−1.41	0.174
X4	0.04000	0.03541	1.13	0.273
X5	0.06500	0.03541	1.84	0.082
X6	−0.14000	0.03541	−3.95	0.001
X7	−0.08500	0.03541	−2.40	0.027
X8	0.03000	0.03541	0.85	0.407
X9	0.04000	0.03541	1.13	0.273
X10	−0.08000	0.03541	−2.26	0.036
X11	0.08000	0.03541	2.26	0.036
X12	0.01000	0.03541	0.28	0.781
X13	−0.02500	0.03541	−0.71	0.489
X14	−0.12000	0.03541	−3.39	0.003
X15	−0.05000	0.03541	−1.41	0.174
X16	−0.08500	0.03541	−2.40	0.027
X17	−0.07500	0.03541	−2.12	0.048
X18	−0.04500	0.03541	−1.27	0.219
X19	0.06500	0.03541	1.84	0.082
X20	0.09000	0.03541	2.54	0.020

S = 0.03541 R−Sq = 96.0% R−Sq(adj) = 91.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	20	0.578750	0.028937	23.07	0.000
Residual Error	19	0.023828	0.001254		
Total	39	0.602577			

effect of blocking can be carried out using (2.70). The alternatives here are:

$$H_0\colon\ \beta_2 = \cdots = \beta_{20} = 0$$
$$H_a\colon\ \text{not all } \beta_k\ (k = 2, 3, \ldots, 20) \text{ equal zero} \tag{15.20}$$

For these data, it can be shown that blocking was effective in significantly reducing the error variance.

15.6 Concluding Remarks

In this chapter, we have outlined the basic differences between observational and experimental studies, and we have described how experimental studies lead to a much firmer basis for making inferences concerning cause and effect. We have also previewed the main types of designed observational and experimental studies. In doing so, we have shown that the statistical models studied in Chapters 1–14 provide the bases for statistical analysis of well-designed studies.

In the chapters to follow, we will consider the design and analysis of experimental and observational studies in greater detail. Design issues not yet discussed, such as sample size planning and power considerations, will be taken up for each design type. There will also be an increased emphasis on the analysis of categorical factors. The linear model for that case is called the analysis of variance (ANOVA) model. While standard regression approaches can always be used, we will see that when the study design is balanced, the use of ANOVA greatly simplifies the analysis. If the study is not balanced, we will simply return to the regression approach. Finally, when all factors are treated as categorical, the analysis frequently focuses on comparisons among treatments or factor-level combinations. A discussion of such *multiple comparison* procedures will accompany nearly every class of study design.

Cited References

15.1. Cochran, W. G., and G. M. Cox. *Experimental Designs*. 2nd ed. New York: John Wiley & Sons, 1992.

15.2. Cochran, W. G. *Planning and Analysis of Observational Studies*. New York: John Wiley & Sons, 1983.

Problems

15.1. In an experiment to study the effect of the location of a product display in drugstores of a chain, the manager of one of the drugstores rearranged the displays of other products so as to increase the traffic flow at the experimental display. Does this action potentially lead to either selection bias or measurement bias? Discuss.

15.2. In a study of the effect of size of team on the volume of communications within the team, can a double-blind procedure be utilized? A single-blind procedure? Discuss.

15.3. Four treatments (T_1, T_2, T_3, T_4) are to be studied in an experiment with a completely randomized design using three replicates. Obtain the randomized assignments of treatments to experimental units.

15.4. Three treatments (T_1, T_2, T_3) are to be studied in an experiment with a completely randomized design using five replicates. Obtain the randomized assignments of treatments to experimental units.

15.5. Give an example of an experiment where a control group would not be necessary.

15.6. Five treatments (T_1, T_2, T_3, T_4, T_5) are to be studied in a randomized complete block design with four blocks. Obtain the randomized assignments of treatments to experimental units.

15.7. In a study to evaluate the quality of three alternative recipes for salsa, six containers of salsa—two from each of the three recipes—were randomly assigned to six taste panels. Each taste panel consisted of a team of four trained taste-testers. Each panel reached a consensus score for the assigned recipe. What is the experimental unit in this study? Why?

15.8. Three high schools participated in a study to evaluate the effectiveness of a new computer-based mathematics curriculum. In each school, four 24-student sections of freshman algebra were available for the study. The two types of instruction (standard curriculum, computer-based curriculum) were randomly assigned to the four sections in each of the three schools. At the end of the term, a standard mathematics achievement test was given to each of the 24 students in each section.

a. Is this study experimental, observational, or mixed experimental and observational? Why?

b. Identify all factors, factor levels, and factor-level combinations.

c. What type of study design is being implemented here?

d. What is the basic unit of study?

*15.9. An economist compiled data on productivity improvements last year for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditures for research and development in the past three years (low, moderate, high).

a. Is this study experimental, observational, or mixed experimental and observational? Why?

b. Identify all factors, factor levels, and factor-level combinations.

c. What type of study design is being implemented here?

d. What is the basic unit of study?

15.10. In a study to investigate the effect of color of paper (blue, green, orange) on response rates for questionnaires distributed by the "windshield method" in supermarket parking lots, four supermarket parking lots were chosen in a metropolitan area and 10 questionnaires of each color were assigned at random to cars in the parking lots.

a. Is this study experimental, observational, or mixed? Why?

b. Identify all factors, factor levels, and factor-level combinations.

c. What type of study design is being implemented here?

d. What is the basic unit of study?

15.11. A rehabilitation center researcher was interested in examining the relationship between physical fitness prior to surgery of persons undergoing corrective knee surgery and the time required in physical therapy until successful rehabilitation. Data on the number of days required for successful completion of physical therapy and the prior physical fitness status (below average, average, above average) were collected.

a. Is this study experimental, observational or mixed? Why?

b. Identify all factors, factor levels, and factor-level combinations.

c. What type of study design is being implemented here?

d. What is the basic unit of study?

15.12. In a study of the effect of applicant's eye contact (yes, no) and personnel officer's gender (male, female) on the personnel officer's assessment of likely job success of an applicant, personnel officers were shown a front view photograph of an applicant's face and were asked

to give the person in the photograph a success rating score. Half of the officers in each gender group were chosen at random to receive a version of the photograph in which the applicant made eye contact with the counselors. The other half received a version in which there was no eye contact. Data were collected on success ratings.

a. Is this study experimental, observational, or mixed? Why?

b. Identify all factors, factor levels, and factor-level combinations.

c. What type of study design is being implemented here?

d. What is the basic unit of study?

15.13. An automotive engineer was interested in the effect of four alternative rubber compounds on the life of automobile tires. To carry out the study, five tires were manufactured from each of the four compounds and five automobiles were obtained for testing. With each automobile, the four tire types were assigned at random to the four wheels. Each automobile was driven for 40,000 miles and the amount of wear on each of the four tires was recorded.

a. What type of study is this, experimental, observational, or mixed? Why?

b. What is the basic unit of study?

c. What factors and factor levels are being studied here?

d. What type of study design is being implemented here?

e. Suppose that six compounds were under study instead of four. What type of study design is suggested?

*15.14. A research laboratory was developing a new compound for the relief of severe cases of hay fever. The amounts of two active ingredients (low, medium, high) in the compound were varied at three levels each using 18 volunteers. Randomization was used in assigning volunteers to each of the treatment combinations. Data were collected on hours of relief.

a. Is this study experimental, observational, or mixed? Why?

b. Identify all factors, factor levels, and factor-level combinations.

c. Describe how randomization would be performed in this study.

d. What type of study design is being implemented here?

e. What is the basic unit of study?

15.15. Kidney failure patients are commonly treated on dialysis machines that filter toxic substances from the blood. The approximate dose for effective treatment depends on, among other things, duration of treatment and weight gains between treatments as a result of fluid buildup. To study the effects on the number of days hospitalized (attributable to the disease) during a year, a random sample of patients who had undergone dialysis treatment at a large dialysis facility was obtained. Treatment duration was categorized into two groups (short duration, long duration). Average weight gain between treatments during the year was categorized in three groups (slight, moderate, substantial).

a. Is this study purely experimental or observational or mixture of both? Why?

b. Identify all factors, factor levels, and factor-level combinations.

c. What type of study design is being implemented here?

d. What is the basic unit of study?

15.16. In a study of recall memory, three different questionnaires (A, B, C) were administered to nine subjects at three different times three months apart about the number of trips to a shopping center during the preceding three months. Each time a different questionnaire was used and the order of the assignments of questionnaires for each subject was randomized.

a. Is this study purely experimental or observational or mixture of both? Why?

b. Identify all factors, factor levels, and factor-level combinations.

c. What type of study design is being implemented here?

d. What is the basic unit of study?

15.17. A chemical company wished to study the consistency of the strength of one of its liquid chemical products. The product is made in batches in large vats and then is barreled. The barrels are subsequently stored for a period of time in a warehouse. To examine the consistency of the strength of the chemical, an analyst randomly selected five different batches of the product from the warehouse and then selected four barrels per batch at random. Three determinations per barrel were made.

a. Is this study purely experimental or observational or mixture of both? Why?

b. Identify all factors, factor levels, and factor-level combinations.

c. What type of study design is being implemented here?

d. What is the basic unit of study?

15.18. A study was undertaken in an effort to reduce the occurrence of dents in a windshield molding manufacturing process. The dents are caused by pieces of metal or plastic that are carried into the dies during stamping and forming operations. Four factors were identified for use in an eight-run experiment: poly-film thickness—used to protect the metal strip during manufacturing to reduce surface blemishes (low, high), oil mixture ratio for surface lubrication (low, high), operator glove type (cotton, nylon), underside oil coating (no coating, coating). During each run of the experiment, 1,000 moldings were fabricated in a batch; the response (Y) is the number of defect-free moldings produced.

a. Is this study purely experimental or observational or mixture of both? Why?

b. Identify all factors, factor levels, and factor-level combinations.

c. What type of study design is being implemented here?

d. What is the basic unit of study?

15.19. Assemblers in an electronics firm attach components to a newly developed "board" to be used in automatic-control equipment in manufacturing plants. A study was conducted to determine the effect of sequence of assembling the components (sequence 1, sequence 2, sequence 3) on the mean time to assemble a board. Potential nuisance factors are gender of the assembler (male, female) and amount of the assembler's prior experience (under 18 months, 18 months or more). Assume that the following assemblers are available for the study: four males with under 18 months experience, three females with under 18 months experience, five male assemblers with 18 months or more experience, and four females with 18 months or more experience.

a. Suggest an experimental design that accounts for the two nuisance factors. What type of study design did you recommend?

b. Show how the randomization is to be carried out for your study design in part (a).

c. What is the experimental unit in your study design?

*15.20. An experiment involving the case hardening of lightweight shafts machined from bars of an alloy was run to study the effects of the amount of chemical agent added to the alloy in a molten state (low, high), the temperature of the hardening process (low, high), and the time duration of the hardening process (low, high). Outcome data measured the hardness of the rods tested. It will be possible to machine 16 bars in the study.

a. Suggest an experimental plan for the study. What type of study design did you recommend?

b. Show how the randomization is to be carried out for your study design in part (a).

c. What is the experimental unit in your study design?

15.21. An experiment is to be conducted to compare the effectiveness of four household detergents. The response is to be the degree of stain removal from a section of clothing on a 10-point scale (1 = no stain removed, 10 = stain completely removed).

a. Identify the experimental unit.

b. Identify the experimental factor(s), levels, and any factor-level combinations if present.

c. Name two potential blocking factors.

d. Propose an experiment to accomplish the objectives of the study. How would you carry out the randomization?

15.22. An experiment is to be carried out to determine the optimal combination of microwave oven settings for microwave popcorn. Cooking time has three possible settings (3, 4, and 5 minutes) and cooking power has two settings (low power, high power). The response (to be minimized) is the number of burned plus the number of unpopped kernels.

a. Identify the experimental unit.

b. Identify the experimental factor(s), levels, and any factor-level combinations if present.

c. Name two potential blocking factors.

d. Propose an experiment to accomplish the objectives of the study. How would you carry out the randomization?

*15.23. Refer to the skin sensitivity example data in Table 15.1.

a. Test the hypothesis that the mean within-subject difference is zero using the t test for paired observations in (A.69) using $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of your test? Do your results agree with those obtained on page 671? Should they agree?

b. Conduct the test for block effects using $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of your test? Is your conclusion of primary interest in this study? Why or why not?

Exercise

15.24. Show that (15.5b) follows from (15.5a) for model (15.4).

Chapter 6

Single-Factor Studies

In the last chapter, we presented a general introduction to the design of experimental and observational studies. In this and the next two chapters, we shall focus on the design and analysis of single-factor studies. This includes the development of single-factor analysis of variance (ANOVA) model, the analysis and interpretation of factor level means, assessment of model adequacy, and the use of remedial measures when necessary.

In this chapter, we briefly review the design of single-factor studies and the associated linear models, then discuss the relation between regression and analysis of variance. In the next few sections we introduce in detail the single-factor ANOVA model and the associated F test for equality of factor level means. We then consider alternative formulations of the ANOVA model, followed by a regression approach to the single-factor ANOVA model. In the last few sections, we consider a nonparametric randomization test as an alternative to the ANOVA test, and, finally, we present two methods for the planning of sample sizes in single-factor studies.

16.1 Single-Factor Experimental and Observational Studies

Single-factor experimental and observational studies are the most basic form of comparative studies used in practice. In a single-factor experimental study, the treatments correspond to the levels of the factor, and randomization is used to assign the treatments to the experimental units. In the following we present three examples of single-factor studies. The first two examples are experimental studies, and the third is a cross-sectional observational study. We then briefly review the approach described in Chapter 15 for modeling a single-factor study.

Example 1 A hospital research staff wished to determine the best dosage level for a standard type of drug therapy to treat a medical condition. In order to compare the effectiveness of three dosage levels, 30 patients with the medical problem were recruited to participate in a pilot study. Each patient was randomly assigned to one of the three drug dosage levels. Randomization was performed in such a way that an equal number of patients ended up being evaluated for each drug dosage level, i.e., with exactly 10 patients studied in each drug dosage level group. This is an example of completely randomized design, based on a single, three-level quantitative factor. This particular design is said to be *balanced,* because each treatment is replicated the same number of times.

Example 2 In an experiment to investigate absorptive properties of four different formulations of a paper towel, five sheets of paper towel were randomly selected from each of the four types (formulation 1, formulation 2, formulation 3, and formulation 4) of paper towel. Twenty 6-ounce beakers of water were prepared, and the twenty paper towel sheets were randomly assigned to the beakers. Paper towels were then fully submerged in the beaker water for 10 seconds, withdrawn, and the amount of water absorbed by each paper towel sheet was determined. This is an example of a completely randomized design, based on a single, four-level qualitative factor.

Example 3 Four machines in a plant were studied with respect to the diameters of ball bearings they produced. The purpose of the study was to determine whether substantial differences in the diameters of ball bearings existed between the machines. If so, the machines would need to be calibrated. This is an example of an observational study, as no randomization of treatments to experimental units occurred.

As we noted in Chapter 15, although the first two examples are experimental studies and the third is an observational study, the methods used for statistical analysis are generally the same. If the single factor has r levels, one approach to constructing a linear statistical model employs $r - 1$ indicator variables as predictors. Then the response for the jth replicate of the ith treatment or factor level is modeled:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \cdots + \beta_{r-1} X_{ij,r-1} + \varepsilon_{ij}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if treatment 1} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij2} = \begin{cases} 1 & \text{if treatment 2} \\ 0 & \text{otherwise} \end{cases}$$

$$\cdots$$

$$X_{ij,r-1} = \begin{cases} 1 & \text{if treatment } r-1 \\ 0 & \text{otherwise} \end{cases}$$

Recall that because all of the predictors are indicator variables, this model is sometimes referred to as an *analysis of variance* model.

For the first example, we have an alternative. Because the factor—dosage level—is quantitative with three levels, we could also model its effect using a second-order (or lower-order) polynomial regression model, as described in Section 8.1. Specifically, two choices for the first example are:

$$Y_{ij} = \beta_0 + \beta_1 X_{ij1} + \beta_2 X_{ij2} + \varepsilon_{ij} \qquad \text{ANOVA Model}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if treatment 1} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij2} = \begin{cases} 1 & \text{if treatment 2} \\ 0 & \text{otherwise} \end{cases}$$

or, employing second-order polynomial model (8.1):

$$Y_{ij} = \beta_0 + \beta_1 x_{ij} + \beta_{11} x_{ij}^2 + \varepsilon_{ij} \qquad \text{Regression Model}$$

where:

x_{ij} = centered dosage level amount for the ijth case

In the next section, we discuss the choice between the two types of models.

16.2 Relation between Regression and Analysis of Variance

Regression analysis, as we have seen, is concerned with the statistical relation between one or more predictor variables and a response variable. Both the predictor and response variables in ordinary regression models are quantitative. The regression function describes the nature of the statistical relation between the mean response and the levels of the predictor variable(s).

We encountered the use of analysis of variance in our consideration of regression. It was used there for a variety of tests concerning the regression coefficients, the fit of the regression model, and the like. The analysis of variance is actually much more general than its use with regression models indicated. Analysis of variance models are a basic type of statistical model. They are concerned, like regression models, with the statistical relation between one or more predictor variables and a response variable. Like regression models, analysis of variance models are appropriate for both observational data and data based on formal experiments. Further, as in the usual regression models, the response variable for analysis of variance models is a quantitative variable. Analysis of variance models differ from ordinary regression models in two key respects:

1. The explanatory or predictor variables in analysis of variance models may be qualitative (gender, geographic location, plant shift, etc.).
2. If the predictor variables are quantitative, no assumption is made in analysis of variance models about the nature of the statistical relation between them and the response variable. Thus, the need to specify the nature of the regression function encountered in ordinary regression analysis does not arise in analysis of variance models.

Illustrations

Figure 16.1 illustrates the essential differences between regression and analysis of variance models for the case where the predictor variable is quantitative. Shown in Figure 16.1a is the regression model for a pricing study involving three different price levels, $X = \$50$, $\$60$, $\$70$. Note that the XY plane has been rotated from its usual position so that the Y axis faces the viewer. For each level of the predictor variable, there is a probability distribution of sales volumes. The means of these probability distributions fall on the regression curve, which describes the statistical relation between price and mean sales volume.

The analysis of variance model for the same study is illustrated in Figure 16.1b. The three price levels are treated as separate populations, each leading to a probability distribution of sales volumes. The quantitative differences in the three price levels and their statistical relation to expected sales volume are not considered by the analysis of variance model.

FIGURE 16.1 Relation between Regression and Analysis of Variance Models.

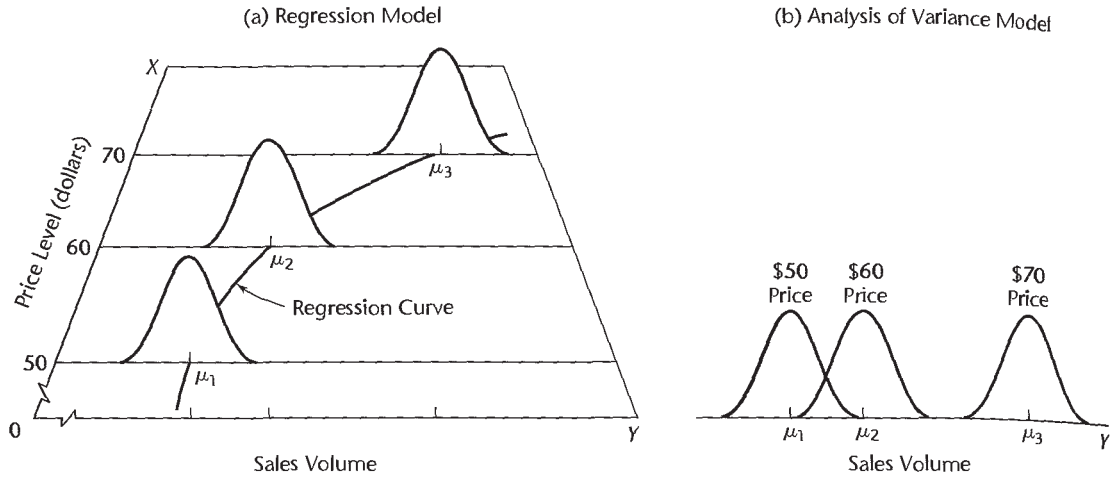


FIGURE 16.2 Analysis of Variance Model Representation —Incentive Pay Example.

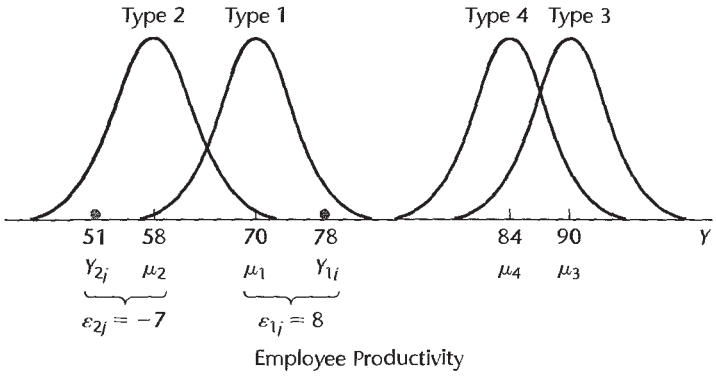


Figure 16.2 illustrates the analysis of variance model for a study of the effects of four different types of incentive pay systems on employee productivity. Here, each type of incentive pay system corresponds to a different population, and there is associated with each a probability distribution of employee productivities (Y). Since type of incentive pay system is a qualitative variable, Figure 16.2 does not contain a corresponding regression model representation.

Choice between Two Types of Models

As we have seen in Chapter 8, regression analysis can handle qualitative predictor variables by means of indicator variables. When indicator variables are so used with regression models, the regression results will be identical to those obtained with analysis of variance models. The reason why analysis of variance exists as a distinct statistical methodology is that the structure of the predictor indicator variables permits computational simplifications that are explicitly recognized in the statistical procedures for the analysis of variance.

Hence, there is no fundamental choice between regression and analysis of variance models when the predictor variables are qualitative.

On the other hand, there is a choice in modeling when the predictor variables are quantitative. One possibility is to recognize the quantitative nature of the predictor variables explicitly; this can only be done by a regression model. The other possibility is to set up classes for each quantitative variable and then employ either indicator variables in a regression model or an analysis of variance model. As we mentioned in Chapter 8, the strategy of setting up classes for quantitative variables is sometimes followed in large-scale studies as a means of obtaining a nonparametric regression fit when there is substantial doubt about the nature of the statistical relation. Here again, analysis of variance models and regression models with indicator variables will lead to identical results.

3 Single-Factor ANOVA Model

sic Ideas

The basic elements of the ANOVA model for a single-factor study are quite simple. Corresponding to each factor level, there is a probability distribution of responses. For example, in a study of the effects of four types of incentive pay on employee productivity, there is a probability distribution of employee productivities for each type of incentive pay. The ANOVA model assumes that:

1. Each probability distribution is normal.
2. Each probability distribution has the same variance.
3. The responses for each factor level are random selections from the corresponding probability distribution and are independent of the responses for any other factor level.

Figure 16.2 illustrates these conditions. Note the normality of the probability distributions and the constant variability. The probability distributions differ only with respect to their means. Differences in the means therefore reflect the essential factor level effects, and it is for this reason that the analysis of variance focuses on the mean responses for the different factor levels.

The analysis of the sample data from the factor level probability distributions usually proceeds in two steps:

1. Determine whether or not the factor level means are the same.
2. If the factor level means differ, examine how they differ and what the implications of the differences are.

In this chapter, we consider step 1, the testing procedure for determining whether or not the factor level means are the same. In the next chapter, we take up the analysis of the factor level means when the means differ.

Cell Means Model

Before stating the ANOVA model for single-factor studies, we need to develop some notation. We shall denote by r the number of levels of the factor under study (e.g., $r = 4$ types of incentive pay), and we shall denote any one of these levels by the index i ($i = 1, \ldots, r$). The number of cases for the ith factor level is denoted by n_i, and the total number of cases

in the study is denoted by n_T, where:

$$n_T = \sum_{i=1}^{r} n_i \tag{16.1}$$

This notation differs from that used earlier for regression models, where the subscript i identifies the case or trial.

For analysis of variance models we shall always use the last subscript to represent the case or trial for a given factor level or treatment. Here, the index j will be used to identify the given case or trial for a particular factor level. We shall let Y_{ij} denote the value of the response variable in the jth trial for the ith factor level. For instance, Y_{ij} is the productivity of the jth employee in the ith incentive plan, or the sales volume of the jth store featuring the ith type of shelf display. Since the number of cases or trials for the ith factor level is denoted by n_i, we have $j = 1, \ldots, n_i$.

The ANOVA model can now be stated as follows:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{16.2}$$

where:

Y_{ij} is the value of the response variable in the jth trial for the ith factor level or treatment

μ_i are parameters

ε_{ij} are independent $N(0, \sigma^2)$

$i = 1, \ldots, r;\ j = 1, \ldots, n_i$

This model is called the *cell means model* for reasons to be explained shortly. This model may be used for data from observational studies or for data from experimental studies based on a completely randomized design.

Important Features of Model

1. The observed value of Y in the jth trial for the ith factor level or treatment is the sum of two components: (a) a constant term μ_i and (b) a random error term ε_{ij}.

2. Since $E\{\varepsilon_{ij}\} = 0$, it follows that:

$$E\{Y_{ij}\} = \mu_i \tag{16.3}$$

Thus, all responses or observations Y_{ij} for the ith factor level have the same expectation μ_i, and this parameter is the mean response for the ith factor level or treatment.

3. Since μ_i is a constant, it follows from (A.16a) that:

$$\sigma^2\{Y_{ij}\} = \sigma^2\{\varepsilon_{ij}\} = \sigma^2 \tag{16.4}$$

Thus, all observations have the same variance, regardless of factor level.

4. Since each ε_{ij} is normally distributed, so is each Y_{ij}. This follows from (A.36) because Y_{ij} is a linear function of ε_{ij}.

5. The error terms are assumed to be independent. Hence, the error term for the outcome on any one trial has no effect on the error term for the outcome of any other trial for the

same factor level or for a different factor level. Since the ε_{ij} are independent, so are the responses Y_{ij}.

6. In view of these features, ANOVA model (16.2) can be restated as follows:

$$Y_{ij} \text{ are independent } N(\mu_i, \sigma^2) \tag{16.5}$$

mple

Suppose that ANOVA model (16.2) is applicable to the earlier incentive pay study illustration and that the parameters are as follows:

$$\mu_1 = 70 \qquad \mu_2 = 58 \qquad \mu_3 = 90 \qquad \mu_4 = 84 \qquad \sigma = 4$$

Figure 16.2 contains a representation of this model. Note that employee productivities for incentive pay type 1 according to this model are normally distributed with mean $\mu_1 = 70$ and standard deviation $\sigma = 4$.

Suppose that in the jth trial of incentive pay type 1, the observed productivity is $Y_{1j} = 78$. In that case, the error term value is $\varepsilon_{1j} = 8$, for we have:

$$\varepsilon_{1j} = Y_{1j} - \mu_1 = 78 - 70 = 8$$

Figure 16.2 shows this observation Y_{1j}. Note that the deviation of Y_{1j} from the mean μ_1 represents the error term ε_{1j}. This figure also shows the observation $Y_{2j} = 51$, for which the error term value is $\varepsilon_{2j} = -7$.

The ANOVA Model Is a Linear Model

ANOVA model (16.2) is a linear model because it can be expressed in matrix terms in the form (6.19), i.e., as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. We illustrate this for a study involving $r = 3$ treatments, and for which $n_1 = n_2 = n_3 = 2$. $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$ are then defined as follows here:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{bmatrix} \tag{16.6}$$

Note the simple structure of the $\mathbf{X}$ matrix and that the $\boldsymbol{\beta}$ vector consists of the means μ_i.

To see that these matrices yield ANOVA model (16.2), recall from (6.20) that the vector of expected values $E\{Y_{ij}\}$ is given by $\mathbf{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}$. We thus obtain:

$$\mathbf{E}\{\mathbf{Y}\} = \begin{bmatrix} E\{Y_{11}\} \\ E\{Y_{12}\} \\ E\{Y_{21}\} \\ E\{Y_{22}\} \\ E\{Y_{31}\} \\ E\{Y_{32}\} \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \end{bmatrix} \tag{16.7}$$

This indicates properly that $E\{Y_{ij}\} = \mu_i$. Hence, ANOVA model (16.2)—$Y_{ij} = \mu_i + \varepsilon_{ij}$—in matrix form is given by $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} = \mathbf{X\beta} + \boldsymbol{\varepsilon} = \begin{bmatrix} \mu_1 \\ \mu_1 \\ \mu_2 \\ \mu_2 \\ \mu_3 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{bmatrix} \tag{16.8}$$

Since the error terms in the model have the same structure as those in general linear regression model (6.19)—namely, independence and constant variance—the variance-covariance matrix of the error terms in the ANOVA model is the same as in (6.19):

$$\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\} = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2\mathbf{I} \tag{16.9}$$

In addition, like for general linear regression model (6.19), the variance-covariance matrix of the Y responses is the same as that of the error terms:

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \sigma^2\mathbf{I} \tag{16.10}$$

When ANOVA model (16.2) is expressed as a linear model, as in (16.8), it can be seen why it is called the cell means model, because the $\boldsymbol{\beta}$ vector contains the means of the "cells"—here factor levels. In Section 16.7 we discuss an equivalent ANOVA model called the factor effects model, where the $\boldsymbol{\beta}$ vector contains components of the factor level means.

Interpretation of Factor Level Means

Observational Data. In an observational study, the factor level means μ_i correspond to the means for the different factor level populations. For instance, in a study of the productivity of employees in each of three shifts operated in a plant, the populations consist of the employee productivities for each of the three shifts. The population mean μ_1 is the mean productivity for employees in shift 1, and μ_2 and μ_3 are interpreted similarly. The variance σ^2 refers to the variability of employee productivities within a shift.

Experimental Data. In an experimental study, the factor level mean μ_i stands for the mean response that would be obtained if the ith treatment were applied to all units in the population of experimental units about which inferences are to be drawn. Similarly, the variance σ^2 refers to the variability of responses if any given experimental treatment were applied to the entire population of experimental units. For instance, in a completely randomized design to study the effects of three different training programs on employee productivity, in which 90 employees participate, a third of these employees is assigned at random to each of the three programs. The mean μ_1 here denotes the mean productivity if training program 1 were given to each employee in the population of experimental units; the means μ_2 and μ_3 are interpreted correspondingly. The variance σ^2 denotes the variability in productivities if any one training program were given to each employee in the population of experimental units.

Distinction between ANOVA Models I and II

We shall consider two single-factor analysis of variance models. For brevity, we shall refer to these as ANOVA models I and II. ANOVA model I, which was stated in (16.2), applies to such cases as a comparison of five different advertisements or a comparison of four different rust inhibitors, where the conclusions pertain to just those factor levels included in the study. ANOVA model II, to be discussed in Chapter 25, applies to a different type of situation, namely, where the conclusions extend to a population of factor levels of which the levels in the study are a sample. Consider, for instance, a company that owns several hundred retail stores throughout the country. Seven of these stores are selected at random, and a sample of employees from each store is then chosen and asked in a confidential interview for an evaluation of the management of the store. The seven stores in the study constitute the seven levels of the factor under study, namely, retail store. In this case, however, management is not just interested in the seven stores included in the study but wishes to generalize the study results to all of the retail stores it owns. Another example when ANOVA model II is applicable is when three machines out of 75 in a plant are selected at random and their daily output is studied for a period of 10 days. The three machines constitute the three factor levels in this study, but interest is not just in the three machines in the study but in all machines in the plant.

Thus, the essential difference between situations where ANOVA models I and II are applicable is that model I is relevant when the factor levels are chosen because of intrinsic interest in them (e.g., five different advertisements) and they are not considered to be a sample from a larger population. ANOVA model II is appropriate when the factor levels constitute a sample from a larger population (e.g., three machines out of 75) and interest is in this larger population. Thus, ANOVA model I is also referred as the *fixed effects* model, and ANOVA model II is called the *random effects* model. In this and the next two chapters, we focus on ANOVA model I. For brevity, we omit the word "fixed" or "model I" and simply refer to the model as the ANOVA model.

Comment

The ANOVA model (16.2) for single-factor studies, like any other statistical model, is not likely to be met exactly by any real-world situation. However, it will be met approximately in many cases. As we shall note later, the statistical procedures based on ANOVA model (16.2) are quite robust, so that even if the actual conditions differ substantially from those of the model, the statistical analysis may still be an appropriate approximation. ■

16.4 Fitting of ANOVA Model

The parameters of ANOVA model (16.2) are ordinarily unknown and must be estimated from sample data. As with normal error regression models, the method of least squares and the method of maximum likelihood lead to the same estimators of the model parameters μ_i in normal error ANOVA model (16.2). Before turning to these estimators, we shall describe an example to be used in this chapter and the next, and we shall develop needed additional notation.

Example The Kenton Food Company wished to test four different package designs for a new breakfast cereal. Twenty stores, with approximately equal sales volumes, were selected as the experimental units. Each store was randomly assigned one of the package designs, with each

TABLE 16.1 **Number of Cases Sold by Stores for Each of Four Package Designs—Kenton Food Company Example.**

Package Design	Store (j) 1	2	3	4	5	Total	Mean	Number of Stores
i	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	Y_{i5}	$Y_{i.}$	$\bar{Y}_{i.}$	n_i
1	11	17	16	14	15	73	14.6	5
2	12	10	15	19	11	67	13.4	5
3	23	20	18	17		78	19.5	4
4	27	33	22	26	28	136	27.2	5
All designs						$Y_{..} = 354$	$\bar{Y}_{..} = 18.63$	19

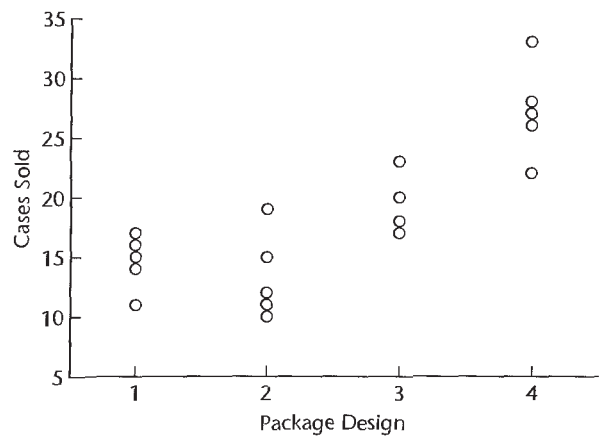


FIGURE 16.3 **JMP Scatter Plot of Number of Cases Sold by Package Design—Kenton Food Company Example.**

package design assigned to five stores. A fire occurred in one store during the study period, so this store had to be dropped from the study. Hence, one of the designs was tested in only four stores. The stores were chosen to be comparable in location and sales volume. Other relevant conditions that could affect sales, such as price, amount and location of shelf space, and special promotional efforts, were kept the same for all of the stores in the experiment. Sales, in number of cases, were observed for the study period, and the results are recorded in Table 16.1. This study is a completely randomized design with package design as the single, four-level factor.

Figure 16.3 contains a JMP scatter plot of the number of cases sold versus package design number. We readily see that designs 3 and 4 led to the largest sales, and that designs 1 and 2 led to smaller sales. We also see that the variability in store sales appears to be about the same for the four designs, consistent with ANOVA model (16.2). To make more formal inferences, we first need to develop some additional notation.

Notation

As explained earlier, Y_{ij} represents the observation or response for the jth sample unit for the ith factor level. For the Kenton Food Company example, Y_{ij} denotes the number of cases sold by the jth store assigned to the ith package design. For instance, Y_{11} represents the sales of the first store assigned package design 1. For our example, $Y_{11} = 11$ cases. Similarly, sales of the second store assigned package design 3 are $Y_{32} = 20$ cases.

The total of the observations for the ith factor level is denoted by $Y_{i\cdot}$:

$$Y_{i\cdot} = \sum_{j=1}^{n_i} Y_{ij} \tag{16.11}$$

Note that the dot in $Y_{i\cdot}$ indicates an aggregation over the j index; in our example, the aggregation is over all stores assigned to the ith package design. For instance, the total sales for all stores assigned package design 1 are, according to Table 16.1, $Y_{1\cdot} = 73$ cases. Similarly, total sales for all stores assigned package design 4 are $Y_{4\cdot} = 136$ cases.

The sample mean for the ith factor level is denoted by $\bar{Y}_{i\cdot}$:

$$\bar{Y}_{i\cdot} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} = \frac{Y_{i\cdot}}{n_i} \tag{16.12}$$

In our example, the mean number of cases sold by stores assigned package design 1 is $\bar{Y}_{1\cdot} = 73/5 = 14.6$. Note that the dot in the subscript $\bar{Y}_{1\cdot}$ indicates that the averaging is done over j (stores).

The total of all observations in the study is denoted by $Y_{\cdot\cdot}$:

$$Y_{\cdot\cdot} = \sum_{i=1}^{r} \sum_{j=1}^{n_i} Y_{ij} \tag{16.13}$$

where the two dots indicate aggregation over both the j and i indexes (in our example, over all stores for any one package design and then over all package designs). In our example, the total sales for all stores for all designs are $Y_{\cdot\cdot} = 354$.

Finally, the overall mean for all responses is denoted by $\bar{Y}_{\cdot\cdot}$:

$$\bar{Y}_{\cdot\cdot} = \frac{\sum_i \sum_j Y_{ij}}{n_T} = \frac{Y_{\cdot\cdot}}{n_T} \tag{16.14}$$

The two dots here indicate that the averaging is done over both i and j. For our example, we have from Table 16.1 that $\bar{Y}_{\cdot\cdot} = 354/19 = 18.63$. Note that the overall mean (16.14) can be written as a weighted average of the factor level means in (16.12):

$$\bar{Y}_{\cdot\cdot} = \sum_{i=1}^{r} \frac{n_i}{n_T} \bar{Y}_{i\cdot} \tag{16.14a}$$

Least Squares and Maximum Likelihood Estimators

According to the least squares criterion, the sum of the squared deviations of the observations around their expected values must be minimized with respect to the parameters. For ANOVA model (16.2), we know from (16.3) that the expected value of observation Y_{ij} is $E\{Y_{ij}\} = \mu_i$. Hence, the quantity to be minimized is:

$$Q = \sum_i \sum_j (Y_{ij} - \mu_i)^2 \tag{16.15}$$

Now (16.15) can be written as follows:

$$Q = \sum_j (Y_{1j} - \mu_1)^2 + \sum_j (Y_{2j} - \mu_2)^2 + \cdots + \sum_j (Y_{rj} - \mu_r)^2 \tag{16.15a}$$

Note that each of the parameters appears in only one of the component sums in (16.15a). Hence, Q can be minimized by minimizing each of the component sums separately. It is well known that the sample mean minimizes a sum of squared deviations. Hence, the least squares estimator of μ_i, denoted by $\hat{\mu}_i$, is:

$$\hat{\mu}_i = \bar{Y}_{i\cdot} \tag{16.16}$$

Thus, the *fitted value* for observation Y_{ij}, denoted by $\hat{Y}_{ij}$ for regression models, is simply the corresponding factor level sample mean here:

$$\hat{Y}_{ij} = \bar{Y}_{i\cdot} \tag{16.17}$$

The same estimators are obtained by the method of maximum likelihood. The likelihood function here corresponds to that in (1.26) for the normal error simple linear regression model, except that the regression model expected value $\beta_0 + \beta_1 X_i$ is replaced here by μ_i:

$$L(\mu_1, \ldots, \mu_r, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_i \sum_j (Y_{ij} - \mu_i)^2\right] \tag{16.18}$$

Maximizing this likelihood function with respect to the parameters μ_i is equivalent to minimizing the sum $\sum\sum(Y_{ij} - \mu_i)^2$ in the exponent, which is the least squares criterion in (16.15).

Example

For the Kenton Food Company example, the least squares and maximum likelihood estimates of the model parameters are as follows according to Table 16.1:

Parameter	Estimate
μ_1	$\hat{\mu}_1 = \bar{Y}_{1\cdot} = 14.6$
μ_2	$\hat{\mu}_2 = \bar{Y}_{2\cdot} = 13.4$
μ_3	$\hat{\mu}_3 = \bar{Y}_{3\cdot} = 19.5$
μ_4	$\hat{\mu}_4 = \bar{Y}_{4\cdot} = 27.2$

Thus, the mean sales per store with package design 1 are estimated to be 14.6 cases for the population of stores under study, and the fitted value for each of the observations for package design 1 is $\hat{Y}_{1j} = \bar{Y}_{1\cdot} = 14.6$. Similarly, the mean sales for package design 2 are estimated to be 13.4 cases per store, and the fitted values for each response for this package design is $\hat{Y}_{2j} = \bar{Y}_{2\cdot} = 13.4$.

Comments

1. The least squares and maximum likelihood estimators in (16.16) have all of the desirable properties mentioned in Chapter 1 for the regression estimators. For example, they are minimum variance unbiased estimators.

2. To derive the least squares estimator of μ_i, we need to minimize, with respect to μ_i, the ith component sum of squares in (16.15a):

$$Q_i = \sum_j (Y_{ij} - \mu_i)^2 \tag{16.19}$$

Differentiating with respect to μ_i, we obtain:

$$\frac{dQ_i}{d\mu_i} = \sum_j -2(Y_{ij} - \mu_i)$$

When we set this derivative equal to zero and replace the parameter μ_i by the least squares estimator $\hat{\mu}_i$, we obtain the result in (16.16):

$$-2\sum_{j=1}^{n_i}(Y_{ij} - \hat{\mu}_i) = 0$$

$$\sum_j Y_{ij} = n_i\hat{\mu}_i$$

$$\hat{\mu}_i = \bar{Y}_{i\cdot}$$

■

esiduals

Residuals are highly useful for examining the aptness of ANOVA models. The residual e_{ij} is again defined, as for regression models, as the difference between the observed and fitted values:

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i\cdot} \tag{16.20}$$

Thus, a residual here represents the deviation of an observation from its estimated factor level mean.

An important property of the residuals for ANOVA model (16.2) is that they sum to zero for each factor level i:

$$\sum_j e_{ij} = 0 \qquad i = 1, \ldots, r \tag{16.21}$$

As for regression analysis, residuals for ANOVA models are useful for examining the appropriateness of the ANOVA model. We shall discuss this use of residuals in Chapter 18.

Example

Table 16.2 contains the residuals for the Kenton Food Company example. For instance, from Table 16.1, we find:

$$e_{11} = Y_{11} - \bar{Y}_{1\cdot} = 11 - 14.6 = -3.6$$

$$e_{21} = Y_{21} - \bar{Y}_{2\cdot} = 12 - 13.4 = -1.4$$

Note from Table 16.2 that the residuals sum to zero for each factor level, as expected.

TABLE 16.2 Residuals—Kenton Food Company Example.

Package Design	Store (j)					
i	1	2	3	4	5	Total
1	−3.6	2.4	1.4	−.6	.4	0
2	−1.4	−3.4	1.6	5.6	−2.4	0
3	3.5	.5	−1.5	−2.5		0
4	−.2	5.8	−5.2	−1.2	.8	0
All designs						0

16.5 Analysis of Variance

Just as the analysis of variance for a regression model partitions the total sum of squares into the regression sum of squares and the error sum of squares, so a corresponding partitioning exists for ANOVA model (16.2).

Partitioning of *SSTO*

The total variability of the Y_{ij} observations, not using any information about factor levels, is measured in terms of the total deviation of each observation, i.e., the deviation of Y_{ij} around the overall mean $\bar{Y}_{..}$:

$$Y_{ij} - \bar{Y}_{..} \tag{16.22}$$

When we utilize information about the factor levels, the deviations reflecting the uncertainty remaining in the data are those of each observation Y_{ij} around its respective estimated factor level mean $\bar{Y}_{i.}$:

$$Y_{ij} - \bar{Y}_{i.} \tag{16.23}$$

The difference between the deviations (16.22) and (16.23) reflects the difference between the estimated factor level mean and the overall mean:

$$(Y_{ij} - \bar{Y}_{..}) - (Y_{ij} - \bar{Y}_{i.}) = \bar{Y}_{i.} - \bar{Y}_{..} \tag{16.24}$$

Note from (16.24) that we can decompose the total deviation $Y_{ij} - \bar{Y}_{..}$ into two components:

$$\underbrace{Y_{ij} - \bar{Y}_{..}}_{\substack{\text{Total}\\ \text{deviation}}} = \underbrace{\bar{Y}_{i.} - \bar{Y}_{..}}_{\substack{\text{Deviation of}\\ \text{estimated}\\ \text{factor level}\\ \text{mean around}\\ \text{overall mean}}} + \underbrace{Y_{ij} - \bar{Y}_{i.}}_{\substack{\text{Deviation}\\ \text{around}\\ \text{estimated}\\ \text{factor}\\ \text{level mean}}} \tag{16.25}$$

Thus, the total deviation $Y_{ij} - \bar{Y}_{..}$ can be viewed as the sum of two components:

1. The deviation of the estimated factor level mean around the overall mean.
2. The deviation of Y_{ij} around its estimated factor level mean, which is simply the residual e_{ij} according to (16.20).

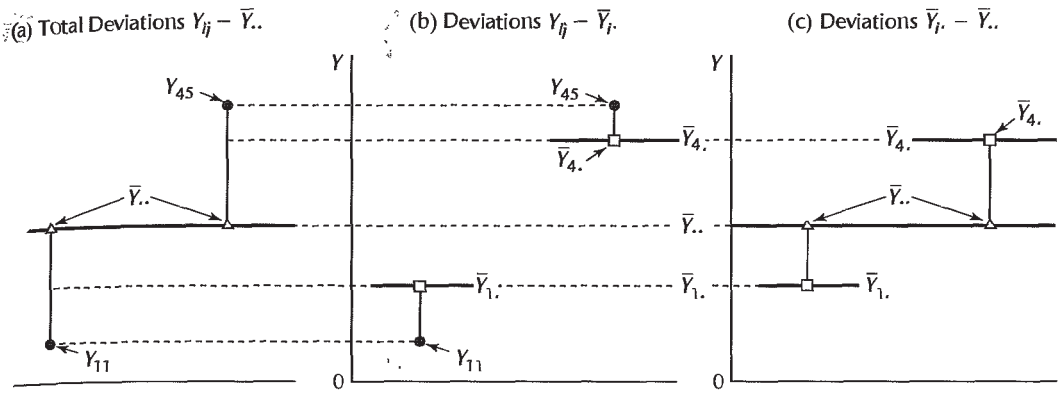
Figure 16.4 illustrates this decomposition for the Kenton Food Company example for two of the observations, Y_{11} and Y_{45}.

When we square both sides in (16.25) and then sum, the cross products on the right drop out and we obtain:

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = \sum_i n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2 \tag{16.26}$$

The term on the left measures the total variability of the Y_{ij} observations and is denoted, as

FIGURE 16.4 Illustration of Partitioning of Total Deviations $Y_{ij} - \bar{Y}_{..}$—Kenton Food Company Example (not drawn to scale; only observations Y_{11} and Y_{45} are shown).



for regression, by *SSTO* for *total sum of squares:*

$$SSTO = \sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 \tag{16.27}$$

The first term on the right in (16.26) will be denoted by *SSTR*, standing for *treatment sum of squares:*

$$SSTR = \sum_i n_i(\bar{Y}_{i.} - \bar{Y}_{..})^2 \tag{16.28}$$

The second term on the right in (16.26) will be denoted by *SSE*, standing for *error sum of squares:*

$$SSE = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2 = \sum_i \sum_j e_{ij}^2 \tag{16.29}$$

Thus, (16.26) can be written equivalently:

$$SSTO = SSTR + SSE \tag{16.30}$$

The correspondence to the regression decomposition in (2.50) is readily apparent.

The total sum of squares for the analysis of variance model is therefore made up of these two components:

1. *SSE:* A measure of the random variation of the observations around the respective estimated factor level means. The less variation among the observations for each factor level, the smaller is *SSE*. If $SSE = 0$, the observations for any given factor level are all the same, and this holds for all factor levels. The more the observations for each factor level differ among themselves, the larger will be *SSE*.

2. *SSTR:* A measure of the extent of differences between the estimated factor level means, based on the deviations of the estimated factor level means $\bar{Y}_{i.}$ around the overall mean $\bar{Y}_{..}$. If all estimated factor level means $\bar{Y}_{i.}$ are the same, then $SSTR = 0$. The more the estimated factor level means differ, the larger will be *SSTR*.

Example

The analysis of variance breakdown of the total sum of squares for the Kenton Food Company example in Table 16.1 is obtained as follows, using (16.27), (16.28), and (16.29):

$$\begin{aligned} SSTO &= (11 - 18.63)^2 + (17 - 18.63)^2 + (16 - 18.63)^2 + \cdots + (28 - 18.63)^2 \\ &= 746.42 \\ SSTR &= 5(14.6 - 18.63)^2 + 5(13.4 - 18.63)^2 + 4(19.5 - 18.63)^2 + 5(27.2 - 18.63)^2 \\ &= 588.22 \\ SSE &= (11 - 14.6)^2 + (17 - 14.6)^2 + (16 - 14.6)^2 + \cdots + (28 - 27.2)^2 \\ &= 158.20 \end{aligned}$$

Thus, the decomposition of *SSTO* is:

$$746.42 = 588.22 + 158.20$$
$$SSTO = SSTR + SSE$$

Note that much of the total variation in the observations is associated with variation between the estimated factor level means.

Comments

1. To prove (16.26), we begin by considering (16.25):

$$Y_{ij} - \bar{Y}_{..} = (\bar{Y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.})$$

Squaring both sides we obtain:

$$(Y_{ij} - \bar{Y}_{..})^2 = (\bar{Y}_{i.} - \bar{Y}_{..})^2 + (Y_{ij} - \bar{Y}_{i.})^2 + 2(\bar{Y}_{i.} - \bar{Y}_{..})(Y_{ij} - \bar{Y}_{i.})$$

When we sum over all sample observations in the study (i.e., over both i and j), we obtain:

$$\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2 = \sum_i \sum_j (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.})^2 + \sum_i \sum_j 2(\bar{Y}_{i.} - \bar{Y}_{..})(Y_{ij} - \bar{Y}_{i.}) \tag{16.31}$$

The first term on the right in (16.31) equals:

$$\sum_i \sum_j (\bar{Y}_{i.} - \bar{Y}_{..})^2 = \sum_i n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 \tag{16.32}$$

since $(\bar{Y}_{i.} - \bar{Y}_{..})^2$ is constant when summed over j; hence, n_i such terms are picked up for the summation over j.

The third term on the right in (16.31) equals zero:

$$\sum_i \sum_j 2(\bar{Y}_{i.} - \bar{Y}_{..})(Y_{ij} - \bar{Y}_{i.}) = 2 \sum_i (\bar{Y}_{i.} - \bar{Y}_{..}) \sum_j (Y_{ij} - \bar{Y}_{i.}) = 0 \tag{16.33}$$

This follows because $\bar{Y}_{i.} - \bar{Y}_{..}$ is constant for the summation over j; hence, it can be brought in front of the summation sign over j. Further, $\sum_j (Y_{ij} - \bar{Y}_{i.}) = 0$ for all i, since the sum of the deviations around the arithmetic mean is always zero.

Thus, (16.31) reduces to (16.26).

2. The squared estimated factor level mean deviations $(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$ in *SSTR* in (16.28) are weighted by the number of cases n_i for that factor level. The reason is that for each observation Y_{ij} at factor level i, the deviation component $\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$ is the same. ■

kdown of Degrees of Freedom

Corresponding to the decomposition of the total sum of squares, we can also obtain a breakdown of the associated degrees of freedom.

SSTO has $n_T - 1$ degrees of freedom associated with it. There are altogether n_T deviations $Y_{ij} - \bar{Y}_{\cdot\cdot}$, but one degree of freedom is lost because the deviations are not independent in that they must sum to zero; i.e., $\sum\sum(Y_{ij} - \bar{Y}_{\cdot\cdot}) = 0$.

SSTR has $r - 1$ degrees of freedom associated with it. There are r estimated factor level mean deviations $\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$, but one degree of freedom is lost because the deviations are not independent in that the weighted sum must equal zero; i.e., $\sum n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}) = 0$.

SSE has $n_T - r$ degrees of freedom associated with it. This can be readily seen by considering the component of *SSE* for the ith factor level:

$$\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2 \tag{16.34}$$

The expression in (16.34) is the equivalent of a total sum of squares considering only the ith factor level. Hence, there are $n_i - 1$ degrees of freedom associated with this sum of squares. Since *SSE* is a sum of component sums of squares such as the one in (16.34), the degrees of freedom associated with *SSE* are the sum of the component degrees of freedom:

$$(n_1 - 1) + (n_2 - 1) + \cdots + (n_r - 1) = n_T - r \tag{16.35}$$

xample For the Kenton Food Company example, for which $n_T = 19$ and $r = 4$, the degrees of freedom associated with the three sums of squares are as follows:

SS	*df*
SSTO	$19 - 1 = 18$
SSTR	$4 - 1 = 3$
SSE	$19 - 4 = 15$

Note that degrees of freedom, like sums of squares, are additive:

$$18 = 3 + 15$$

Mean Squares

The mean squares, as usual, are obtained by dividing each sum of squares by its associated degrees of freedom. We therefore have:

$$MSTR = \frac{SSTR}{r - 1} \tag{16.36a}$$

$$MSE = \frac{SSE}{n_T - r} \tag{16.36b}$$

Here, *MSTR* stands for *treatment mean square* and *MSE*, as before, stands for *error mean square*.

Example

For the Kenton Food Company example, we obtain from earlier results:

$$MSTR = \frac{588.22}{3} = 196.07$$

$$MSE = \frac{158.20}{15} = 10.55$$

Note that the two mean squares do not add to $SSTO/(n_T - 1) = 746.42/18 = 41.47$. Thus, the mean squares here, as in regression, are not additive.

Analysis of Variance Table

The breakdowns of the total sum of squares and degrees of freedom, together with the resulting mean squares, are presented in an ANOVA table such as Table 16.3. The ANOVA table for the Kenton Food Company example is presented in Figure 16.5 which contains the JMP output for single-factor analysis of variance. Note that the output contains the overall mean response ($\bar{Y} = 18.63158$), the number of observations, the ANOVA table, and the estimated factor level means $\bar{Y}_{i\cdot}$. In this table, the line for the treatments source of variation is labeled "Package Design." The results in the JMP output are shown to more decimal places than we have shown, but are consistent with our calculations. Note also that the JMP ANOVA table shows the degrees of freedom column before the sum of squares column. The columns labeled "Std Error," "Lower 95%," and "Upper 95%" will be discussed in Chapter 17.

Expected Mean Squares

The expected values of *MSE* and *MSTR* can be shown to be as follows:

$$E\{MSE\} = \sigma^2 \tag{16.37a}$$

$$E\{MSTR\} = \sigma^2 + \frac{\sum n_i(\mu_i - \mu_\cdot)^2}{r-1} \tag{16.37b}$$

where:

$$\mu_\cdot = \frac{\sum n_i \mu_i}{n_T} \tag{16.37c}$$

is referred to as the weighted mean. These expected values are shown in the $E\{MS\}$ column of Table 16.3.

TABLE 16.3 ANOVA Table for Single-Factor Study.

Source of Variation	*SS*	*df*	*MS*	$E\{MS\}$
Between treatments	$SSTR = \sum n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$r-1$	$MSTR = \frac{SSTR}{r-1}$	$\sigma^2 + \frac{\sum n_i(\mu_i - \mu_\cdot)^2}{r-1}$
Error (within treatments)	$SSE = \sum\sum(Y_{ij} - \bar{Y}_{i\cdot})^2$	$n_T - r$	$MSE = \frac{SSE}{n_T - r}$	σ^2
Total	$SSTO = \sum\sum(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	$n_T - 1$		

GURE 16.5 rtion of JMP tput for gle-Factor alysis of Variance—enton Food mpany xample.

Oneway Anova

Summary of Fit

Rsquare	0.788055
Adj Rsquare	0.745666
Root Mean Square Error	3.247563
Mean of Response	18.63158
Observations (or Sum Wgts)	19

Analysis of Variance

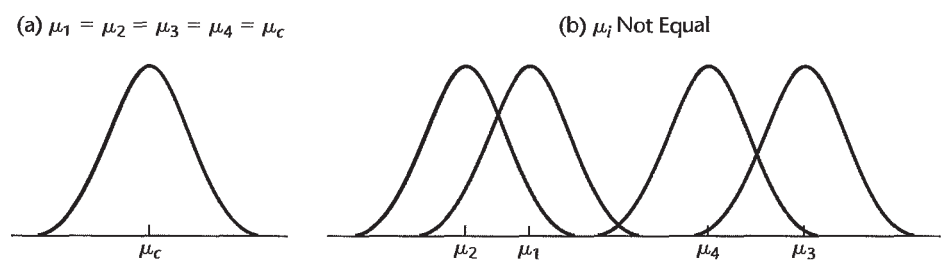
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Package Design	3	588.22105	196.074	18.5911	<.0001
Error	15	158.20000	10.547		
C. Total	18	746.42105			

Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
1	5	14.6000	1.4524	11.504	17.696
2	5	13.4000	1.4524	10.304	16.496
3	4	19.5000	1.6238	16.039	22.961
4	5	27.2000	1.4524	24.104	30.296

Std Error uses a pooled estimate of error variance

FIGURE 16.6 Sampling Distributions of $\bar{Y}_{i\cdot}$ for Four Treatments ($n_i \equiv n$).



Two important features of the expected mean squares deserve attention:

1. *MSE* is an unbiased estimator of σ^2, the variance of the error terms ε_{ij}, whether or not the factor level means μ_i are equal. This is intuitively reasonable since the variability of the observations within each factor level is not affected by the magnitudes of the estimated factor level means for normal populations.

2. When all factor level means μ_i are equal and hence equal to the weighted mean $\mu_{\cdot}$, then $E\{MSTR\} = \sigma^2$ since the second term on the right in (16.37b) becomes zero. Hence, *MSTR* and *MSE* both estimate the error variance σ^2 when all factor level means μ_i are equal. When, however, the factor level means are not equal, *MSTR* tends on the average to be larger than *MSE*, since the second term in (16.37b) will then be positive. This is intuitively reasonable, as illustrated in Figure 16.6 for four treatments. The situation portrayed there assumes that all sample sizes are equal, i.e., $n_i \equiv n$. When all μ_i are equal, then all $\bar{Y}_{i\cdot}$ follow the same sampling distribution, with common mean μ_c and variance σ^2/n; this is portrayed in

Figure 16.6a. When the μ_i are not equal, on the other hand, the $\bar{Y}_{i\cdot}$ follow different sampling distributions, each with the same variability σ^2/n but centered on different means μ_i. One such possibility is shown in Figure 16.6b. Hence, the $\bar{Y}_{i\cdot}$ will tend to differ more from each other when the μ_i differ than when the μ_i are equal, and consequently *MSTR* will tend to be larger when the factor level means are not the same than when they are equal. This property of *MSTR* is utilized in constructing the statistical test discussed in the next section to determine whether or not the factor level means μ_i are the same. If *MSTR* and *MSE* are of the same order of magnitude, this is taken to suggest that the factor level means μ_i are equal. If *MSTR* is substantially larger than *MSE*, this is taken to suggest that the μ_i are not equal.

Comments

1. To find the expected value of *MSE*, we first note that *MSE* can be expressed as follows:

$$MSE = \frac{1}{n_T - r}\sum_i\sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2$$

$$= \frac{1}{n_T - r}\sum_i \left[(n_i - 1)\frac{\sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2}{n_i - 1}\right] \tag{16.38}$$

Now let us denote the ordinary sample variance of the observations for the ith factor level by s_i^2:

$$s_i^2 = \frac{\sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2}{n_i - 1} \tag{16.39}$$

Hence, (16.38) can be expressed as follows:

$$MSE = \frac{1}{n_T - r}\sum_i (n_i - 1)s_i^2 \tag{16.40}$$

Since it is well known that the sample variance (16.39) is an unbiased estimator of the population variance, which in our case is σ^2 for all factor levels, we obtain:

$$E\{MSE\} = \frac{1}{n_T - r}\sum_i (n_i - 1)E\{s_i^2\}$$

$$= \frac{1}{n_T - r}\sum_i (n_i - 1)\sigma^2$$

$$= \sigma^2$$

2. We shall derive the expected value of *MSTR* for the special case when all sample sizes n_i are the same, namely, when $n_i \equiv n$. The general result in (16.37b) becomes for this special case:

$$E\{MSTR\} = \sigma^2 + \frac{n\sum(\mu_i - \mu_\cdot)^2}{r - 1} \qquad \text{when } n_i \equiv n \tag{16.41}$$

Further, when all factor level sample sizes are n, *MSTR* as defined in (16.28) and (16.36a) becomes:

$$MSTR = \frac{n\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}{r - 1} \qquad \text{when } n_i \equiv n \tag{16.42}$$

To derive (16.41), consider the model formulation for Y_{ij} in (16.2):

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

Averaging the Y_{ij} for the ith factor level, we obtain:

$$\bar{Y}_{i\cdot} = \mu_i + \bar{\varepsilon}_{i\cdot} \tag{16.43}$$

where $\bar{\varepsilon}_{i\cdot}$ is the average of the ε_{ij} for the ith factor level:

$$\bar{\varepsilon}_{i\cdot} = \frac{\sum_j \varepsilon_{ij}}{n} \tag{16.44}$$

Averaging the Y_{ij} over all factor levels, we obtain:

$$\bar{Y}_{\cdot\cdot} = \mu_{\cdot} + \bar{\varepsilon}_{\cdot\cdot} \tag{16.45}$$

where $\mu_{\cdot}$, which is defined in (16.37c), becomes for $n_i \equiv n$:

$$\mu_{\cdot} = \frac{n\sum \mu_i}{nr} = \frac{\sum \mu_i}{r} \qquad \text{when } n_i \equiv n \tag{16.46}$$

and $\bar{\varepsilon}_{\cdot\cdot}$ is the average of all ε_{ij}:

$$\bar{\varepsilon}_{\cdot\cdot} = \frac{\sum\sum \varepsilon_{ij}}{nr} \tag{16.47}$$

Since the sample sizes are equal, we also have:

$$\bar{Y}_{\cdot\cdot} = \frac{\sum \bar{Y}_{i\cdot}}{r} \qquad \bar{\varepsilon}_{\cdot\cdot} = \frac{\sum \bar{\varepsilon}_{i\cdot}}{r} \tag{16.48}$$

Using (16.43) and (16.45), we obtain:

$$\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} = (\mu_i + \bar{\varepsilon}_{i\cdot}) - (\mu_{\cdot} + \bar{\varepsilon}_{\cdot\cdot}) = (\mu_i - \mu_{\cdot}) + (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}) \tag{16.49}$$

When we square $\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot}$ and sum over the factor levels, we obtain:

$$\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = \sum(\mu_i - \mu_{\cdot})^2 + \sum(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2 + 2\sum(\mu_i - \mu_{\cdot})(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}) \tag{16.50}$$

We now wish to find $E\{\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\}$, and therefore need to find the expected value of each term on the right in (16.50):

a. Since $\sum(\mu_i - \mu_{\cdot})^2$ is a constant, its expectation is:

$$E\left\{\sum(\mu_i - \mu_{\cdot})^2\right\} = \sum(\mu_i - \mu_{\cdot})^2 \tag{16.51}$$

b. Before finding the expectation of the second term on the right, consider first the expression:

$$\frac{\sum(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2}{r-1}$$

This is an ordinary sample variance, since $\bar{\varepsilon}_{\cdot\cdot}$ is the sample mean of the r terms $\bar{\varepsilon}_{i\cdot}$ per (16.48). We further know that the sample variance is an unbiased estimator of the variance of the variable, in this case the variable being $\bar{\varepsilon}_{i\cdot}$. But $\bar{\varepsilon}_{i\cdot}$ is just the mean of n independent error terms ε_{ij} by (16.44). Hence:

$$\sigma^2\{\bar{\varepsilon}_{i\cdot}\} = \frac{\sigma^2\{\varepsilon_{ij}\}}{n} = \frac{\sigma^2}{n}$$

Therefore:

$$E\left\{\frac{\sum(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2}{r-1}\right\} = \frac{\sigma^2}{n}$$

so that:

$$E\left\{\sum(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2\right\} = \frac{(r-1)\sigma^2}{n} \tag{16.52}$$

c. Since both $\bar{\varepsilon}_{i\cdot}$ and $\bar{\varepsilon}_{\cdot\cdot}$ are means of ε_{ij} terms, all of which have expectation 0, it follows that:

$$E\{\bar{\varepsilon}_{i\cdot}\} = 0 \qquad E\{\bar{\varepsilon}_{\cdot\cdot}\} = 0$$

Hence:

$$E\left\{2\sum(\mu_i - \mu_\cdot)(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})\right\} = 2\sum(\mu_i - \mu_\cdot)E\{\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot}\} = 0 \tag{16.53}$$

We have thus shown, by (16.51), (16.52), and (16.53), that:

$$E\left\{\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2\right\} = \sum(\mu_i - \mu_\cdot)^2 + \frac{(r-1)\sigma^2}{n}$$

But then (16.41) follows at once:

$$E\{MSTR\} = E\left\{\frac{n\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}{r-1}\right\} = \frac{n}{r-1}\left[\sum(\mu_i - \mu_\cdot)^2 + \frac{(r-1)\sigma^2}{n}\right]$$

$$= \sigma^2 + \frac{n\sum(\mu_i - \mu_\cdot)^2}{r-1} \qquad \text{when } n_i \equiv n$$

■

16.6 *F* Test for Equality of Factor Level Means

It is customary to begin the analysis of a single-factor study by determining whether or not the factor level means μ_i are equal. If, for instance, the four package designs in the Kenton Food Company example lead to the same mean sales volumes, there is no need for further analysis, such as to determine which design is best or how two particular designs compare in stimulating sales.

Thus, the alternative conclusions we wish to consider are:

$$\begin{aligned} &H_0\colon \mu_1 = \mu_2 = \cdots = \mu_r \\ &H_a\colon \text{not all } \mu_i \text{ are equal} \end{aligned} \tag{16.54}$$

Test Statistic

The test statistic to be used for choosing between the alternatives in (16.54) is:

$$F^* = \frac{MSTR}{MSE} \tag{16.55}$$

Note that *MSTR* here plays the role corresponding to *MSR* for a regression model.

Large values of F^* support H_a, since *MSTR* will tend to exceed *MSE* when H_a holds, as we saw from (16.37). Values of F^* near 1 support H_0, since both *MSTR* and *MSE* have the same expected value when H_0 holds. Hence, the appropriate test is an upper-tail one.

ution of F^*

When all treatment means μ_i are equal, each response Y_{ij} has the same expected value. In view of the additivity of sums of squares and degrees of freedom, Cochran's theorem (2.61) then implies:

$$\text{When } H_0 \text{ holds, } \frac{SSE}{\sigma^2} \text{ and } \frac{SSTR}{\sigma^2} \text{ are independent } \chi^2 \text{ variables}$$

It follows in the same fashion as for regression:

$$\text{When } H_0 \text{ holds, } F^* \text{ is distributed as } F(r-1, n_T - r)$$

When H_a holds, that is, when the μ_i are not all equal, F^* does *not* follow the F distribution. Rather, it follows a complex distribution called the *noncentral F distribution.* We shall make use of the noncentral F distribution when we discuss the power of the F test in Section 16.10.

Comment

SSTR and *SSE* are independent even if all μ_i are not equal. *SSTR* is solely based on the estimated factor level means $\bar{Y}_{i\cdot}$. On the other hand, *SSE* reflects the variability within the factor level samples, and this within-sample variability is not affected by the magnitudes of the estimated factor level means when the error terms are normally distributed. ■

onstruction of Decision Rule

Usually, the risk of making a Type I error is controlled in constructing the decision rule. This provides protection against making further, more detailed, analyses of the factor effects when in fact there are no differences in the factor level means. The Type II error can also be controlled, as we shall see later in Section 16.10, through sample size determination.

Since we know that F^* is distributed as $F(r-1, n_T - r)$ when H_0 holds and that large values of F^* lead to conclusion H_a, the appropriate decision rule to control the level of significance at α is:

$$\begin{aligned} &\text{If } F^* \leq F(1-\alpha; r-1, n_T - r), \text{ conclude } H_0 \\ &\text{If } F^* > F(1-\alpha; r-1, n_T - r), \text{ conclude } H_a \end{aligned} \tag{16.56}$$

where $F(1-\alpha; r-1, n_T - r)$ is the $(1-\alpha)100$ percentile of the appropriate F distribution.

Example

For the Kenton Food Company example, we wish to test whether or not mean sales are the same for the four package designs:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_a: \text{not all } \mu_i \text{ are equal}$$

Management wishes to control the risk of making a Type I error at $\alpha = .05$. We therefore require $F(.95; 3, 15)$, where the degrees of freedom are those shown in Figure 16.5. From Table B.4 in Appendix B, we find $F(.95; 3, 15) = 3.29$. Hence, the decision rule is:

$$\text{If } F^* \leq 3.29, \text{ conclude } H_0$$
$$\text{If } F^* > 3.29, \text{ conclude } H_a$$

Using the data in the ANOVA table in Figure 16.5, we obtain the test statistic:

$$F^* = \frac{MSTR}{MSE} = \frac{196.07}{10.55} = 18.6$$

Since $F^* = 18.6 > 3.29$, we conclude H_a, that the factor level means μ_i are not equal, or that the four different package designs do not lead to the same mean sales volume. Thus, we conclude that there is a relation between package design and sales volume.

The P-value for the test statistic is the probability $P\{F(3, 15) > F^* = 18.6\}$, which is .00003. This P-value again indicates that the data from the experiment are not consistent with all designs having the same effect on sales volume.

The conclusion of a relation between package design and sales volume did not surprise the sales manager of the Kenton Food Company. The study was conducted in the first place because the sales manager expected the four package designs to have different effects on sales volume and was interested in finding out the nature of these differences. In the next chapter, we discuss the second stage of the analysis, namely, how to study the nature of the factor level means when differences exist.

Comments

1. If there are only two factor levels so that $r = 2$, it can easily be shown that the test employing F^* in (16.55) is the equivalent of the two-population, two-sided t test in Table A.2a. The F test here has $(1, n_T - 2)$ degrees of freedom, and the t test has $n_1 + n_2 - 2$ or $n_T - 2$ degrees of freedom; thus both tests lead to equivalent critical regions. For comparing two population means, the t test generally is to be preferred since it can be used to conduct both two-sided and one-sided tests (Table A.2); the F test can be used only for two-sided tests.

2. Since the F test for testing the alternatives (16.54) is a test of a linear statistical model, it can be obtained by the general linear test approach explained in Section 2.8:

a. The full model is ANOVA model (16.2):

$$Y_{ij} = \mu_i + \varepsilon_{ij} \qquad \text{Full model} \tag{16.57}$$

Fitting the full model by either the method of least squares or the method of maximum likelihood leads to the fitted values $\hat{Y}_{ij} = \bar{Y}_{i\cdot}$, per (16.17), and to the resulting error sum of squares:

$$SSE(F) = \sum\sum(Y_{ij} - \hat{Y}_{ij})^2 = \sum\sum(Y_{ij} - \bar{Y}_{i\cdot})^2$$

$SSE(F)$ has $df_F = n_T - r$ degrees of freedom associated with it because r parameter values $(\mu_1, \ldots, \mu_r)$ have to be estimated.

b. The reduced model under H_0 is:

$$Y_{ij} = \mu_c + \varepsilon_{ij} \qquad \text{Reduced model} \tag{16.58}$$

where μ_c is the common mean for all factor levels. Fitting the reduced model leads to the estimator $\hat{\mu}_c = \bar{Y}_{\cdot\cdot}$ so that all fitted values are $\hat{Y}_{ij} \equiv \bar{Y}_{\cdot\cdot}$, and the resulting error sum of squares is:

$$SSE(R) = \sum\sum(Y_{ij} - \hat{Y}_{ij})^2 = \sum\sum(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$$

The degrees of freedom associated with $SSE(R)$ are $df_R = n_T - 1$ because one parameter (μ_c) had to be estimated.

c. Since, according to (16.27) and (16.29), respectively:

$$SSE(R) = SSTO$$

$$SSE(F) = SSE$$

and since by (16.30) $SSTO - SSE = SSTR$, the general linear test statistic (2.70) becomes here:

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

$$= \frac{SSTO - SSE}{(n_T - 1) - (n_T - r)} \div \frac{SSE}{n_T - r} = \frac{SSTR}{r - 1} \div \frac{SSE}{n_T - r} = \frac{MSTR}{MSE}$$ ■

6.7 Alternative Formulation of Model

actor Effects Model

At times, an alternative but completely equivalent formulation of the single-factor ANOVA model in (16.2) is used. This alternative formulation is called the *factor effects model*. With this alternative formulation, the treatment means μ_i are expressed in an equivalent fashion by means of the identity:

$$\mu_i \equiv \mu_. + (\mu_i - \mu_.) \tag{16.59}$$

where $\mu_.$ is a constant that can be defined to fit the purpose of the study. We shall denote the difference $\mu_i - \mu_.$ by τ_i:

$$\tau_i = \mu_i - \mu_. \tag{16.60}$$

so that (16.59) can be expressed in equivalent fashion as:

$$\mu_i \equiv \mu_. + \tau_i \tag{16.61}$$

The difference $\tau_i = \mu_i - \mu_.$ is called the *ith factor level effect* or the *ith treatment effect*.

The ANOVA model in (16.2) can now be stated equivalently as follows:

$$Y_{ij} = \mu_. + \tau_i + \varepsilon_{ij} \tag{16.62}$$

where:

$\mu_.$ is a constant component common to all observations

τ_i is the effect of the ith factor level (a constant for each factor level)

ε_{ij} are independent $N(0, \sigma^2)$

$i = 1, \ldots, r;\ j = 1, \ldots, n_i$

ANOVA model (16.62) is called a factor effects model because it is expressed in terms of the factor effects τ_i, in distinction to the cell means model (16.2), which is expressed in terms of the cell (treatment) means μ_i.

Factor effects model (16.62) is a linear model, like the equivalent cell means model (16.2). We shall demonstrate this in the next section.

Definition of $\mu_.$

The splitting up of the factor level mean μ_i into two components, an overall constant $\mu_.$ and a factor level or treatment effect τ_i, depends on the definition of $\mu_.$, which can be defined in many ways. We now explain two basic ways to define $\mu_.$.

Unweighted Mean. Often, a definition of $\mu_.$ as the unweighted average of all factor level means μ_i is found to be useful:

$$\mu_. = \frac{\sum_{i=1}^{r} \mu_i}{r} \tag{16.63}$$

This definition implies that:

$$\sum_{i=1}^{r} \tau_i = 0 \tag{16.64}$$

because by (16.60) we have:

$$\sum \tau_i = \sum (\mu_i - \mu_.) = \sum \mu_i - r\mu_.$$

and by (16.63) we have:

$$\sum \mu_i = r\mu_.$$

Thus, the definition of the overall constant $\mu_.$ in (16.63) implies a restriction on the τ_i, in this case that their sum must be zero.

Example

For the earlier incentive pay example in Figure 16.2, we have $\mu_1 = 70$, $\mu_2 = 58$, $\mu_3 = 90$, and $\mu_4 = 84$. When $\mu_.$ is defined according to (16.63), we obtain:

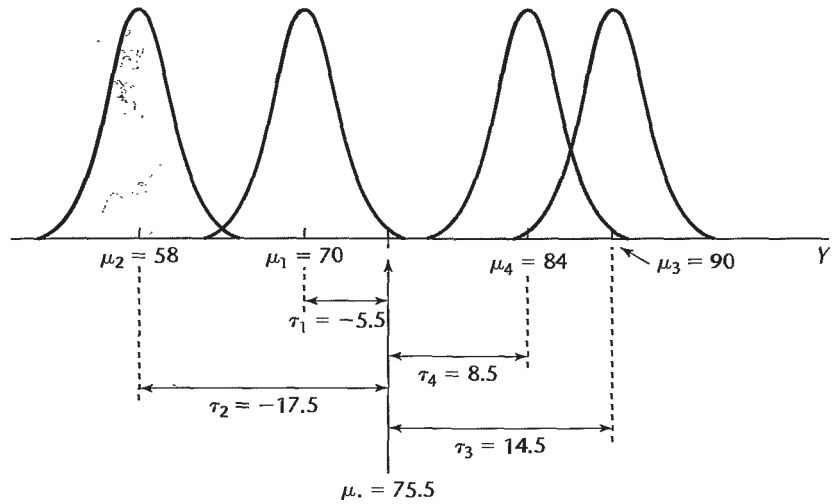
$$\mu_. = \frac{70 + 58 + 90 + 84}{4} = 75.5$$

Hence:

$$\begin{aligned} \tau_1 &= 70 - 75.5 = -5.5 \\ \tau_2 &= 58 - 75.5 = -17.5 \\ \tau_3 &= 90 - 75.5 = 14.5 \\ \tau_4 &= 84 - 75.5 = 8.5 \end{aligned}$$

The first treatment effect $\tau_1 = -5.5$, for instance, indicates that the mean employee productivity for incentive pay type 1 is 5.5 units less than the average productivity for all four types of incentive pay. Figure 16.7 provides an illustration of these treatment effects.

FIGURE 16.7 Illustration of Treatment Effects—Incentive Pay Example.



Weighted Mean The constant $\mu_.$ can also be defined as some weighted average of the factor level means μ_i:

$$\mu_. = \sum_{i=1}^{r} w_i \mu_i \qquad \text{where } \sum_{i=1}^{r} w_i = 1 \tag{16.65}$$

Note that the w_i are weights defined so that their sum is 1. The restriction on the τ_i implied by definition (16.65) is:

$$\sum_{i=1}^{r} w_i \tau_i = 0 \tag{16.66}$$

This follows in the same fashion as (16.64).

The choice of weights w_i should depend on the meaningfulness of the resulting overall mean $\mu_.$. We present now two examples where different weightings are appropriate: (1) weighting according to a known measure of importance and (2) weighting according to sample size.

Example 1 A car rental firm wanted to estimate the average fuel consumption (in miles per gallon) for its large fleet of cars, which consists of 50 percent compacts, 30 percent sedans, and 20 percent station wagons. Here, a meaningful measure of $\mu_.$ might be in terms of overall mean fuel consumption:

$$\mu_. = .5\mu_1 + .3\mu_2 + .2\mu_3 \tag{16.67}$$

where μ_1, μ_2, and μ_3 are the mean fuel consumptions for the three types of cars in the fleet. An estimate of $\mu_.$ here is:

$$\hat{\mu}_. = .5\bar{Y}_{1.} + .3\bar{Y}_{2.} + .2\bar{Y}_{3.} \tag{16.68}$$

Example 2 When exact weights are unknown, the subgroup sample sizes may be useful as weights of relative importance. For instance, the proportions of households in a city with no children, one child, and more than one child are not known. A random sample of n_T households was

selected, which contained n_1 households with no child, n_2 households with one child, and n_3 households with more than one child. For testing whether mean entertainment expenditures are the same for the three types of households, use of the proportions n_1/n_T, n_2/n_T, and n_3/n_T as weights might be meaningful. The resulting definition of the overall entertainment expenditures constant $\mu_.$ would then be:

$$\mu_. = \frac{n_1}{n_T}\mu_1 + \frac{n_2}{n_T}\mu_2 + \frac{n_3}{n_T}\mu_3 \tag{16.69}$$

This quantity would be estimated by $\bar{Y}_{..}$:

$$\hat{\mu}_. = \frac{n_1}{n_T}\bar{Y}_{1.} + \frac{n_2}{n_T}\bar{Y}_{2.} + \frac{n_3}{n_T}\bar{Y}_{3.} = \bar{Y}_{..} \tag{16.70}$$

When all sample sizes are equal, $\mu_.$ as defined in (16.69) reduces to the unweighted mean (16.63).

Test for Equality of Factor Level Means

Since the factor effects model (16.62) is equivalent to the cell means model (16.2), the test for equality of factor level means uses the same test statistic F^* in (16.55). The only difference is in the statement of the alternatives. For the cell means model (16.2), the alternatives are as specified in (16.54):

$$H_0\colon \mu_1 = \mu_2 = \cdots = \mu_r$$
$$H_a\colon \text{not all } \mu_i \text{ are equal}$$

For the factor effects model (16.62), these same alternatives in terms of the factor effects are:

$$\begin{aligned} &H_0\colon \tau_1 = \tau_2 = \cdots = \tau_r = 0 \\ &H_a\colon \text{not all } \tau_i \text{ equal zero} \end{aligned} \tag{16.71}$$

The equivalence of the two forms can be readily established. The equality of the factor level means $\mu_1 = \mu_2 = \cdots = \mu_r$ implies that all τ_i are equal. The equalities of the τ_i follow from (16.61) since the constant term $\mu_.$ is common to all factor level effects τ_i. The equality of the factor level means in turn implies that all $\tau_i = 0$, whether the restriction on the τ_i is of the form in (16.64) or (16.66). In either case, the restriction can be satisfied in only one way given the equality of the τ_i, namely, that $\tau_i \equiv 0$. Thus, it is equivalent to state that all factor level means μ_i are equal or that all factor level effects τ_i equal zero.

16.8 Regression Approach to Single-Factor Analysis of Variance

We noted earlier that cell means model (16.2) is a linear model, and that we can obtain test statistic F^* for testing the equality of the factor level means μ_i by means of the general linear test (2.70). We shall now explain the regression approach to single-factor analysis of variance for three alternative models: (1) the factor effects model with unweighted mean, (2) the factor effects model with weighted mean, and (3) the cell means model. It is important to emphasize that the choice of model affects the definition of the model parameters, and not the outcome of the test for equality of factor level means.

Factor Effects Model with Unweighted Mean

To state ANOVA model (16.62):

$$Y_{ij} = \mu. + \tau_i + \varepsilon_{ij}$$

as a regression model, we need to represent the parameters $\mu., \tau_1, \ldots, \tau_r$ in the model. However, constraint (16.64) for the case of equal weightings:

$$\sum_{i=1}^{r} \tau_i = 0$$

implies that one of the r parameters τ_i is not needed since it can be expressed in terms of the other $r - 1$ parameters. We shall drop the parameter τ_r, which according to constraint (16.64) can be expressed in terms of the other $r - 1$ parameters τ_i as follows:

$$\tau_r = -\tau_1 - \tau_2 - \cdots - \tau_{r-1} \tag{16.72}$$

Thus, we shall use only the parameters $\mu., \tau_1, \ldots, \tau_{r-1}$ for the linear model.

To illustrate how a linear model is developed with this approach, consider a single-factor study with $r = 3$ factor levels when $n_1 = n_2 = n_3 = 2$. The $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$ matrices for this case are as follows:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \mu. \\ \tau_1 \\ \tau_2 \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{bmatrix} \tag{16.73}$$

Note that the vector of expected values, $\mathbf{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta}$, yields the following:

$$\mathbf{E}\{\mathbf{Y}\} = \begin{bmatrix} E\{Y_{11}\} \\ E\{Y_{12}\} \\ E\{Y_{21}\} \\ E\{Y_{22}\} \\ E\{Y_{31}\} \\ E\{Y_{32}\} \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu. \\ \tau_1 \\ \tau_2 \end{bmatrix} = \begin{bmatrix} \mu. + \tau_1 \\ \mu. + \tau_1 \\ \mu. + \tau_2 \\ \mu. + \tau_2 \\ \mu. - \tau_1 - \tau_2 \\ \mu. - \tau_1 - \tau_2 \end{bmatrix} \tag{16.74}$$

Since $\tau_3 = -\tau_1 - \tau_2$ according to (16.72), we see that $E\{Y_{31}\} = E\{Y_{32}\} = \mu. + \tau_3$. Thus, the above $\mathbf{X}$ matrix and $\boldsymbol{\beta}$ vector representation provides in all cases the appropriate expected values:

$$E\{Y_{ij}\} = \mu. + \tau_i$$

The illustration in (16.73) indicates how we need to define in general the multiple regression model so that it is the equivalent of the single-factor ANOVA model (16.62). Note that we require indicator variables that take on values 0, 1, or -1. This coding was discussed in Section 8.1. While this coding is not as simple as a 0, 1 coding, it is desirable

here because it leads to regression coefficients in the $\boldsymbol{\beta}$ vector that are the parameters in the factor effects ANOVA model, i.e., $\mu_\cdot, \tau_1 \ldots, \tau_{r-1}$.

We shall let X_{ij1} denote the value of indicator variable X_1 for the jth case from the ith factor level, X_{ij2} the value of indicator variable X_2 for this same case, and so on, using altogether $r - 1$ indicator variables in the model. The multiple regression model then is as follows:

$$Y_{ij} = \mu_\cdot + \tau_1 X_{ij1} + \tau_2 X_{ij2} + \cdots + \tau_{r-1} X_{ij,r-1} + \varepsilon_{ij} \qquad \text{Full model} \tag{16.75}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if case from factor level 1} \\ -1 & \text{if case from factor level } r \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$X_{ij,r-1} = \begin{cases} 1 & \text{if case from factor level } r-1 \\ -1 & \text{if case from factor level } r \\ 0 & \text{otherwise} \end{cases}$$

Note how the ANOVA model parameters play the role of regression function parameters in (16.75); the intercept term is $\mu_\cdot$, and the regression coefficients are $\tau_1, \tau_2, \ldots, \tau_{r-1}$.

The least squares estimator of $\mu_\cdot$ is the average of the cell sample means:

$$\hat{\mu}_\cdot = \frac{\sum_{i=1}^{r} \bar{Y}_{i\cdot}}{r} \tag{16.75a}$$

Note that this quantity is generally not the same as the overall mean $\bar{Y}_{\cdot\cdot}$ unless the cell sample sizes are equal. Also, the least squares estimator of the ith factor effect is:

$$\hat{\tau}_i = \bar{Y}_{i\cdot} - \hat{\mu}_\cdot \tag{16.75b}$$

To test the equality of the treatment means μ_i by means of the regression approach, we state the alternatives in the equivalent formulation (16.71), noting that τ_r must equal zero when $\tau_1 = \tau_2 = \cdots = \tau_{r-1} = 0$ according to (16.72):

$$\begin{aligned} &H_0\colon \tau_1 = \tau_2 = \cdots = \tau_{r-1} = 0 \\ &H_a\colon \text{not all } \tau_i \text{ equal zero} \end{aligned} \tag{16.76}$$

Note that H_0 states that all regression coefficients in regression model (16.75) are zero, and the reduced model is therefore:

$$Y_{ij} = \mu_\cdot + \varepsilon_{ij} \qquad \text{Reduced model} \tag{16.77}$$

Thus, we employ the usual test statistic (6.39b) for testing whether or not there is a regression relation:

$$F^* = \frac{MSR}{MSE} \tag{16.78}$$

Example

To test the equality of mean sales for the four cereal package designs in the Kenton Food Company example by means of the regression approach, we shall employ the regression

model:

$$Y_{ij} = \mu. + \tau_1 X_{ij1} + \tau_2 X_{ij2} + \tau_3 X_{ij3} + \varepsilon_{ij} \tag{16.79}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if case from factor level 1} \\ -1 & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij2} = \begin{cases} 1 & \text{if case from factor level 2} \\ -1 & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij3} = \begin{cases} 1 & \text{if case from factor level 3} \\ -1 & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

A portion of the data in Table 16.1 is repeated in Table 16.4a, together with the coding of the indicator variables X_1, X_2, and X_3. For observation Y_{11}, for instance, note that $X_1 = 1$, $X_2 = 0$, and $X_3 = 0$; hence, we obtain from (16.79):

$$E\{Y_{11}\} = \mu. + \tau_1$$

TABLE 16.4 Regression Approach to the Analysis of Variance—Kenton Food Company Example.

(a) Data for Regression Model (16.79)

i	j	Y_{ij}	X_{ij1}	X_{ij2}	X_{ij3}
1	1	11	1	0	0
1	2	17	1	0	0
1	3	16	1	0	0
1	4	14	1	0	0
1	5	15	1	0	0
2	1	12	0	1	0
...	...	...	...	...	...
4	4	26	−1	−1	−1
4	5	28	−1	−1	−1

(b) Fitted Regression Function

$$\hat{Y} = 18.675 - 4.075X_1 - 5.275X_2 + .825X_3$$

(c) Regression Analysis of Variance Table

Source of Variation	SS	df	MS
Regression	$SSR = 588.22$	3	$MSR = 196.07$
Error	$SSE = 158.20$	15	$MSE = 10.55$
Total	$SSTO = 746.42$	18	

Similarly, for observation Y_{45} we have $X_1 = -1$, $X_2 = -1$, and $X_3 = -1$; hence:

$$E\{Y_{45}\} = \mu. - \tau_1 - \tau_2 - \tau_3 = \mu. + \tau_4$$

since $\tau_4 = -\tau_1 - \tau_2 - \tau_3$.

Note that we employ the following codings in the indicator variables for cases from each of the four factor levels:

	Coding		
Factor Level	X_1	X_2	X_3
1	1	0	0
2	0	1	0
3	0	0	1
4	−1	−1	−1

A computer run of the multiple regression of Y on X_1, X_2, and X_3 yielded the fitted regression function and analysis of variance table presented in Tables 16.4b and 16.4c. Test statistic (16.78) therefore is:

$$F^* = \frac{MSR}{MSE} = \frac{196.07}{10.55} = 18.6$$

This is the same test statistic obtained earlier based on the analysis of variance calculations. Indeed, the analysis of variance table in Table 16.4c obtained with the regression approach is the same as the one in Figure 16.5 obtained with the analysis of variance approach except that the treatment sum of squares and mean square are called the regression sum of squares and mean square in Table 16.4c. From this point on, the test procedure based on the regression approach parallels the analysis of variance test procedure explained earlier.

Note that in the fitted regression function in Table 16.4b, the intercept term $\hat{\mu}. = 18.675$ is the unweighted average of the estimated factor level means $\bar{Y}_{i\cdot}$, not the overall mean $\bar{Y}_{\cdot\cdot}$, because $\mu.$ was defined as the unweighted average of the factor level means μ_i. The regression coefficient $b_1 = \hat{\tau}_1 = \bar{Y}_{i\cdot} - \hat{\mu}. = 14.6 - 18.675 = -4.075$ is simply the difference between the estimated mean in the first cell and the unweighted overall mean. b_2 and b_3 represent similar differences between the estimated factor level mean and the overall unweighted mean.

Comment

The regression approach is not utilized generally for ordinary analysis of variance problems. The reason is that the **X** matrix for analysis of variance problems usually is of a very simple structure, as we have seen earlier. This simple structure permits computational simplifications that are explicitly recognized in the statistical procedures for analysis of variance. We take up the regression approach to analysis of variance here, and in later chapters, for two principal reasons. First, we see that analysis of variance models are encompassed by the general linear statistical model (6.19). Second, the regression approach is very useful for analyzing some multifactor studies when the structure of the **X** matrix is not simple.

actor Effects Model with Weighted Mean

When the factor effects model (16.62) is used with a weighted mean, a modification of the coding scheme in (16.75) is required. The new coding scheme leads to changes in the definitions of the regression coefficients. We describe the new coding scheme and summarize the changes in the context of the proportional sample size weights, $w_i = n_i/n_T$.

When the constant $\mu.$ is the weighted average of the factor level means using proportional sample size weights, we have, from (16.65):

$$\mu. = \sum_{i=1}^{r} w_i \mu_i = \sum_{i=1}^{r} \frac{n_i}{n_T} \mu_i \tag{16.80a}$$

From (16.66), the restriction on the τ_i is:

$$\sum_{i=1}^{r} \frac{n_i}{n_T} \tau_i = 0$$

Solving for τ_r, we find:

$$\tau_r = -\frac{n_1}{n_r}\tau_1 - \frac{n_2}{n_r}\tau_2 - \cdots - \frac{n_{r-1}}{n_r}\tau_{r-1} \tag{16.80b}$$

This leads to the weighted model:

$$Y_{ij} = \mu. + \tau_1 X_{ij1} + \tau_2 X_{ij2} + \cdots + \tau_{r-1} X_{ij,r-1} + \varepsilon_{ij} \qquad \text{Full model} \tag{16.81}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if case from factor level 1} \\ -\dfrac{n_1}{n_r} & \text{if case from factor level } r \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$X_{ij,r-1} = \begin{cases} 1 & \text{if case from factor level } r-1 \\ -\dfrac{n_{r-1}}{n_r} & \text{if case from factor level } r \\ 0 & \text{otherwise} \end{cases}$$

Note that if all cell sample sizes are equal, the mean $\mu.$ is the unweighted mean, and the coding scheme above is the same as the unweighted coding scheme used in (16.75), since $-n_i/n_r = -1$ for $i = 1, \ldots, r-1$.

When the sample sizes are not all equal, as noted in (16.70), the least squares estimate of the weighted mean $\mu.$ is the overall mean $\bar{Y}_{..}$, and the least squares estimate of the ith factor effect τ_i is $\bar{Y}_{i.} - \bar{Y}_{..}$.

Example

In the Kenton Food Company example, weighted mean model (16.81) is:

$$Y_{ij} = \mu. + \tau_1 X_{ij1} + \tau_2 X_{ij2} + \tau_3 X_{ij3} + \varepsilon_{ij} \tag{16.82}$$

where:

$$X_{ij1} = \begin{cases} 1 & \text{if case from factor level 1} \\ -\frac{5}{5} & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij2} = \begin{cases} 1 & \text{if case from factor level 2} \\ -\frac{5}{5} & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

$$X_{ij3} = \begin{cases} 1 & \text{if case from factor level 3} \\ -\frac{4}{5} & \text{if case from factor level 4} \\ 0 & \text{otherwise} \end{cases}$$

The fitted regression function is:

$$\hat{Y} = 18.63 - 4.03X_1 - 5.23X_2 + .87X_3$$

and the following relations hold:

$$\begin{aligned} \hat{\mu}_. &= b_0 = \bar{Y}_{..} = 18.63 \\ \hat{\tau}_1 &= b_1 = \bar{Y}_{1.} - \bar{Y}_{..} = 14.6 - 18.63 = -4.03 \\ \hat{\tau}_2 &= b_2 = \bar{Y}_{2.} - \bar{Y}_{..} = 13.4 - 18.63 = -5.23 \\ \hat{\tau}_3 &= b_3 = \bar{Y}_{3.} - \bar{Y}_{..} = 19.5 - 18.63 = .87 \\ \hat{\tau}_4 &= -\frac{n_1}{n_4}\hat{\tau}_1 - \frac{n_2}{n_4}\hat{\tau}_2 - \frac{n_3}{n_4}\hat{\tau}_3 = 8.56. \end{aligned}$$

A general linear test of the alternatives:

$$\begin{aligned} &H_0\colon \tau_1 = \tau_2 = \tau_3 = 0 \\ &H_a\colon \text{not all } \tau_i = 0 \end{aligned}$$

is conducted using the full model in (16.82) and forming the reduced model by setting $\tau_1 = \tau_2 = \tau_3 = 0$ in full model (16.82). The test statistic (16.78) for the presence of a regression relation again yields:

$$F^* = \frac{MSR}{MSE} = \frac{196.07}{10.55} = 18.6$$

As expected, the results are identical to those obtained earlier for the ANOVA F test.

Cell Means Model

When the analysis of variance test is to be conducted by means of the regression approach based on the cell means model (16.2):

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

the $\boldsymbol{\beta}$ vector can be defined to contain all r treatment means μ_i:

$$\boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_r \end{bmatrix} \tag{16.83}$$

and r indicator variables $X_1, X_2, \ldots, X_r$ are utilized, each defined as a 0, 1 variable as illustrated in Chapter 8:

$$\begin{aligned} X_1 &= \begin{cases} 1 & \text{if case from factor level 1} \\ 0 & \text{otherwise} \end{cases} \\ &\vdots \\ X_r &= \begin{cases} 1 & \text{if case from factor level r} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{16.84}$$

The regression model therefore is:

$$Y_{ij} = \mu_1 X_{ij1} + \mu_2 X_{ij2} + \cdots + \mu_r X_{ijr} + \varepsilon_{ij} \qquad \text{Full model} \tag{16.85}$$

with the μ_i playing the role of regression coefficients.

The $\mathbf{X}$ matrix with this approach contains only 0 and 1 entries. For example, for $r = 3$ factor levels with $n_1 = n_2 = n_3 = 2$ cases, the $\mathbf{X}$ matrix (observations in order Y_{11}, Y_{12}, Y_{21}, etc.) and $\boldsymbol{\beta}$ vector would be as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}$$

Note that regression model (16.85) has no intercept term. When a computer regression package is to be employed for this case, it is important that a fit with no intercept term be specified.

The ANOVA table obtained with regression model (16.85) is different from the one with the single-factor ANOVA model in (16.2) because the regression model (16.85) has no intercept term. Thus, the F test obtained with the regression model cannot be used to test the equality of factor level means. The test of whether the factor level means are equal, i.e., $\mu_1 = \mu_2 = \cdots = \mu_r$, asks only whether or not the regression coefficients in (16.83) are equal, not whether or not they equal zero. Hence, we need to fit the full model and then the reduced model to conduct this test. The reduced model when H_0: $\mu_1 = \cdots = \mu_r$ holds is:

$$Y_{ij} = \mu_c + \varepsilon_{ij} \qquad \text{Reduced model} \tag{16.86}$$

where μ_c is the common value of all μ_i under H_0. The $\mathbf{X}$ matrix here consists simply of a column of 1s. The $\mathbf{X}$ matrix and $\boldsymbol{\beta}$ vector for the reduced model in our example

would be:

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \boldsymbol{\beta} = [\mu_c]$$

After the full and reduced models are fitted and the error sums of squares are obtained for each fit, the usual general linear test statistic (2.70) is then calculated.

Example

For the Kenton Food Company example, the regression fit for the cell means model in (16.85) is:

$$\hat{Y} = 14.6X_1 + 13.4X_2 + 19.5X_3 + 27.2X_4$$

It can be readily seen that the coefficient of X_i is equal to the estimated factor level mean $\bar{Y}_i$. for $i = 1, \ldots, 4$.

A general linear test of the alternatives:

$$H_0\colon \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_a\colon \text{not all } \mu_i \text{ are equal}$$

is conducted using the full and reduced models in (16.85) and (16.86). Here we again find that $SSE(R) = 746.42$ and that $SSE(F) = 158.2$. From (2.70) we have:

$$F^* = \frac{746.42 - 158.2}{4 - 1} \div \frac{158.2}{19 - 4} = 18.6$$

This demonstrates that the test for equality of means using the regression approach is, as expected, the same as that obtained earlier for the ANOVA F test.

16.9 Randomization Tests

Randomization can provide the basis for making inferences without requiring assumptions about the distribution of the error terms ε. Consider factor effects model (16.62) for a single-factor study:

$$Y_{ij} = \mu_. + \tau_i + \varepsilon_{ij}$$

Rather than assume that the ε_{ij} are independent normal random variables with mean zero and constant variance σ^2, we shall now consider each ε_{ij} to be a fixed effect associated with the experimental unit. In this framework, we view the n_T experimental units to be a finite population, and associated with each unit is the unit-specific effect ε_{ij}. When randomization assigns this experimental unit to treatment i, the observed response will be $Y_{ij} = \mu_. + \tau_i + \varepsilon_{ij}$. The response Y_{ij} is still a random variable, but under the randomization view the randomness arises because the treatment effect τ_i is the result of a random assignment of the experimental unit to treatment i.

If there are no treatment effects, that is, if all $\tau_i = 0$, then the response $Y_{ij} = \mu_. + \varepsilon_{ij}$ depends only on the experimental unit. Since with randomization the experimental unit is

equally likely to be assigned to any treatment, the observed response Y_{ij}, if there are no treatment effects, could with equal likelihood have been observed for any of the treatments. Thus, when there are no treatment effects, randomization will lead to an assignment of the finite population of n_T observations Y_{ij} to the treatments such that all treatment combinations of observations are equally likely. This, in turn, leads to an exact sampling distribution of the test statistic under H_0: $\tau_i \equiv 0$, sometimes termed the *randomization distribution* of the test statistic. Percentiles of the randomization distribution can then be used to test for the presence of factor effects. This use of the randomization distribution provides the basis of a nonparametric test for treatment effects.

To illustrate the concept of a randomization distribution, consider a single-factor experiment consisting of two treatments and two replications. In this experiment, the alternatives of interest are:

$$H_0: \tau_1 = \tau_2 = 0$$

$$H_a: \text{not both } \tau_1 \text{ and } \tau_2 \text{ equal zero}$$

Test statistic F^* in (16.55) will be used to conduct the test. The sample results are:

Treatment 1 Y_{1j}	Treatment 2 Y_{2j}
3	8
7	10

For these data, $F^* = 3.20$.

Since the treatments are assigned to experimental units at random, it would have been just as likely, if there are no treatment effects, to have observed 3 and 8 for treatment 1 and 7 and 10 for treatment 2. In that event, the test statistic would have been $F^* = 1.06$. In fact, any division of the four observations into two groups of size two is equally likely with randomization if there are no treatment effects. Because this experiment is small, we can easily list all $4!/(2!2!) = 6$ possible outcomes of the experiment, assuming no treatment effects are present:

Randomization	Treatment 1	Treatment 2	F^*	Probability
1	3, 7	8, 10	3.20	1/6
2	3, 8	7, 10	1.06	1/6
3	3, 10	8, 7	.08	1/6
4	8, 7	3, 10	.08	1/6
5	7, 10	3, 8	1.06	1/6
6	8, 10	3, 7	3.20	1/6

The last two columns give the randomization distribution of test statistic F^* under H_0. Randomization assures us that, when H_0 is true, each possible value of the test statistic has probability 1/6. From the randomization distribution, we see that the P-value for the test

is the probability:

$$P\{F^* \geq 3.20\} = \frac{2}{6} = .33$$

This P-value is somewhat different than the usual (normal theory) P-value:

$$P\{F(1, 2) \geq 3.20\} = .22$$

In this instance, because the sample sizes are very small, the F distribution does not provide a particularly good approximation to the exact sampling distribution of F^* under H_0. However, both empirical and theoretical studies have shown that the F distribution is a good approximation to the exact randomization distribution when the sample sizes are not small. Thus, randomization alone can justify the F test as a good approximate test, without requiring any assumption of independent, normal error terms. We shall next demonstrate the use of the randomization test in a more realistic setting.

Comments

1. Because of the discreteness of the randomization distribution, it is conservative to define the P-value as the probability of equaling or exceeding the observed value of the test statistic when H_0 holds. For continuous sampling distributions, it does not matter whether the P-value is defined as the probability of exceeding the observed value of the test statistic or as the probability of equaling or exceeding it. For instance, $P\{F(1, 2) > 3.20\} = P\{F(1, 2) \geq 3.20\}$. When more than one treatment combination yields the value of the test statistic F^*, some authors suggest that the P-value be calculated as $P\{F > F^*\} + P\{F = F^*\}/2$. This leads to a less conservative P-value.

2. The randomization test is sometimes referred to as a *permutation* test, although permutation tests are also applied to nonrandomized studies. Because of the conservativeness of permutation (or randomization) tests for small samples, their virtues continue to be debated in the literature. See Reference 16.1. ■

Example

A manufacturer of children's plastic toys considered the introduction of statistical process control (SPC) and engineering process control (EPC) in order to reduce the volume of scrap and rework at each of its nine manufacturing plants. To assess the effects of these quality practices, a single-factor experiment was conducted for a six-month period. The treatments were:

Treatment i	Quality Practice
1	None (control group)
2	SPC
3	Both SPC and EPC

The three treatments were each randomly assigned to three of the nine available plants. The response of interest was the reduction in the defect rate at the end of the six-month trial period. The results are given in the first row (randomization 1) in Table 16.5. Management wishes to test whether or not the mean reduction in the defect rate is the same for the three

TABLE 16.5 Randomization Samples and Test Statistics—Quality Control Example.

Randomization	Treatment 1	Treatment 2	Treatment 3	F^*	Probability
1	1.1, .5, −2.1	4.2, 3.7, .8	3.2, 2.8, 6.3	4.39	1/1,680
2	1.1, .5, −2.1	4.2, 3.7, 3.2	.8, 2.8, 6.3	3.74	1/1,680
3	1.1, .5, −2.1	4.2, 3.7, 2.8	3.2, .8, 6.3	3.67	1/1,680
...	...	...	...	...	...
1,680	3.2, 2.8, 6.3	4.2, 3.7, .8	1.1, .5, −2.1	4.39	1/1,680

treatments:

$$H_0\text{: } \tau_1 = \tau_2 = \tau_3 = 0$$
$$H_a\text{: not all } \tau_i \text{ equal zero}$$

The risk of a Type I error is to be controlled at $\alpha = .10$. We shall now conduct this test by obtaining the exact randomization distribution.

In this experimental study, there are $9!/(3!3!3!) = 1{,}680$ possible combinations of assigning the nine experimental units to the three treatments. A computer program was utilized to enumerate these 1,680 combinations and to calculate the F^* statistic for each. A partial listing of results is presented in Table 16.5.

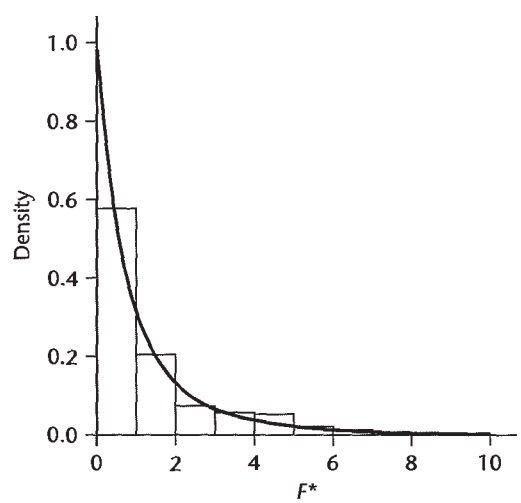
Of the 1,680 possible values of the test statistic F^*, 120 were equal to or greater than the observed value 4.39. Thus, from the randomization distribution we find:

$$P\text{-value} = P\{F^* \geq 4.39\} = \frac{120}{1{,}680} = .071$$

Since $.071 < \alpha = .10$, we conclude that the mean reduction in the defect rate is not the same for the three treatments.

Even though the sample sizes are not very large here, the exact randomization distribution is well approximated by the F distribution. Figure 16.8 shows both the randomization

FIGURE 16.8 Randomization Distribution of F^* and Corresponding F Distribution—Quality Control Example.



distribution in the form of a histogram and the density function for the corresponding F distribution, $F(2, 6)$. Note how well the F distribution approximates the randomization distribution. The P-value according to the F distribution is $P\{F(2, 6) \geq 4.39\} = .067$. This is very close to the randomization P-value of .071.

16.10 Planning of Sample Sizes with Power Approach

For analysis of variance studies, as for other statistical studies, it is important to plan the sample sizes so that needed protection against both Type I and Type II errors can be obtained, or so that the estimates of interest have sufficient precision to be useful. This planning is necessary for both observational and experimental studies to ensure that the sample sizes are large enough to detect important differences with high probability. At the same time, the sample sizes should not be so large that the cost of the study becomes excessive and that unimportant differences become statistically significant with high probability. Planning of sample sizes is therefore an integral part of the design of a study.

We shall generally assume in our discussion of planning sample sizes that all treatments are to have equal sample sizes, reflecting that they are about equally important. Indeed, when major interest lies in pairwise comparisons of all treatment means, it can be shown that equal sample sizes maximize the precision of the comparisons. Another reason for equal sample sizes is that certain departures from the assumed ANOVA model are less troublesome if all factor levels have the same sample size, as noted earlier.

There will be times, however, when unequal sample sizes are appropriate. For instance, when four experimental treatments are each to be compared to a control, it may be reasonable to make the sample size for the control larger. We shall comment later on the planning of sample sizes for such a case.

Planning of sample sizes can be approached in terms of (1) controlling the risks of making Type I and Type II errors, (2) controlling the widths of desired confidence intervals, or (3) a combination of these two. The procedures for planning sample sizes that we shall discuss here are applicable to both observational studies and to experimental studies based on a completely randomized single-factor design. In later chapters, we shall consider the planning of sample sizes for other study designs. In this section, we consider planning of sample sizes with the power approach, which permits controlling the risks of making Type I and Type II errors. In Section 16.11 we discuss planning of sample sizes when the best treatment is to be identified. Later, in Section 17.8, we take up planning of sample sizes to control the precision of estimates of important effects. We shall consider planning of sample sizes for multifactor studies in Section 24.7.

Before we can discuss planning of sample sizes with the power approach, we need to consider the power of the F test.

Power of F Test

By the power of the F test for a single-factor study, we refer to the probability that the decision rule will lead to conclusion H_a, that the treatment means differ, when in fact H_a holds. Specifically, the power is given by the following expression for the cell means model (16.2):

$$\text{Power} = P\{F^* > F(1-\alpha; r-1, n_T - r) \mid \phi\} \tag{16.87}$$

where ϕ is the *noncentrality parameter*, that is, a measure of how unequal the treatment means μ_i are:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{\sum n_i(\mu_i - \mu_.)^2}{r}} \tag{16.87a}$$

and:

$$\mu_. = \frac{\sum n_i \mu_i}{n_T} \tag{16.87b}$$

When all factor level samples are of equal size n, the parameter ϕ becomes:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{n}{r}\sum(\mu_i - \mu_.)^2} \qquad \text{when } n_i \equiv n \tag{16.88}$$

where:

$$\mu_. = \frac{\sum \mu_i}{r} \tag{16.88a}$$

Power probabilities are determined by utilizing the noncentral F distribution since this is the sampling distribution of F^* when H_a holds. The resulting calculations are quite complex. We present a series of tables in Appendix Table B.11 that can be used readily to look up power probabilities directly. The proper table to use depends on the number of factor levels and the level of significance employed in the decision rule. Specifically, Table B.11 is used as follows:

1. Each page refers to a different ν_1, the number of degrees of freedom for the numerator of F^*. For ANOVA model (16.2), $\nu_1 = r - 1$, or the number of factor levels minus one. Table B.11 contains power tables for $\nu_1 = 2, 3, 4, 5$, and 6, as shown at the top of each page.
2. Two levels of significance, denoted by α, are presented in Table B.11, namely, $\alpha = .05$ and $\alpha = .01$. The upper table on each page refers to $\alpha = .05$ and the lower table to $\alpha = .01$.
3. Within each table, the rows refer to different values of ν_2, the degrees of freedom for the denominator of F^*. The columns refer to different values of ϕ, the noncentrality parameter defined in (16.87a). For ANOVA model (16.2), $\nu_2 = n_T - r$.

Examples

1. Consider the case where $\nu_1 = 2$, $\nu_2 = 10$, $\phi = 3$, and $\alpha = .05$. We then find from Table B.11 (p. 1337) that the power is $1 - \beta = .98$.

2. Suppose that for the Kenton Food Company example, the analyst wishes to determine the power of the decision rule in the example on page 699 when there are substantial differences between the factor level means. Specifically, the analyst wishes to consider the case when $\mu_1 = 12.5$, $\mu_2 = 13$, $\mu_3 = 18$, and $\mu_4 = 21$. The weighted mean in (16.87b) therefore is:

$$\mu_. = \frac{5(12.5) + 5(13) + 4(18) + 5(21)}{19} = 16.03$$

Thus, the specified value of ϕ is:

$$\phi = \frac{1}{\sigma}\left[\frac{5(-3.53)^2 + 5(-3.03)^2 + 4(1.97)^2 + 5(4.97)^2}{4}\right]^{1/2}$$

$$= \frac{1}{\sigma}(7.86)$$

Note that we still need to know σ, the standard deviation of the error terms ε_{ij} in the model. Suppose that from past experience it is known that $\sigma = 3.5$ cases approximately. Then we have:

$$\phi = \frac{1}{3.5}(7.86) = 2.25$$

Further, we have for this example:

$$\nu_1 = r - 1 = 3 \qquad \nu_2 = n_T - r = 15 \qquad \alpha = .05$$

Table B.11 on page 1338 indicates that the power is $1 - \beta = .91$. In other words, there are 91 chances in 100 that the decision rule, based on the sample sizes employed, will lead to the detection of differences in the mean sales volumes for the four package designs when the differences are the ones specified earlier.

Comments

1. Any given value of ϕ encompasses many different combinations of factor level means μ_i. Thus, in the Kenton Food Company example, the means $\mu_1 = 12.5$, $\mu_2 = 13$, $\mu_3 = 18$, $\mu_4 = 21$ and the means $\mu_1 = 21$, $\mu_2 = 12.5$, $\mu_3 = 18$, $\mu_4 = 13$ lead to the same value of $\phi = 2.25$ and hence to the same power.

2. The larger ϕ—that is, the larger the differences between the factor level means—the greater the power and hence the smaller the probability of making a Type II error for a given risk α of making a Type I error. Also, the smaller the specified α risk, the smaller is the power for any given ϕ, and hence the larger the risk of a Type II error.

3. Since many single-factor studies are undertaken because of the expectation that the factor level means differ and it is desired to investigate these differences, the α risk used in constructing the decision rule for determining whether or not the factor level means are equal is often set relatively high (e.g., .05 or .10 instead of .01) so as to increase the power of the test.

4. The power table for $\nu_1 = 1$ is not reproduced in Table B.11 since this case corresponds to the comparison of two population means. As noted previously, the F test is the equivalent of the two-sided t test for this case, and the power tables for the two-sided t test presented in Table B.5 can then be used, with noncentrality parameter:

$$\delta = \frac{|\mu_1 - \mu_2|}{\sigma\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \tag{16.89}$$

and degrees of freedom $n_1 + n_2 - 2$. ■

Use of Table B.12 for Single-Factor Studies

The power approach in planning sample sizes can be implemented by use of the power tables for F tests presented in Table B.11. A trial-and-error process is required, however, with these tables. Instead, we shall use other tables that furnish the appropriate sample sizes directly. Table B.12 presents sample size determinations that are applicable when all treatments are to have equal sample sizes and all effects are fixed.

The planning of sample sizes for single-factor studies with fixed factor levels using Table B.12 is done in terms of the noncentrality parameter (16.88) for equal sample sizes. However, instead of requiring a direct specification of the levels of μ_i for which it is important to control the risk of making a Type II error, Table B.12 only requires a specification

of the minimum range of factor level means for which it is important to detect differences between the μ_i with high probability. This minimum range is denoted by Δ:

$$\Delta = \max(\mu_i) - \min(\mu_i) \tag{16.90}$$

The following three specifications need to be made in using Table B.12:

1. The level α at which the risk of making a Type I error is to be controlled.
2. The magnitude of the minimum range Δ of the μ_i which is important to detect with high probability. The magnitude of σ, the standard deviation of the probability distributions of Y, must also be specified since entry into Table B.12 is in terms of the ratio:

$$\frac{\Delta}{\sigma} \tag{16.91}$$

3. The level β at which the risk of making a Type II error is to be controlled for the specification given in 2. Entry into Table B.12 is in terms of the power $1 - \beta$.

When using Table B.12, four α levels are available at which the risk of making a Type I error can be controlled ($\alpha = .2, .1, .05, .01$). The Type II error risk can be controlled at one of four β levels ($\beta = .3, .2, .1, .05$) through the specification of the power $1 - \beta$. Table B.12 provides necessary sample sizes for studies consisting of $r = 2, \ldots, 10$ factor levels or treatments.

Example

A company owning a large fleet of trucks wishes to determine whether or not four different brands of snow tires have the same mean tread life (in thousands of miles). It is important to conclude that the four brands of snow tires have different mean tread lives when the difference between the means of the best and worst brands is 3 (thousand miles) or more. Thus, the minimum range specification is $\Delta = 3$. It is known from past experience that the standard deviation of the tread lives of these tires is $\sigma = 2$ (thousand miles), approximately. Management would like to control the risks of making incorrect decisions at the following levels:

$$\alpha = .05$$

$$\beta = .10 \qquad \text{or} \qquad \text{Power} = 1 - \beta = .90$$

Entering Table B.12 for $\Delta/\sigma = 3/2 = 1.5$, $\alpha = .05$, $1 - \beta = .90$, and $r = 4$, we find $n = 14$. Hence, 14 snow tires of each brand need to be tested in order to control the risks of making incorrect decisions at the desired levels.

Specification of Δ/σ Directly. Table B.12 can also be used when the minimum range is specified directly in units of the standard deviation σ. Let the specification of Δ in this case be $k\sigma$ so that we have by (16.91):

$$\frac{\Delta}{\sigma} = \frac{k\sigma}{\sigma} = k$$

Hence, Table B.12 is entered directly for the specified value k with this approach.

Example

Suppose it is specified in the snow tires example that it is important to detect differences between the mean tread lives if the range of the mean tread lives is $k = 2$ standard deviations

or more. Suppose also that the other specifications are:

$$\alpha = .10$$
$$\beta = .05 \qquad \text{or} \qquad \text{Power} = 1 - \beta = .95$$

From Table B.12, we find for $k = 2$ and $r = 4$ that $n = 9$ tires will need to be tested for each brand in order that the specified risk protection will be achieved.

Comment

While specifying Δ/σ directly does not require an advance planning value of the standard deviation σ, this is not of as much advantage as it might seem because a meaningful specification of Δ in units of σ will frequently require knowledge of the approximate magnitude of the standard deviation. ■

Some Further Observations on Use of Table B.12

1. The exact specification of Δ/σ has great effect on the sample sizes n when Δ/σ is small, but it has much less effect when Δ/σ is large. For instance, when $r = 3$, $\alpha = .05$, and $\beta = .10$, we have from Table B.12:

Δ/σ	n
1.0	27
1.5	13
2.0	8
2.5	6

Thus, unless Δ/σ is quite small, one need not be too concerned about some imprecision in specifying Δ/σ.

2. Reducing either the specified α or β risks or both increases the required sample sizes. For instance, when $r = 4$, $\alpha = .10$, and $\Delta/\sigma = 1.25$, we have:

β	$1 - \beta$	n
.20	.80	13
.10	.90	16
.05	.95	20

3. A moderate error in the advance planning value of σ can cause a substantial miscalculation of required sample sizes. For instance, when $r = 5$, $\alpha = .05$, $\beta = .10$, and $\Delta = 3$, we have:

σ	Δ/σ	n
1	3.0	5
2	1.5	15
3	1.0	32

In view of the usual approximate nature of the advance planning value of σ, it is generally desirable to investigate the needed sample sizes for a range of likely values of σ before deciding on the sample sizes to be employed.

4. Table B.12 is based on the noncentrality parameter ϕ in (16.88) even though no specification is made of the individual factor level means μ_i for which it is important to conclude that the factor level means differ. To see how Table B.12 utilizes the noncentrality parameter ϕ, consider again the snow tires example where $r = 4$ brands are to be tested and a minimum range of $\Delta = 3$ (thousand miles) of the four mean tread lives μ_i is to be detected with high probability. The following are some possible sets of values of the μ_i, each of which has range $\Delta = 3$:

Case	μ_1	μ_2	μ_3	μ_4	$\sum(\mu_i - \mu_.)^2$
1	24	27	25	26	5.00
2	25	25	26	23	4.75
3	25	25	25	28	6.75
4	25	25	26.5	23.5	4.50

The term $\sum(\mu_i - \mu_.)^2$ of the noncentrality parameter ϕ in (16.88) differs for each of these four possibilities and hence the power differs, even though the range is the same in all cases. Note that the term $\sum(\mu_i - \mu_.)^2$ is the smallest for case 4, where two factor level means are at $\mu_.$ and the other two are equally spaced around $\mu_.$. It can be shown that for a given range Δ, the term $\sum(\mu_i - \mu_.)^2$ is minimized when all but two factor level means are at $\mu_.$ and the two remaining factor level means are equally spaced around $\mu_.$. Thus, we have:

$$\min \sum_{i=1}^{r}(\mu_i - \mu_.)^2 = \left(\frac{\Delta}{2}\right)^2 + \left(-\frac{\Delta}{2}\right)^2 + 0 + \cdots + 0 = \frac{\Delta^2}{2} \qquad \textbf{(16.92)}$$

Since the power of the test varies directly with $\sum(\mu_i - \mu_.)^2$, use of (16.92) in calculating Table B.12 ensures that the power is at least $1 - \beta$ for any combination of μ_i values with range Δ.

16.11 Planning of Sample Sizes to Find "Best" Treatment

There are occasions when the chief purpose of the study is to ascertain the treatment with the highest or lowest mean. In the snow tires example, for instance, it may be desired to determine which of the four brands has the longest mean tread life.

Table B.13, developed by Bechhofer, enables us to determine the necessary sample sizes so that with probability $1-\alpha$ the highest (lowest) estimated treatment mean is from the treatment with the highest (lowest) population mean. We need to specify the probability $1-\alpha$, the standard deviation σ, and the smallest difference λ between the highest (lowest) and second highest (second lowest) treatment means that it is important to recognize. Table B.13 assumes that equal sample sizes are to be used for all r treatments.

Example

Suppose that in the snow tires example, the chief objective is to identify the brand with the longest mean tread life. There are $r = 4$ brands. We anticipate, as before, that $\sigma = 2$ (thousand

miles). Further, we are informed that a difference $\lambda = 1$ (thousand miles) between the highest and second highest brand means is important to recognize, and that the probability is to be $1 - \alpha = .90$ or greater that we identify correctly the brand with the highest mean tread life when $\lambda \geq 1$.

The entry in Table B.13 is $\lambda\sqrt{n}/\sigma$. For $r = 4$ and probability $1 - \alpha = .90$, we find from Table B.13 that $\lambda\sqrt{n}/\sigma = 2.4516$. Hence, since the λ specification is $\lambda = 1$, we obtain:

$$\frac{(1)\sqrt{n}}{2} = 2.4516$$

$$\sqrt{n} = 4.9032 \qquad \text{or} \qquad n = 25$$

Thus, when the mean tread life for the best brand exceeds that of the second best by at least 1 (thousand miles) and when $\sigma = 2$ (thousand miles), sample sizes of 25 tires for each brand provide an assurance of at least .90 that the brand with the highest estimated mean $\bar{Y}_{i\cdot}$ is the brand with the highest population mean.

Comment

If the planning value for the standard deviation is not accurate, the probability of identifying the population with the highest (lowest) mean correctly is, of course, affected. This is no different from the other approaches, where a misjudgment of the standard deviation affects the risks of making a Type II error. ■

Cited Reference

16.1. Berger, V. W. "Pros and Cons of Permutation Tests in Clinical Trials," *Statistics in Medicine* 19 (2000), pp. 1319–1328.

Problems

16.1. Refer to Figure 16.1a. Could you determine the mean sales level when the price level is \$68 if you knew the true regression function? Could you make this determination from Figure 16.1b if you only knew the values of the parameters μ_1, μ_2, and μ_3 of ANOVA model (16.2)? What distinction between regression models and ANOVA models is demonstrated by your answers?

16.2. A market researcher, having collected data on breakfast cereal expenditures by families with 1, 2, 3, 4, and 5 children living at home, plans to use an ordinary regression model to estimate the mean expenditures at each of these five family size levels. However, the researcher is undecided between fitting a linear or a quadratic regression model, and the data do not give clear evidence in favor of one model or the other. A colleague suggests: "For your purposes you might simply use an ANOVA model." Is this a useful suggestion? Explain.

16.3. In a study of intentions to get flu-vaccine shots in an area threatened by an epidemic, 90 persons were classified into three groups of 30 according to the degree of risk of getting flu. Each group was together when the persons were asked about the likelihood of getting the shots, on a probability scale ranging from 0 to 1.0. Unavoidably, most persons overheard the answers of nearby respondents. An analyst wishes to test whether the mean intent scores are the same for the three risk groups. Consider each assumption for ANOVA model (16.2) and explain whether this assumption is likely to hold in the present situation.

16.4. A company, studying the relation between job satisfaction and length of service of employees, classified employees into three length-of-service groups (less than 5 years, 5–10 years, more than 10 years). Suppose $\mu_1 = 65$, $\mu_2 = 80$, $\mu_3 = 95$, and $\sigma = 3$, and that ANOVA model (16.2) is applicable.

a. Draw a representation of this model in the format of Figure 16.2.

b. Find $E\{MSTR\}$ and $E\{MSE\}$ if 25 employees from each group are selected at random for intensive interviewing about job satisfaction. Is $E\{MSTR\}$ substantially larger than $E\{MSE\}$ here? What is the implication of this?

16.5. In a study of length of hospital stay (in number of days) of persons in four income groups, the parameters are as follows: $\mu_1 = 5.1$, $\mu_2 = 6.3$, $\mu_3 = 7.9$, $\mu_4 = 9.5$, $\sigma = 2.8$. Assume that ANOVA model (16.2) is appropriate.

a. Draw a representation of this model in the format of Figure 16.2.

b. Suppose 100 persons from each income group are randomly selected for the study. Find $E\{MSTR\}$ and $E\{MSE\}$. Is $E\{MSTR\}$ substantially larger than $E\{MSE\}$ here? What is the implication of this?

c. If $\mu_2 = 5.6$ and $\mu_3 = 9.0$, everything else remaining the same, what would $E\{MSTR\}$ be? Why is $E\{MSTR\}$ substantially larger here than in part (b) even though the range of the factor level means is the same?

16.6. A student asks: "Why is the F test for equality of factor level means not a two-tail test since any differences among the factor level means can occur in either direction?" Explain, utilizing the expressions for the expected mean squares in (16.37).

*16.7. **Productivity improvement.** An economist compiled data on productivity improvements last year for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditures for research and development in the past three years (low, moderate, high). The results of the study follow (productivity improvement is measured on a scale from 0 to 100). Assume that ANOVA model (16.2) is appropriate.

		j											
	i	1	2	3	4	5	6	7	8	9	10	11	12
1	Low	7.6	8.2	6.8	5.8	6.9	6.6	6.3	7.7	6.0			
2	Moderate	6.7	8.1	9.4	8.6	7.8	7.7	8.9	7.9	8.3	8.7	7.1	8.4
3	High	8.5	9.7	10.1	7.8	9.6	9.5						

a. Prepare aligned dot plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals. Do they sum to zero in accord with (16.21)?

d. Obtain the analysis of variance table.

e. Test whether or not the mean productivity improvement differs according to the level of research and development expenditures. Control the α risk at .05. State the alternatives, decision rule, and conclusion.

f. What is the P-value of the test in part (e)? How does it support the conclusion reached in part (e)?

g. What appears to be the nature of the relationship between research and development expenditures and productivity improvement?

16.8. **Questionnaire color.** In an experiment to investigate the effect of color of paper (blue, green, orange) on response rates for questionnaires distributed by the "windshield method"

in supermarket parking lots, 15 representative supermarket parking lots were chosen in a metropolitan area and each color was assigned at random to five of the lots. The response rates (in percent) follow. Assume that ANOVA model (16.2) is appropriate.

		j				
i		1	2	3	4	5
1	Blue	28	26	31	27	35
2	Green	34	29	25	31	29
3	Orange	31	25	27	29	28

a. Prepare aligned dot plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals.

d. Obtain the analysis of variance table.

e. Conduct a test to determine whether or not the mean response rates for the three colors differ. Use level of significance $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. When informed of the findings, an executive said: "See? I was right all along. We might as well print the questionnaires on plain white paper, which is cheaper." Does this conclusion follow from the findings of the study? Discuss.

16.9. **Rehabilitation therapy.** A rehabilitation center researcher was interested in examining the relationship between physical fitness prior to surgery of persons undergoing corrective knee surgery and time required in physical therapy until successful rehabilitation. Patient records in the rehabilitation center were examined, and 24 male subjects ranging in age from 18 to 30 years who had undergone similar corrective knee surgery during the past year were selected for the study. The number of days required for successful completion of physical therapy and the prior physical fitness status (below average, average, above average) for each patient follow.

		j									
i		1	2	3	4	5	6	7	8	9	10
1	Below Average	29	42	38	40	43	40	30	42		
2	Average	30	35	39	28	31	31	29	35	29	33
3	Above Average	26	32	21	20	23	22				

Assume that ANOVA model (16.2) is appropriate.

a. Prepare aligned dot plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals. Do they sum to zero in accord with (16.21)?

d. Obtain the analysis of variance table.

e. Test whether or not the mean number of days required for successful rehabilitation is the same for the three fitness groups. Control the α risk at .01. State the alternatives, decision rule, and conclusion.

f. Obtain the P-value for the test in part (e). Explain how the same conclusion reached in part (e) can be obtained by knowing the P-value.

g. What appears to be the nature of the relationship between physical fitness status and duration of required physical therapy?

*16.10. **Cash offers.** A consumer organization studied the effect of age of automobile owner on size of cash offer for a used car by utilizing 12 persons in each of three age groups (young, middle, elderly) who acted as the owner of a used car. A medium price, six-year-old car was selected for the experiment, and the "owners" solicited cash offers for this car from 36 dealers selected at random from the dealers in the region. Randomization was used in assigning the dealers to the "owners." The offers (in hundred dollars) follow. Assume that ANOVA model (16.2) is applicable.

		j											
i		1	2	3	4	5	6	7	8	9	10	11	12
1	Young	23	25	21	22	21	22	20	23	19	22	19	21
2	Middle	28	27	27	29	26	29	27	30	28	27	26	29
3	Elderly	23	20	25	21	22	23	21	20	19	20	22	21

a. Prepare aligned dot plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals.

d. Obtain the analysis of variance table.

e. Conduct the F test for equality of factor level means; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. What appears to be the nature of the relationship between age of owner and mean cash offer?

*16.11. **Filling machines.** A company uses six filling machines of the same make and model to place detergent into cartons that show a label weight of 32 ounces. The production manager has complained that the six machines do not place the same amount of fill into the cartons. A consultant requested that 20 filled cartons be selected randomly from each of the six machines and the content of each carton carefully weighed. The observations (stated for convenience as deviations from 32.00 ounces) follow. Assume that ANOVA model (16.2) is applicable.

	j						
i	1	2	3	...	18	19	20
1	−.14	.20	.07	...	.07	−.01	−.19
2	.46	.11	.12	...	.02	.11	.12
3	.21	.78	.32	...	.50	.20	.61
4	.49	.58	.52	...	.42	.45	.20
5	−.19	.27	.06	...	.14	.35	−.18
6	.05	−.05	.28	...	.35	−.09	.05

a. Prepare aligned box plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals. Do they sum to zero in accord with (16.21)?

d. Obtain the analysis of variance table.

e. Test whether or not the mean fill differs among the six machines; control the α risk at .05. State the alternatives, decision rule, and conclusion. Does your conclusion support the production manager's complaint?

f. What is the P-value of the test in part (e)? Is this value consistent with your conclusion in part (e)? Explain.

g. Based on the box plots obtained in part (a), does the variation between the mean fills for the six machines appear to be large relative to the variability in fills between cartons for any given machine? Explain.

16.12. **Premium distribution.** A soft-drink manufacturer uses five agents (1, 2, 3, 4, 5) to handle premium distributions for its various products. The marketing director desired to study the timeliness with which the premiums are distributed. Twenty transactions for each agent were selected at random, and the time lapse (in days) for handling each transaction was determined. The results follow. Assume that ANOVA model (16.2) is appropriate.

	j						
i	1	2	3	...	18	19	20
1	24	24	29		27	26	25
2	18	20	20	..	26	22	21
3	10	11	8	...	9	11	12
4	15	13	18	...	17	14	16
5	33	22	28	...	26	30	29

a. Prepare aligned box plots of the data. Do the factor level means appear to differ? Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

b. Obtain the fitted values.

c. Obtain the residuals. Do they sum to zero in accord with (16.21)?

d. Obtain the analysis of variance table.

e. Test whether or not the mean time lapse differs for the five agents; use $\alpha = .10$. State the alternatives, decision rule, and conclusion.

f. What is the P-value of the test in part (e)? Explain how the same conclusion as in part (e) can be reached by knowing the P-value.

g. Based on the box plots obtained in part (a), does there appear to be much variation in the mean time lapse for the five agents? Is this variation necessarily the result of differences in the efficiency of operations of the five agents? Discuss.

16.13. Refer to **Questionnaire color** Problem 16.8. Explain how you would make the random assignments of supermarket parking lots to colors in this single-factor study. Make all appropriate randomizations.

16.14. Refer to **Cash offers** Problem 16.10. Explain how you would make the random assignments of dealers to "owners" in this single-factor study. Make all appropriate randomizations.

16.15. Refer to Problem 16.4. What are the values of τ_1, τ_2, and τ_3 if the ANOVA model is expressed in the factor effects formulation (16.62), and $\mu.$ is defined by (16.63)?

16.16. Refer to Problem 16.5. What are the values of τ_i if the ANOVA model is expressed in the factor effects formulation (16.62), and $\mu.$ is defined by (16.63)?

16.17. Refer to **Premium distribution** Problem 16.12. Suppose that 25 percent of all premium distributions are handled by agent 1, 20 percent by agent 2, 20 percent by agent 3, 20 percent by agent 4, and 15 percent by agent 5.

a. Obtain a point estimate of $\mu.$ when the ANOVA model is expressed in the factor effects formulation (16.62) and $\mu.$ is defined by (16.65), with the weights being the proportions of premium distribution handled by each agent.

b. State the alternatives for the test of equality of factor level means in terms of factor effects model (16.62) for the present case. Would this statement be affected if $\mu.$ were defined according to (16.63)? Explain.

*16.18. Refer to **Productivity improvement** Problem 16.7. Regression model (16.75) is to be employed for testing the equality of the factor level means.

a. Set up the $\mathbf{Y}$, $\mathbf{X}$, and $\boldsymbol{\beta}$ matrices.

b. Obtain $\mathbf{X}\boldsymbol{\beta}$. Develop equivalent expressions of the elements of this vector in terms of the cell means μ_i.

c. Obtain the fitted regression function. What is estimated by the intercept term?

d. Obtain the regression analysis of variance table.

e. Conduct the test for equality of factor level means; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

16.19. Refer to **Questionnaire color** Problem 16.8. Regression model (16.75) is to be employed for testing the equality of the factor level means.

a. Set up the $\mathbf{Y}$, $\mathbf{X}$, and $\boldsymbol{\beta}$ matrices.

b. Obtain $\mathbf{X}\boldsymbol{\beta}$. Develop equivalent expressions of the elements of this vector in terms of the cell means μ_i.

c. Obtain the fitted regression function. What is estimated by the intercept term?

d. Obtain the regression analysis of variance table.

e. Conduct the test for equality of factor level means; use $\alpha = .10$. State the alternatives, decision rule, and conclusion.

16.20. Refer to **Rehabilitation therapy** Problem 16.9. Regression model (16.81) is to be employed for testing the equality of the factor level means.

a. Set up the $\mathbf{Y}$, $\mathbf{X}$, and $\boldsymbol{\beta}$ matrices.

b. Obtain $\mathbf{X}\boldsymbol{\beta}$. Develop equivalent expressions of the elements of this vector in terms of the cell means μ_i.

c. Obtain the fitted regression function. What is estimated by the intercept term?

d. Obtain the regression analysis of variance table.

e. Conduct the test for equality of factor level means; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

*16.21. Refer to **Cash offers** Problem 16.10.

a. Fit regression model (16.75) to the data. What is estimated by the intercept term?

b. Obtain the regression analysis of variance table and test whether or not the factor level means are equal; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

16.22. Refer to **Rehabilitation therapy** Problem 16.9.

a. Fit the full regression model (16.85) to the data. Why would a fitted regression model containing an intercept term not be proper here?

b. Fit the reduced model (16.86) to the data.

c. Use test statistic (2.70) for testing the equality of the factor level means; employ level of significance $\alpha = .01$.

16.23. Refer to Example 1 on page 717. Find the power of the test if $\alpha = .01$, everything else remaining unchanged. How does this power compare with that in Example 1?

16.24. Refer to Example 2 on page 717. The analyst is also interested in the power of the test when $\mu_1 = \mu_2 = 13$ and $\mu_3 = \mu_4 = 18$. Assume that $\sigma = 3.5$.

a. Obtain the power of the test if $\alpha = .05$.

b. What would be the power of the test if $\alpha = .01$?

*16.25. Refer to **Productivity improvement** Problem 16.7. Obtain the power of the test in Problem 16.7e if $\mu_1 = 7.0$, $\mu_2 = 8.0$, and $\mu_3 = 9.0$. Assume that $\sigma = .9$.

16.26. Refer to **Rehabilitation therapy** Problem 16.9. Obtain the power of the test in Problem 16.9e if $\mu_1 = 37$, $\mu_2 = 35$, and $\mu_3 = 28$. Assume that $\sigma = 4.5$.

*16.27. Refer to **Cash offers** Problem 16.10. Obtain the power of the test in Problem 16.10e if the mean cash offers are $\mu_1 = 22$, $\mu_2 = 28$, and $\mu_3 = 22$. Assume that $\sigma = 1.6$.

16.28. Why do you think that the approach to planning sample sizes to find the best treatment by means of Table B.13 does not consider the risk of an incorrect identification when the best two treatment means are the same or practically the same?

*16.29. Consider a single-factor study where $r = 5$, $\alpha = .01$, $\beta = .05$, and $\sigma = 10$, and equal treatment sample sizes are desired by means of the approach in Table B.12.

a. What are the required sample sizes if $\Delta = 10, 15, 20, 30$? What generalization is suggested by your results?

b. What are the required sample sizes for the same values of Δ as in part (a) if $\alpha = .05$, all other specifications remaining the same? How do these sample sizes compare with those in part (a)?

16.30. Consider a single-factor study where $r = 6$, $\alpha = .05$, $\beta = .10$, and $\Delta = 50$, and equal treatment sample sizes are desired by means of the approach in Table B.12.

a. What are the required sample sizes if $\sigma = 50, 25, 20$? What generalization is suggested by your results?

b. What are the required sample sizes for the same values of σ as in part (a) if $r = 4$, all other specifications remaining the same? How do these sample sizes compare with those in part (a)?

16.31. Consider a single-factor study where $r = 5$, $1 - \alpha = .95$, and $\sigma = 20$, and equal sample sizes are desired by means of the approach in Table B.13.

a. What are the required sample sizes if $\lambda = 20, 10, 5$? What generalization is suggested by your results?

b. What are the required sample sizes for the same values of λ as in part (a) if $\sigma = 30$, all other specifications remaining the same? How do these sample sizes compare with those in part (a)?

16.32. Refer to **Questionnaire color** Problem 16.8. Suppose that the sample sizes have not yet been determined but it has been decided to sample the same number of supermarket parking lots for each questionnaire color. A reasonable planning value for the error standard deviation is $\sigma = 3.0$.

a. What would be the required sample sizes if: (1) differences in the response rates are to be detected with probability .90 or more when the range of the treatment means is 4.5, and (2) the α risk is to be controlled at .05?

b. If the sample sizes determined in part (a) were employed, what would be the minimum power of the test for treatment mean differences (using $\alpha = .05$) when the range of the treatment means is 6.0?

c. Suppose the chief objective is to identify the color with the highest mean response rate. The probability should be at least .99 that the best color is recognized correctly when the difference between the response rates for the best and second best colors is 1.5 percent points or more. What are the required sample sizes?

16.33. Refer to **Rehabilitation therapy** Problem 16.9. Suppose that the sample sizes have not yet been determined but it has been decided to use the same number of patients for each physical fitness group. Assume that a reasonable planning value for the error standard deviation is $\sigma = 4.5$ days.

a. What would be the required sample sizes if: (1) differences in the mean times for the three physical fitness categories are to be detected with probability .80 or more when the range of the treatment means is 5.63 days, and (2) the α risk is to be controlled at .01?

b. If the sample sizes determined in part (a) were employed, what would be the power of the test for treatment mean differences when $\mu_1 = 37$, $\mu_2 = 32$, and $\mu_3 = 28$?

c. Suppose the chief objective is to identify the physical fitness group with the smallest mean required time for therapy. The probability should be at least .90 that the correct group is identified when the mean required time for the second best group differs by 2.0 days or more. What are the required sample sizes?

*16.34. Refer to **Filling machines** Problem 16.11. Suppose that the sample sizes have not yet been determined but it has been decided to sample the same number of cartons for each filling machine. Assume that a reasonable planning value for the error standard deviation is $\sigma = .15$ ounce.

a. What would be the required sample sizes if: (1) differences in the mean amount of fill for the six filling machines are to be detected with probability .70 or more when the range of the treatment means is .15 ounce, and (2) the α risk is to be controlled at .05?

b. For the sample sizes determined in part (a), what would be the power of the test if $\mu_1 = .09$, $\mu_2 = .18$, $\mu_3 = .30$, $\mu_4 = .20$, $\mu_5 = .10$, and $\mu_6 = .20$?

c. Suppose the chief objective is to identify the filling machine with the smallest mean fill. The probability should be at least .95 that the filling machine with the smallest mean fill is recognized correctly when the filling machine with the next smallest mean fill differs by .10 ounce or more. What are the required sample sizes?

16.35. Refer to **Premium distribution** Problem 16.12. Suppose that the sample sizes have not yet been determined but it has been decided to sample the same number of premium distributions for each agent. Assume that a reasonable planning value for the error standard deviation is $\sigma = 3.0$ days.

a. What would be the required sample sizes if: (1) differences in the mean time lapse for the five agents are to be detected with probability .95 or more when the range of the treatment means is 3.75 days, and (2) the α risk is to be controlled at .10?

b. Suppose the chief objective is to identify the best agent, i.e., the one with the smallest mean time lapse. The probability should be at least .90 that the best agent is recognized correctly when the mean time lapse for the second best agent differs by 1.0 day or more. What are the required sample sizes?

Exercises

16.36. (Calculus needed.) State the likelihood function for ANOVA model (16.2) when $r = 3$ and $n_i \equiv 2$ and obtain the maximum likelihood estimators.

16.37. Show that when test statistic t^* in Table A.2a is squared, it is equivalent to the F^* test statistic (16.55) for $r = 2$.

16.38. Derive the restriction in (16.66) when the constant $\mu_.$ is defined according to (16.65).

16.39. a. Obtain the least squares estimators of the regression coefficients in full regression model (16.85). What is $SSE(F)$ here?

b. Obtain the least squares estimator of μ_i in reduced regression model (16.86). What is $SSE(R)$ here?

16.40. A completely randomized experiment is to be conducted involving $r = 3$ treatments, with $n = 2$ experimental trials for each treatment. Because the normality of the error terms is strongly in doubt, the test for treatment effects based on the F^* test statistic in (16.55) is to be carried out by means of the randomization distribution.

a. Determine the number of ways that the six experimental units can be divided into three groups of size two. How many unique F^* statistics are possible?

b. Using the results in part (a), what is the smallest P-value that is possible with the randomization test? What does this suggest about the adequacy of the planned sample size?

16.41. (Calculus needed.) Given $\mu_1 = 0$, $\mu_3 = 1$, and $0 \leq \mu_2 \leq 1$, show that $\sum(\mu_i - \mu_.)^2$ is minimized when $\mu_2 = .5$, where $\mu_. = (\mu_1 + \mu_2 + \mu_3)/3$.

Projects

16.42. Refer to the **SENIC** data set in Appendix C.1. Test whether or not the mean infection risk (variable 4) is the same in the four geographic regions (variable 9); use $\alpha = .05$. Assume that ANOVA model (16.2) is applicable. State the alternatives, decision rule, and conclusion.

16.43. Refer to the **SENIC** data set in Appendix C.1. The effect of average age of patient (variable 3) on mean infection risk (variable 4) is to be studied. For purposes of this ANOVA study, average age is to be classified into four categories: Under 50.0, 50.0–54.9, 55.0–59.9, 60.0 and over. Assume that ANOVA model (16.2) is applicable. Test whether or not the mean infection risk differs for the four age groups. Control the α risk at .10. State the alternatives, decision rule, and conclusion.

16.44. Refer to the **CDI** data set in Appendix C.2. The effect of geographic region (variable 17) on the crime rate (variable 10 ÷ variable 5) is to be studied. Assume that ANOVA model (16.2) is applicable. Test whether or not the mean crime rates for the four geographic regions differ; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

16.45. Refer to the **Market share** data set in Appendix C.3. Test whether or not the average monthly market share (variable 2) is the same for the four factor-level combinations associated with the two levels of each factor for discount price (variable 5) and package promotion (variable 6); use $\alpha = .05$. Assume that model (16.2) is applicable. State the alternatives, decision rule, and conclusion.

16.46. Consider a test involving H_0: $\mu_1 = \mu_2 = \mu_3$. Five observations are to be taken for each factor level, and level of significance $\alpha = .05$ is to be employed in the test.

a. Generate five random normal observations when $\mu_1 = 100$ and $\sigma = 12$ to represent the observations for treatment 1. Repeat this for the other two treatments when $\mu_2 = \mu_3 = 100$ and $\sigma = 12$. Finally, calculate F^* test statistic (16.55).

b. Repeat part (a) 100 times.

c. Calculate the mean of the 100 F^* statistics.

d. What proportion of the F^* statistics lead to conclusion H_0? Is this consistent with theoretical expectations?

e. Repeat parts (a) and (b) when $\mu_1 = 80$, $\mu_2 = 60$, $\mu_3 = 160$, and $\sigma = 12$. Calculate the mean of the 100 F^* statistics. How does this mean compare with the mean obtained in part (c) when $\mu_1 = \mu_2 = \mu_3 = 100$? Is this result consistent with the expectation in (16.37b)?

f. What proportion of the 100 test statistics obtained in part (e) lead to conclusion H_a? Does it appear that the test has satisfactory power when $\mu_1 = 80$, $\mu_2 = 60$, and $\mu_3 = 160$?

16.47. A completely randomized experiment involving $r = 2$ treatments was carried out, based on $n = 3$ experimental trials for each treatment. The test for equality of the treatment means is to be carried out by means of the randomization distribution of the F^* test statistic (16.55).

a. Determine the number of ways that the six experimental units can be divided into two groups of size three each. How many unique F^* statistics are possible?

b. For the sample results:

j:	1	2	3
Y_{1j}:	23	34	78
Y_{2j}:	17	29	23

obtain the randomization distribution of the test statistic F^* and the P-value of the randomization test.

c. Obtain the P-value of the normal-theory F^* statistic for the sample results in part (b). How does this P-value compare with the one from the randomization test in part (b)? What does this suggest about the appropriateness of the F distribution here if the error terms are far from normally distributed?

16.48. A completely randomized psychological reinforcement experiment was conducted in which a standard treatment and an experimental treatment were each applied to four subjects. The sample results are:

j:	1	2	3	4
Y_{1j} (standard treatment):	16	14	18	16
Y_{2j} (experimental treatment):	12	15	13	12

The test for equality of treatment means is to be carried out by means of the randomization distribution of the F^* test statistic (16.55), with $\alpha = .10$.

a. Obtain the randomization distribution of the test statistic F^* and carry out the indicated test. State the alternatives, decision rule, and conclusion. What is the P-value of the randomization test?

b. For the randomization distribution in part (a), determine the proportion of F^* values that exceed $F(.90; 1, 6)$, the proportion of F^* values that exceed $F(.95; 1, 6)$, and the proportion that exceed $F(.99; 1, 6)$.

c. How do the proportions obtained in part (b) compare with the probabilities for the normal error model? Discuss.

Case Studies

16.49. Refer to the **Prostate cancer** data set in Appendix C.5. Carry out a one-way analysis of variance of this data set, where the response of interest is PSA level (variable 2) and the single factor is Gleason score (variable 9). The analysis should consider transformations of the response variable. Document steps taken in your analysis, and justify your conclusions.

16.50. Refer to the **Real estate sales** data set in Appendix C.7. Carry out a one-way analysis of variance of this data set, where the response of interest is sales price (variable 2) and the single factor is number of bedrooms (variable 4). Recode the number of bedrooms into four categories: 0–2, 3, 4, and greater than or equal to 5. The analysis should consider transformations of the response variable. Document steps taken in your analysis, and justify your conclusions.

16.51. Refer to the **Ischemic heart disease** data set in Appendix C.9. Carry out a one-way analysis of variance of this data set, where the response of interest is total cost (variable 2) and the single factor is total number of interventions (variable 5). Recode the number of interventions into six categories: 0, 1, 2, 3–4, 5–7, and greater than or equal to 8. The analysis should consider transformations of the response variable. Document steps taken in your analysis, and justify your conclusions.

Chapter 17

Analysis of Factor Level Means

17.1 Introduction

In Chapter 16, we discussed the F test for determining whether or not the factor level means μ_i differ. This is a preliminary test to establish whether detailed analysis of the factor level means is warranted. When this test leads to the conclusion that the factor level means μ_i are equal, and ANOVA model (16.2) is appropriate, no relation between the factor and the response variable is present and usually no further analysis of factor means is therefore indicated. On the other hand, when the F test leads to the conclusion that the factor level means μ_i differ, a relation between the factor and the response variable is present. In this latter case, a thorough analysis of the nature of the factor level means is usually undertaken. This is done in two principal ways:

1. Analysis of the factor level means of interest using estimation techniques.
2. Statistical tests concerning the factor level means of interest.

Often, the analysis of factor level means combines the two approaches. For instance, a two-sided confidence interval may be constructed initially for an effect of interest. A test concerning this effect is then carried out either by determining whether or not the confidence interval contains the hypothesized value or by constructing the appropriate test statistic.

When many related comparisons are to be made, testing often precedes estimation. This occurs, for instance, when each factor level effect is compared with every other one and the number of factor levels is not small. Here, statistical tests are often performed first to determine the *active* or statistically significant set of comparisons. Estimation techniques are then used to construct confidence intervals for the active comparisons.

Special simultaneous estimation and testing procedures, called multiple comparison procedures, are required when a series of interval estimates or tests are performed. These multiple comparison procedures preserve the overall confidence coefficient $1 - \alpha$, or the overall significance level α, for the family of inferences.

We first discuss three simple graphical methods for displaying the factor level means. Much of the remainder of the chapter is devoted to a consideration of important multiple comparison procedures. In Section 16.10 we introduced methods for determining sample

TABLE 17.1 Summary of Results—Kenton Food Company Example.

	Package Design (i)				
	1	2	3	4	Total
n_i	5	5	4	5	19
$Y_{i\cdot}$	73	67	78	136	354
$\bar{Y}_{i\cdot}$	14.6	13.4	19.5	27.2	18.63

Source of Variation	*SS*	*df*	*MS*
Between designs	588.22	3	196.07
Error	158.20	15	10.55
Total	746.42	18	

Package Design	Characteristics
1	3 colors, with cartoons
2	3 colors, without cartoons
3	5 colors, with cartoons
4	5 colors, without cartoons

sizes in single-factor studies based on the power approach. This chapter concludes with a discussion of the estimation approach to sample size planning.

Throughout this chapter, we continue to assume the usual single-factor ANOVA model. The cell means version of this model was given in (16.2):

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{17.1}$$

where:

μ_i are parameters
ε_{ij} are independent $N(0, \sigma^2)$

Our discussion of the analysis of factor means will be illustrated by two examples. The first is the Kenton Food Company example. Data for this example are provided in Table 16.1 on page 686, and the ANOVA table is displayed in Figure 16.5 on page 695. For convenience, we repeat the main results in Table 17.1. The second example, the rust inhibitor example, is described next.

Example

In a study of the effectiveness of different rust inhibitors, four brands (A, B, C, D) were tested. Altogether, 40 experimental units were randomly assigned to the four brands, with 10 units assigned to each brand. A portion of the results after exposing the experimental units to severe weather conditions is given in coded form in Table 17.2a. The higher the coded value, the more effective is the rust inhibitor. This study is a completely randomized design, where the levels of the single factor correspond to the four rust inhibitor brands.

The analysis of variance is shown in Table 17.2b. For level of significance $\alpha = .05$ for testing whether or not the four rust inhibitors differ in effectiveness, we require

17.2 d sis of re ts—Rust tor ple (data coded).

(a) Data

	Rust Inhibitor Brand			
	A	B	C	D
j	$i = 1$	$i = 2$	$i = 3$	$i = 4$
1	43.9	89.8	68.4	36.2
2	39.0	87.1	69.3	45.2
3	46.7	92.7	68.5	40.7
...	...	...	...	...
8	38.9	88.1	65.2	38.7
9	43.6	90.8	63.8	40.9
10	40.0	89.1	69.2	39.7
$\bar{Y}_{i\cdot}$	43.14	89.44	67.95	40.47

$\bar{Y}_{\cdot\cdot} = 60.25$

(b) Analysis of Variance

Source of Variation	SS	df	MS
Between brands	15,953.47	3	5,317.82
Error	221.03	36	6.140
Total	16,174.50	39	

$F(.95; 3, 36) = 2.87$. Using the mean squares from Table 17.2b, we obtain the test statistic:

$$F^* = \frac{MSTR}{MSE} = \frac{5{,}317.82}{6.140} = 866.1$$

Since $F^* = 866.1 > 2.87$, we conclude that the four rust inhibitors differ in effectiveness. The P-value of the test is 0+. We therefore wish to analyze the nature of the factor level effects, particularly whether one rust inhibitor is substantially more effective than the others.

17.2 Plots of Estimated Factor Level Means

Before undertaking formal analysis of the nature of the factor level effects, it is usually helpful to examine these factor effects informally from a plot of the estimated factor level means $\bar{Y}_{i\cdot}$. We shall take up three types of plots: (1) a line plot, (2) a bar graph, and (3) a main effects plot. All three plots are appropriate whether the sample sizes n_i are equal or not.

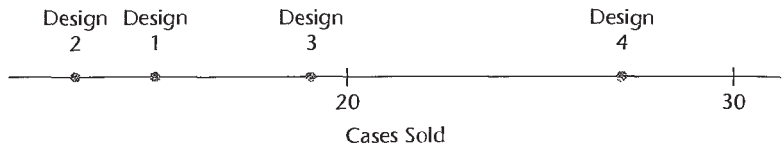
Line Plot

A line plot of the estimated factor level means simply shows the positions of the $\bar{Y}_{i\cdot}$ on a line scale. It is a very simple, but effective, device for indicating when one or several factor level means may differ substantially from the others.

Example

In Figure 17.1 we present a line plot of the estimated factor level means $\bar{Y}_{i\cdot}$ for the Kentorf Food Company example. It is clear from Figure 17.1 that design 4 led by far to the highest

FIGURE 17.1 Line Plot of Estimated Factor Level Means—Kenton Food Company Example.



mean sales in the study, and that package designs 1 and 2 led to the smallest mean sales which did not differ much from each other. The purpose of the formal inference procedures to be taken up shortly is to determine whether the pattern noted here reflects underlying differences in the factor level means μ_i or is simply the result of random variation.

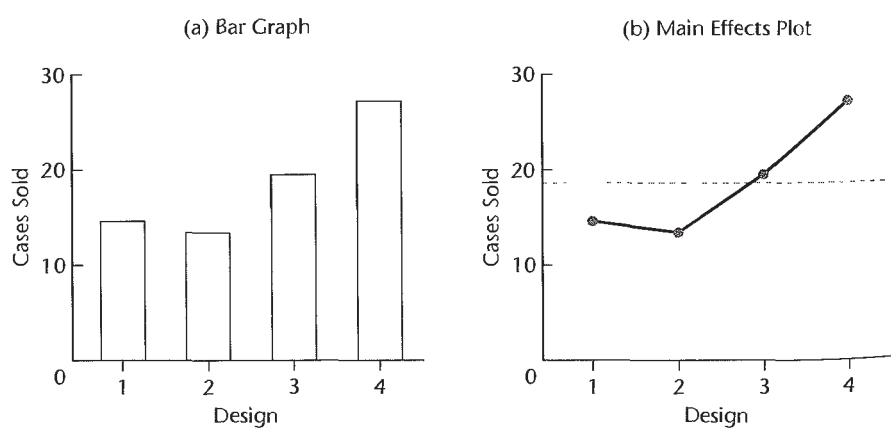
Bar Graph and Main Effects Plot

Bar graphs and main effects plots are frequently used to display the estimated factor level means in two dimensions. Both can be used to compare the magnitudes of different factor level means. In a bar graph, vertical bars are used to display the estimated factor level means. In a main effects plot, a scatter plot of the estimated factor level means is provided, and the plot symbols are connected by straight lines, to visibly highlight potential trends in the cell means. Note that these trend lines are not particularly meaningful for qualitative factors. For this reason, main effects plots are most appropriate for quantitative factors. In some packages, the main effects plot also displays the overall mean using a horizontal line, permitting visual comparisons of the factor-level means with the overall mean.

Example

A bar graph and a main effects plot of the estimated factor level means for the Kenton Food Company example are displayed in Figure 17.2. Because package design is a qualitative factor, the bar graph in Figure 17.2a is the recommended graphic here. An advantage of the main effects plot in Figure 17.2b is that it permits a visual comparison of the estimated factor level means and the overall mean. Here it shows that designs 3 and 4 had higher mean sales than the overall mean, while designs 1 and 2 both had smaller means sales than the overall mean.

FIGURE 17.2 MINITAB Bar Graph and Main Effects Plot of Estimated Factor Level Means—Kenton Food Company Example.



Comments

1. In Section 16.7 we defined the difference of the factor level mean and the overall mean as the factor level effect. In our discussion of multifactor studies in Chapter 19 and beyond, we shall refer to factor level effects as main effects. For this reason, the plot in Figure 17.2b is frequently referred to as a main effects plot.

2. None of the three plots provides information on the standard errors. Without such information, we cannot easily tell whether differences between factor level means are statistically significant. Later in this chapter, we shall enhance all three plots by including the information on the standard errors.

3. The normal probability plot introduced in Chapter 3 can also be used to compare the estimated factor level means. A normal probability plot is appropriate when the sample sizes n_i are equal and the number of factors r is sufficiently large. We recommend that a normal probability plot of factor level means be considered if $r \geq 10$. ■

17.3 Estimation and Testing of Factor Level Means

Inferences for factor level means are generally concerned with one or more of the following:

1. A single factor level mean μ_i
2. A difference between two factor level means
3. A contrast among factor level means
4. A linear combination of factor level means

We discuss each of these types of inferences in turn.

Inferences for Single Factor Level Mean

Estimation. An unbiased point estimator of the factor level mean μ_i is given in (16.16):

$$\hat{\mu}_i = \bar{Y}_{i\cdot} \tag{17.2}$$

This estimator has mean and variance:

$$E\{\bar{Y}_{i\cdot}\} = \mu_i \tag{17.3a}$$

$$\sigma^2\{\bar{Y}_{i\cdot}\} = \frac{\sigma^2}{n_i} \tag{17.3b}$$

The latter result follows because (16.43) indicates that $\bar{Y}_{i\cdot} = \mu_i + \bar{\varepsilon}_{i\cdot}$, the sum of a constant plus a mean of n_i independent ε_{ij} error terms, each of which has variance σ^2. Further, $\bar{Y}_{i\cdot}$ is normally distributed because the error terms ε_{ij} are independent normal random variables.

The estimated variance of $\bar{Y}_{i\cdot}$ is denoted by $s^2\{\bar{Y}_{i\cdot}\}$ and is obtained as usual by replacing σ^2 in (17.3b) by the unbiased point estimator *MSE*:

$$s^2\{\bar{Y}_{i\cdot}\} = \frac{MSE}{n_i} \tag{17.4}$$

The estimated standard deviation $s\{\bar{Y}_{i\cdot}\}$ is the positive square root of (17.4).

It can be shown that:

$$\frac{\bar{Y}_{i\cdot} - \mu_i}{s\{\bar{Y}_{i\cdot}\}} \text{ is distributed as } t(n_T - r) \text{ for ANOVA model (17.1)} \tag{17.5}$$

where the degrees of freedom are those associated with MSE. The result (17.5) follows from the definition of t in (A.44) since: (1) $\bar{Y}_{i\cdot}$ is normally distributed and (2) MSE/σ^2 is distributed independently of $\bar{Y}_{i\cdot}$ as $\chi^2(n_T - r)/(n_T - r)$ according to the following theorem:

For ANOVA model (17.1), SSE/σ^2 is distributed as χ^2 with $n_T - r$ degrees of freedom, and is independent of $\bar{Y}_{1\cdot}, \ldots, \bar{Y}_{r\cdot}$. **(17.6)**

It follows directly from (17.5) that the $1 - \alpha$ confidence limits for μ_i are:

$$\bar{Y}_{i\cdot} \pm t(1 - \alpha/2; n_T - r)s\{\bar{Y}_{i\cdot}\} \tag{17.7}$$

Testing. The confidence interval based on the limits in (17.7) can be used to test a hypothesis of the form:

$$\begin{aligned} H_0&: \mu_i = c \\ H_a&: \mu_i \neq c \end{aligned} \tag{17.8}$$

where c is an appropriate constant. We conclude H_0, at level of significance α, when c is contained in the confidence interval, and we conclude H_a when the confidence interval does not contain c. Equivalently, one can compute the test statistic:

$$t^* = \frac{\bar{Y}_{i\cdot} - c}{s\{\bar{Y}_{i\cdot}\}} \tag{17.9}$$

Test statistic t^* follows a t distribution with $n_T - r$ degrees of freedom when H_0 is true, according to (17.5). Consequently, we conclude H_0 whenever $|t^*| \leq t(1 - \alpha/2; n_T - r)$; otherwise, we conclude H_a.

Example

In the Kenton Food Company example, the sales manager wished to estimate mean sales for package design 1 with a 95 percent confidence interval. Using the results from Table 17.1, we have:

$$\bar{Y}_{1\cdot} = 14.6 \qquad n_1 = 5 \qquad MSE = 10.55$$

We require $t(.975; 15) = 2.131$. Finally, we need $s\{\bar{Y}_{1\cdot}\}$. We have:

$$s^2\{\bar{Y}_{1\cdot}\} = \frac{MSE}{n_1} = \frac{10.55}{5} = 2.110$$

so that $s\{\bar{Y}_{1\cdot}\} = 1.453$. Hence, we obtain the confidence limits $14.6 \pm 2.131(1.453)$ and the 95 percent confidence interval is:

$$11.5 \leq \mu_1 \leq 17.7$$

Thus, we estimate with confidence coefficient .95 that the mean sales per store for package design 1 are between 11.5 and 17.7 cases.

Graphical Displays. One way to enhance a bar graph or the main effects plot of factor level means is to display the confidence limits in (17.7) for each factor level mean. Figure 17.3 provides two such plots. Figure 17.3a contains a *bar-interval graph*, in which the 95 percent confidence limits are superimposed on a bar graph of the treatment means. Figure 17.3b contains an *interval plot*, in which the 95 percent confidence limits for each factor level

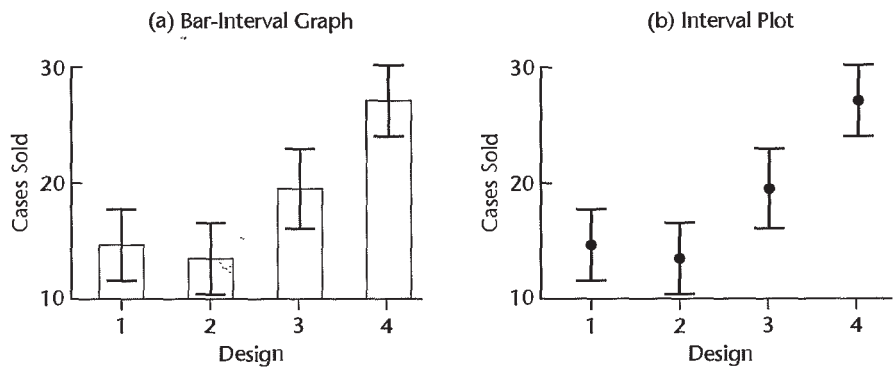


FIGURE 17.3 Bar-Interval Graph and Interval Plot—Kenton Food Company Example.

mean are displayed. Many investigators prefer to simply display limits that correspond to plus-or-minus one standard error—that is, $\bar{Y}_{i\cdot} \pm s\{\bar{Y}_{i\cdot}\}$.

Inferences for Difference between Two Factor Level Means

Estimation. Frequently two treatments or factor levels are to be compared by estimating the difference D between the two factor level means, say, μ_i and $\mu_{i'}$:

$$D = \mu_i - \mu_{i'} \tag{17.10}$$

Such a difference between two factor level means is called a *pairwise comparison*. A point estimator of D in (17.10), denoted by $\hat{D}$, is:

$$\hat{D} = \bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot} \tag{17.11}$$

This point estimator is unbiased:

$$E\{\hat{D}\} = \mu_i - \mu_{i'} \tag{17.12}$$

Since $\bar{Y}_{i\cdot}$ and $\bar{Y}_{i'\cdot}$ are independent, the variance of $\hat{D}$ follows from (A.31b):

$$\sigma^2\{\hat{D}\} = \sigma^2\{\bar{Y}_{i\cdot}\} + \sigma^2\{\bar{Y}_{i'\cdot}\} = \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) \tag{17.13}$$

The estimated variance of $\hat{D}$, denoted by $s^2\{\hat{D}\}$, is given by:

$$s^2\{\hat{D}\} = MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) \tag{17.14}$$

Finally, $\hat{D}$ is normally distributed by (A.40) because $\hat{D}$ is a linear combination of independent normal variables.

It follows from these characteristics, theorem (17.6), and the definition of t in (A.44) that:

$$\frac{\hat{D} - D}{s\{\hat{D}\}} \text{ is distributed as } t(n_T - r) \text{ for ANOVA model (17.1)} \tag{17.15}$$

Hence, the $1 - \alpha$ confidence limits for D are:

$$\hat{D} \pm t(1 - \alpha/2; n_T - r)s\{\hat{D}\} \tag{17.16}$$

Testing. There is often interest in testing whether two factor level means are the same. The alternatives here are of the form:

$$\begin{aligned} H_0&: \mu_i = \mu_{i'} \\ H_a&: \mu_i \neq \mu_{i'} \end{aligned} \tag{17.17}$$

The alternatives in (17.17) can be stated equivalently as follows:

$$\begin{aligned} H_0&: \mu_i - \mu_{i'} = 0 \\ H_a&: \mu_i - \mu_{i'} \neq 0 \end{aligned} \tag{17.17a}$$

Conclusion H_0 is reached at the α level of significance if zero is contained within the confidence limits (17.16); otherwise, conclusion H_a is reached. An equivalent procedure is based on the test statistic:

$$t^* = \frac{\hat{D}}{s\{\hat{D}\}} \tag{17.18}$$

Conclusion H_0 is reached if $|t^*| \leq t(1 - \alpha/2; n_T - r)$; otherwise, H_a is concluded.

Example

For the Kenton Food Company example, package designs 1 and 2 used 3-color printing and designs 3 and 4 used 5-color printing, as shown in Table 17.1. We wish to estimate the difference in mean sales for 5-color designs 3 and 4 using a 95 percent confidence interval. That is, we wish to estimate $D = \mu_3 - \mu_4$. From Table 17.1, we have:

$$\begin{aligned} \bar{Y}_{3\cdot} &= 19.5 \qquad n_3 = 4 \qquad MSE = 10.55 \\ \bar{Y}_{4\cdot} &= 27.2 \qquad n_4 = 5 \end{aligned}$$

Hence:

$$\hat{D} = \bar{Y}_{3\cdot} - \bar{Y}_{4\cdot} = 19.5 - 27.2 = -7.7$$

The estimated variance of $\hat{D}$ is:

$$s^2\{\hat{D}\} = MSE\left(\frac{1}{n_3} + \frac{1}{n_4}\right) = 10.55\left(\frac{1}{4} + \frac{1}{5}\right) = 4.748$$

so that the estimated standard deviation of $\hat{D}$ is $s\{\hat{D}\} = 2.179$. We require $t(.975; 15) = 2.131$. The confidence limits therefore are $-7.7 \pm 2.131(2.179)$, and the desired 95 percent confidence interval is:

$$-12.3 \leq \mu_3 - \mu_4 \leq -3.1$$

Thus, we estimate with confidence coefficient .95 that the mean sales for package design 3 fall short of those for package design 4 by somewhere between 3.1 and 12.3 cases per store.

Note from Table 17.1 that the only difference between package designs 3 and 4 is the presence of cartoons; both designs used 5-color printing. The sales manager may therefore wish to test whether the addition of cartoons affects sales for 5-color designs. The alternatives

here are:

$$H_0: \mu_3 - \mu_4 = 0$$
$$H_a: \mu_3 - \mu_4 \neq 0$$

Since the hypothesized difference zero in H_0 is not contained within the 95 percent confidence limits -12.3 and -3.1, we conclude H_a, that the presence of cartoons has an effect. We could also obtain test statistic (17.18):

$$t^* = \frac{\hat{D}}{s\{\hat{D}\}} = \frac{-7.7}{2.179} = -3.53$$

Since $|t^*| = 3.53 > t(.975; 15) = 2.131$, we conclude H_a. The two-sided P-value for this test is .003.

Inferences for Contrast of Factor Level Means

A *contrast* is a comparison involving two or more factor level means and includes the previous case of a pairwise difference between two factor level means in (17.10). A contrast will be denoted by L, and is defined as a linear combination of the factor level means μ_i where the coefficients c_i sum to zero:

$$L = \sum_{i=1}^{r} c_i \mu_i \qquad \text{where } \sum_{i=1}^{r} c_i = 0 \tag{17.19}$$

Illustrations of Contrasts. In the Kenton Food Company example, package designs 1 and 2 used 3-color printing and designs 3 and 4 used 5-color printing, as shown in Table 17.1. Also, package designs 1 and 3 utilized cartoons while no cartoons were utilized in designs 2 and 4. The following contrasts here may be of interest:

1. Comparison of the mean sales for the two 3-color designs:

$$L = \mu_1 - \mu_2$$

Here, $c_1 = 1$, $c_2 = -1$, $c_3 = 0$, $c_4 = 0$, and $\sum c_i = 0$.

2. Comparison of the mean sales for the 3-color and 5-color designs:

$$L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

Here, $c_1 = 1/2$, $c_2 = 1/2$, $c_3 = -1/2$, $c_4 = -1/2$, and $\sum c_i = 0$.

3. Comparison of the mean sales for designs with and without cartoons:

$$L = \frac{\mu_1 + \mu_3}{2} - \frac{\mu_2 + \mu_4}{2}$$

Here, $c_1 = 1/2$, $c_2 = -1/2$, $c_3 = 1/2$, $c_4 = -1/2$, and $\sum c_i = 0$.

4. Comparison of the mean sales for design 1 with average sales for all four designs:

$$L = \mu_1 - \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4}{4}$$

Here, $c_1 = 3/4$, $c_2 = -1/4$, $c_3 = -1/4$, $c_4 = -1/4$, and $\sum c_i = 0$.

Note that the first contrast is simply a pairwise comparison. In the second and third contrasts, averages of several factor level means are compared. The fourth contrast is the factor effect τ_1 defined by (16.60) and (16.63).

The averages used here are unweighted averages of the means μ_i; these are ordinarily the averages of interest. In special cases one might be interested in weighted averages of the μ_i to describe the mean response for a group of several factor levels. For example, if both 3-color and 5-color designs were to be employed, with 3-color printing used three times as often as 5-color printing, the comparison of the effect of cartoons versus no cartoons might be based on the contrast:

$$L = \frac{3\mu_1 + \mu_3}{4} - \frac{3\mu_2 + \mu_4}{4}$$

Here, $c_1 = 3/4$, $c_2 = -3/4$, $c_3 = 1/4$, $c_4 = -1/4$, and $\sum c_i = 0$.

Estimation. An unbiased estimator of a contrast L is:

$$\hat{L} = \sum_{i=1}^{r} c_i \bar{Y}_{i\cdot} \tag{17.20}$$

Since the $\bar{Y}_{i\cdot}$ are independent, the variance of $\hat{L}$ according to (A.31) is:

$$\sigma^2\{\hat{L}\} = \sum_{i=1}^{r} c_i^2 \sigma^2\{\bar{Y}_{i\cdot}\} = \sum_{i=1}^{r} c_i^2 \left(\frac{\sigma^2}{n_i}\right) = \sigma^2 \sum_{i=1}^{r} \frac{c_i^2}{n_i} \tag{17.21}$$

An unbiased estimator of this variance is:

$$s^2\{\hat{L}\} = MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i} \tag{17.22}$$

$\hat{L}$ is normally distributed by (A.40) because it is a linear combination of independent normal random variables. It can be shown by theorem (17.6), the characteristics of $\hat{L}$ just mentioned, and the definition of t that:

$$\frac{\hat{L} - L}{s\{\hat{L}\}} \text{ is distributed as } t(n_T - r) \text{ for ANOVA model (17.1)} \tag{17.23}$$

Consequently, the $1 - \alpha$ confidence limits for L are:

$$\hat{L} \pm t(1 - \alpha/2; n_T - r) s\{\hat{L}\} \tag{17.24}$$

Testing. The confidence interval based on the limits in (17.24) can be used to test a hypothesis of the form:

$$\begin{aligned} H_0&: L = 0 \\ H_a&: L \neq 0 \end{aligned} \tag{17.25}$$

H_0 is concluded at the α level of significance if zero is contained in the interval; otherwise H_a is concluded. An equivalent procedure is based on the test statistic:

$$t^* = \frac{\hat{L}}{s\{\hat{L}\}} \tag{17.26}$$

If $|t^*| \leq t(1 - \alpha/2; n_T - r)$, H_0 is concluded; otherwise, H_a is concluded.

Example

In the Kenton Food Company example, the mean sales for the 3-color designs are to be compared to the mean sales for the 5-color designs with a 95 percent confidence interval. We wish to estimate:

$$L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

The point estimate is (see data in Table 17.1):

$$\hat{L} = \frac{\bar{Y}_{1\cdot} + \bar{Y}_{2\cdot}}{2} - \frac{\bar{Y}_{3\cdot} + \bar{Y}_{4\cdot}}{2} = \frac{14.6 + 13.4}{2} - \frac{19.5 + 27.2}{2} = -9.35$$

Since $c_1 = 1/2$, $c_2 = 1/2$, $c_3 = -1/2$, and $c_4 = -1/2$, we obtain:

$$\sum \frac{c_i^2}{n_i} = \frac{(1/2)^2}{5} + \frac{(1/2)^2}{5} + \frac{(-1/2)^2}{4} + \frac{(-1/2)^2}{5} = .2125$$

and:

$$s^2\{\hat{L}\} = MSE \sum \frac{c_i^2}{n_i} = 10.55(.2125) = 2.242$$

so that $s\{\hat{L}\} = 1.50$.

For a 95 percent confidence interval, we require $t(.975; 15) = 2.131$. The confidence limits for L therefore are $-9.35 \pm 2.131(1.50)$, and the desired 95 percent confidence interval is:

$$-12.5 \leq L \leq -6.2$$

Therefore, we conclude with confidence coefficient .95 that mean sales for the 3-color designs fall below those for the 5-color designs by somewhere between 6.2 and 12.5 cases per store.

To test the hypothesis of no difference in mean sales for the 3-color and 5-color designs:

$$H_0: L = 0$$
$$H_a: L \neq 0$$

at the $\alpha = .05$ level of significance, we simply note that the hypothesized value zero is not contained in the 95 percent confidence interval. Hence, we conclude H_a, that the mean sales differ. To obtain a P-value of the test, test statistic (17.26) must be obtained. We find:

$$t^* = \frac{-9.35}{1.50} = -6.23$$

and the corresponding two-sided P-value is 0+.

Comment

Many single-factor analysis of variance programs permit the user to specify a contrast of interest and then will furnish the t^* test statistic or the equivalent F^* test statistic. ■

Inferences for Linear Combination of Factor Level Means

Occasionally, we are interested in a linear combination of the factor level means that is not a contrast. For example, suppose that the Kenton Food Company will use all four package designs, one in each of its four major marketing regions, and that these marketing regions

account for 35, 28, 12, and 25 percent of sales, respectively. In that case, there might be interest in the overall mean sales per store for all regions:

$$L = .35\mu_1 + .28\mu_2 + .12\mu_3 + .25\mu_4$$

Note that this linear combination is of the form $L = \sum c_i \mu_i$ but that the coefficients c_i sum to 1.0, not to zero as they must for a contrast.

We define a *linear combination of the factor level means* μ_i as:

$$L = \sum_{i=1}^{r} c_i \mu_i \tag{17.27}$$

with no restrictions on the coefficients c_i. Confidence limits and test statistics for a linear combination L are obtained in exactly the same way as those for a contrast by means of (17.24) and (17.26), respectively. Point estimator (17.20) and estimated variance (17.22) are still applicable when $\sum c_i \neq 0$.

Single Degree of Freedom Tests. The alternatives for tests concerning a factor level mean in (17.8), a difference between two factor level means in (17.17a), and a contrast of factor level means in (17.25) are all special cases of a test concerning a linear combination of factor level means:

$$H_0\colon \sum c_i \mu_i = c$$

$$H_a\colon \sum c_i \mu_i \neq c$$

where the c_i and c are appropriate constants. Test statistics (17.9), (17.18), and (17.26) can each be converted to an equivalent F^* test statistic by means of the relation in (A.50a):

$$F^* = (t^*)^2$$

Test statistic F^* follows the $F(1, n_T - r)$ distribution when H_0 holds. Note that the numerator degrees of freedom are always one. Hence, these tests are often referred to as *single-degree-of-freedom tests*. The t^* version of the test statistic is more versatile because it can also be used for one-sided tests while the F^* version cannot.

17.4 Need for Simultaneous Inference Procedures

The procedures for estimating and testing factor level means discussed up to this point have two important limitations:

1. The confidence coefficient $1-\alpha$ for the estimation procedures described is a statement confidence coefficient and applies only to a particular estimate, not to a series of estimates. Similarly, the specified Type I error rate, α, applies only to a particular test and not to a series of tests.

2. The confidence coefficient $1-\alpha$ and the specified significance level α are appropriate only if the estimate or test was not suggested by the data.

The first limitation is familiar from regression analysis. It is particularly serious for analysis of variance models because frequently many different comparisons are of interest

here, and one needs to piece the different findings together. Consider the very simple case where three different advertisements are being compared for their effectiveness in stimulating sales. The following estimates of their comparative effectiveness have been obtained, each with a 95 percent statement confidence coefficient:

$$59 \leq \mu_2 - \mu_1 \leq 62$$
$$-2 \leq \mu_3 - \mu_1 \leq 3$$
$$58 \leq \mu_2 - \mu_3 \leq 64$$

It would be natural here to piece the different comparisons together and conclude that advertisement 2 leads to highest mean sales, while advertisements 1 and 3 are substantially less effective and do not differ much among themselves. One would therefore like a family confidence coefficient for this family of statements, to provide known assurance that the set of conclusions is correct.

The same concern for assurance of correct conclusions exists when the inferences involve tests. An analysis of factor means by testing procedures usually involves several single-degree-of-freedom tests to answer related questions. For instance, the sales manager of the Kenton Food Company might wish to know both whether the number of colors has an effect on mean sales and whether the use of cartoons has an effect. Whenever several tests are conducted, both the level of significance and the power, insofar as the family of tests is concerned, are affected. Consider, for example, three different t tests, each conducted with $\alpha = .05$. The probability that each of the tests will lead to conclusion H_0 when indeed H_0 is correct in each case, assuming independence of the tests, is $(.95)^3 = .857$. Thus, the level of significance that at least one of the three tests leads to conclusion H_a when H_0 holds in each case would be $1 - .857 = .143$, not .05. We see then that the level of significance and power for a *family* of tests is not the same as that for an *individual* test. Actually, the t^* statistics are dependent when they all are based on the same sample data and use the same MSE value. It is often therefore more difficult to determine the actual level of significance and power for a family of tests.

The second limitation of the procedures for estimating or testing factor level means discussed so far, namely, that the estimate or test must not be suggested by the data, is an important one in exploratory investigations where many new questions are often suggested once the data are being analyzed. The process of studying effects suggested by the data is sometimes called *data snooping*. One form of data snooping is to investigate comparisons where the effect appears to be large from the sample data, for example, testing whether there is a difference between the two treatment means corresponding to the smallest and largest estimated factor level means $\bar{Y}_{l.}$. Choosing the test in this manner implies a larger significance level than the nominal level used in constructing the decision rule. For example, it can be shown for a study with six factor levels that if the analyst will always compare the smallest and largest estimated factor level means by using the confidence limits (17.16) with a 95 percent confidence coefficient, the interval estimate will not contain zero and therefore suggest a real effect 40 percent of the time when indeed there is no difference between any of the factor level means (Ref. 17.1). Hence, the α level for the test is .40, not .05. With a larger number of factor levels, the likelihood of an erroneous indication of a real effect, i.e., the actual α level, would be even greater. The reason for the higher actual level of significance here is that a family of tests is being conducted implicitly since the analyst

does not know in advance which estimated factor level means will be the extreme ones. The situation here is analogous to that in Chapter 10 where the test to determine whether the largest absolute residual is an outlier considers the family of tests for each of the n residuals.

One solution to this problem of making comparisons that are suggested by initial analysis of the data is to use a multiple comparison procedure where the family of inferences includes all the possible inferences that can be anticipated to be of potential interest after the data are examined. For instance, in an investigation where five factor level means are being studied, it is decided in advance that principal interest is in three pairwise comparisons. However, it is also agreed that other pairwise comparisons that will appear interesting should be studied as well. In this case, the family of *all* pairwise comparisons can be used as the basis for obtaining an appropriate family confidence coefficient or significance level for the comparisons suggested by the data.

In the next three sections, we shall discuss three multiple comparison procedures for analysis of variance models that permit the family confidence coefficient and the family α risk to be controlled. Two of these procedures, the Tukey and Scheffé procedures, allow data snooping to be undertaken naturally without affecting the confidence coefficient or significance level. The other procedure, the Bonferroni procedure, is applicable only when the effects to be investigated are identified in advance of the study.

17.5 Tukey Multiple Comparison Procedure

The Tukey multiple comparison procedure that we will consider here applies when:

The family of interest is the set of all pairwise comparisons of factor level means; in other words, the family consists of estimates of all pairs $D = \mu_i - \mu_{i'}$ or of all tests of the form:

$$H_0\colon \mu_i - \mu_{i'} = 0$$
$$H_a\colon \mu_i - \mu_{i'} \neq 0$$

When all sample sizes are equal, the family confidence coefficient for the Tukey method is exactly $1 - \alpha$ and the family significance level is exactly α. When the sample sizes are not equal, the family confidence coefficient is greater than $1 - \alpha$ and the family significance level is less than α. In other words, the Tukey procedure is conservative when the sample sizes are not equal.

Studentized Range Distribution

The Tukey procedure utilizes the *studentized range distribution*. Suppose that we have r independent observations $Y_1, \ldots, Y_r$ from a normal distribution with mean μ and variance σ^2. Let w be the range for this set of observations; thus:

$$w = \max(Y_i) - \min(Y_i) \tag{17.28}$$

Suppose further that we have an estimate s^2 of the variance σ^2 which is based on ν degrees of freedom and is independent of the Y_i. Then, the ratio w/s is called the *studentized range*. It is denoted by:

$$q(r, \nu) = \frac{w}{s} \tag{17.29}$$

where the arguments in parentheses remind us that the distribution of q depends on r and ν. The distribution of q has been tabulated, and selected percentiles are presented in Table B.9.

This table is simple to use. Suppose that $r = 5$ and $\nu = 10$. The 95th percentile is then $q(.95; 5, 10) = 4.65$, which means:

$$P\left\{\frac{w}{s} = q(5, 10) \leq 4.65\right\} = .95$$

Thus, with five normal Y observations, the probability is .95 that their range is not more than 4.65 times as great as an independent sample standard deviation based on 10 degrees of freedom.

Simultaneous Estimation

The Tukey multiple comparison confidence limits for all pairwise comparisons $D = \mu_i - \mu_{i'}$ with family confidence coefficient of at least $1 - \alpha$ are as follows:

$$\hat{D} \pm Ts\{\hat{D}\} \tag{17.30}$$

where:

$$\hat{D} = \bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot} \tag{17.30a}$$

$$s^2\{\hat{D}\} = s^2\{\bar{Y}_{i\cdot}\} + s^2\{\bar{Y}_{i'\cdot}\} = MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) \tag{17.30b}$$

$$T = \frac{1}{\sqrt{2}}q(1 - \alpha; r, n_T - r) \tag{17.30c}$$

Note that the point estimator $\hat{D}$ in (17.30a) and the estimated variance in (17.30b) are the same as those in (17.11) and (17.14) for a single pairwise comparison. Thus, the only difference between the Tukey confidence limits (17.30) for simultaneous comparisons and those in (17.16) for a single comparison is the multiple of the estimated standard deviation.

The family confidence coefficient $1 - \alpha$ pertaining to the multiple pairwise comparisons refers to the proportion of correct families, each consisting of all pairwise comparisons, when repeated sets of samples are selected and all pairwise confidence intervals are calculated each time. A family of pairwise comparisons is considered to be correct if every pairwise comparison in the family is correct. Thus, a family confidence coefficient of $1 - \alpha$ indicates that all pairwise comparisons in the family will be correct in $(1 - \alpha)100$ percent of the repetitions.

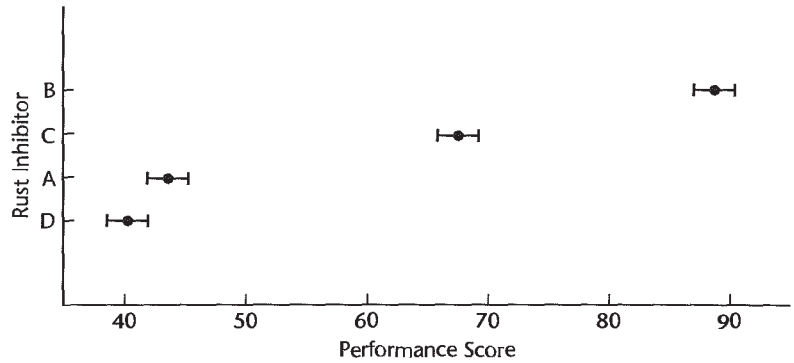
Simultaneous Testing

When we wish to conduct a family of tests of the form:

$$\begin{aligned} H_0&: \mu_i - \mu_{i'} = 0 \\ H_a&: \mu_i - \mu_{i'} \neq 0 \end{aligned} \tag{17.31}$$

for all pairwise comparisons, the family of confidence intervals based on (17.30) may be utilized for this purpose. We simply determine for each interval whether or not zero is contained in the interval. If zero is contained, conclusion H_0 is reached; otherwise, H_a is concluded. By following this procedure, the family level of significance will not exceed α.

FIGURE 17.4 Paired Comparison Plot—Rust Inhibitor Example.



Equivalently, the pairwise tests can be conducted directly by calculating for each pairwise comparison the test statistic:

$$q^* = \frac{\sqrt{2}\hat{D}}{s\{\hat{D}\}} \tag{17.32}$$

where $\hat{D}$ and $s^2\{\hat{D}\}$ are given in (17.30). Conclusion H_0 in (17.31) is reached if $|q^*| \leq q(1-\alpha; r; n_T - r)$; otherwise, H_a is concluded.

A *paired comparison plot* provides still another means of conducting all pairwise tests with the Tukey procedure when all sample sizes are equal, i.e., when $n_i \equiv n$. This plot provides a graphic means of making all pairwise comparisons. Around each estimated treatment mean $\bar{Y}_{i\cdot}$ is plotted an interval whose limits are:

$$\bar{Y}_{i\cdot} \pm \frac{1}{2} T s\{\hat{D}\} \tag{17.33}$$

When the intervals overlap on this plot, the formal test leads to the conclusion that the two treatment means do not differ. When the intervals do not overlap, the formal test leads to the conclusion that the two treatment means differ. In addition, the paired comparison plot shows the direction of the difference.

Figure 17.4 provides an illustration of a paired comparison plot for the rust inhibitor example. There is no overlap between the intervals for rust inhibitors B and C, indicating that the mean performances differ for these two rust inhibitors. Figure 17.4 in addition shows that rust inhibitor B is superior to C since its interval is considerably to the right of that for C, thus providing directional information about the difference in mean performance for the two rust inhibitors. We discuss this plot in greater detail on page 750.

Example 1—Equal Sample Sizes

In the rust inhibitor example in Table 17.2, it was desired to estimate all pairwise comparisons by means of the Tukey procedure, using a family confidence coefficient of 95 percent. Since $r = 4$ and $n_T - r = 36$, we find the required percentile of the studentized range distribution from Table B.9 to be $q(.95; 4, 36) = 3.814$. Hence, by (17.30c), we obtain:

$$T = \frac{1}{\sqrt{2}}(3.814) = 2.70$$

TABLE 17.3 Simultaneous Confidence Intervals and Tests for Pairwise Differences Using the Tukey Procedure—Rust Inhibitor Example.

	Test		
Confidence Interval	H_0	H_a	q^*
$43.3 \leq \mu_2 - \mu_1 \leq 49.3$	$\mu_2 = \mu_1$	$\mu_2 \neq \mu_1$	58.99
$21.8 \leq \mu_3 - \mu_1 \leq 27.8$	$\mu_3 = \mu_1$	$\mu_3 \neq \mu_1$	31.61
$-.3 \leq \mu_1 - \mu_4 \leq 5.7$	$\mu_1 = \mu_4$	$\mu_1 \neq \mu_4$	3.40
$18.5 \leq \mu_2 - \mu_3 \leq 24.5$	$\mu_2 = \mu_3$	$\mu_2 \neq \mu_3$	27.37
$46.0 \leq \mu_2 - \mu_4 \leq 52.0$	$\mu_2 = \mu_4$	$\mu_2 \neq \mu_4$	62.39
$24.5 \leq \mu_3 - \mu_4 \leq 30.5$	$\mu_3 = \mu_4$	$\mu_3 \neq \mu_4$	35.01

Further, we need $s\{\hat{D}\}$. Using (17.30b), we find for any pairwise comparison since equal sample sizes were employed:

$$s^2\{\hat{D}\} = MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right) = 6.140\left(\frac{1}{10} + \frac{1}{10}\right) = 1.23$$

so that $s\{\hat{D}\} = 1.11$. Hence, we obtain for each pairwise comparison:

$$Ts\{\hat{D}\} = 2.70(1.11) = 3.0$$

To illustrate the calculation of the pairwise confidence limits, consider the estimation of the difference between the treatment means for rust inhibitors A and B, $\mu_2 - \mu_1$:

$$\hat{D} = \bar{Y}_{2\cdot} - \bar{Y}_{1\cdot} = 89.44 - 43.14 = 46.3$$

The confidence limits from (17.30) therefore are 46.3 ± 3.0 and the confidence interval is:

$$43.3 \leq \mu_2 - \mu_1 \leq 49.3$$

The complete family of pairwise confidence intervals is listed in the left column of Table 17.3. The pairwise comparisons indicate that all but one of the differences (D and A) are statistically significant (confidence interval does not cover zero).

We incorporate this information in a line plot of the estimated factor level means by underlining nonsignificant comparisons.

The line between D and A indicates that there is no clear evidence whether D or A is the better rust inhibitor. The absence of a line signifies that a difference in performance has been found and the location of the points indicates the direction of the difference. Thus, the multiple comparison procedure permits us to infer with a 95 percent family confidence coefficient for the chain of conclusions that B is the best inhibitor (better by somewhere between 18.5 and 24.5 units than the second best), C is second best, and A and D follow substantially behind with little or no difference between them.

The same conclusions are obtained if we carry out all pairwise tests using the simultaneous testing procedure based on test statistic (17.32). For example, to test:

$$H_0: \mu_2 - \mu_1 = 0$$
$$H_a: \mu_2 - \mu_1 \neq 0$$

we require the test statistic:

$$q^* = \frac{\sqrt{2}(89.44 - 43.14)}{1.11} = 58.99$$

Because $|q^*| = 58.99 > q(.95; 4, 36) = 3.814$, we conclude H_a, that the two treatment means differ. The test statistics q^* for the family of all pairwise tests are listed in the right column of Table 17.3. The absolute values of all test statistics exceed 3.814 except for one, so that all differences are found to be statistically significant except for that involving μ_1 and μ_4 (A and D). For this case, $|q^*| = 3.40$ does not exceed the critical value 3.814.

Figure 17.4 presents a paired comparison plot for the rust inhibitor example. Here are plotted the estimated treatment means $\bar{Y}_{i\cdot}$ with the comparison intervals based on (17.33). For example, for rust inhibitor A, we have from earlier:

$$\bar{Y}_{1\cdot} = 43.14 \qquad T = 2.70 \qquad s\{\hat{D}\} = 1.11$$

so that the comparison limits in (17.33) are:

$$43.14 \pm \frac{1}{2}(2.70)(1.11) \qquad \text{or} \qquad 41.64 \quad \text{and} \quad 44.64$$

We readily see that only the intervals for A and D overlap, that rust inhibitor B is clearly best, that rust inhibitor C is second best, and that rust inhibitors A and D are the least effective.

Example 2—Unequal Sample Sizes

In the Kenton Food Company example in Table 17.1, the sales manager was interested in the comparative performance of the four package designs. The analyst developed all pairwise comparisons by means of the Tukey procedure with a family confidence coefficient of at least 90 percent. Since the sample sizes are not equal here, the estimated standard deviation $s\{\hat{D}\}$ must be recalculated for each pairwise comparison. To compare designs 1 and 2, for instance, we obtain:

$$\hat{D} = \bar{Y}_{1\cdot} - \bar{Y}_{2\cdot} = 14.6 - 13.4 = 1.2$$

$$s^2\{\hat{D}\} = MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right) = 10.55\left(\frac{1}{5} + \frac{1}{5}\right) = 4.22$$

$$s\{\hat{D}\} = 2.05$$

For a 90 percent family confidence coefficient, we require $q(.90; 4, 15) = 3.54$ so that we obtain:

$$T = \frac{1}{\sqrt{2}}(3.54) = 2.50$$

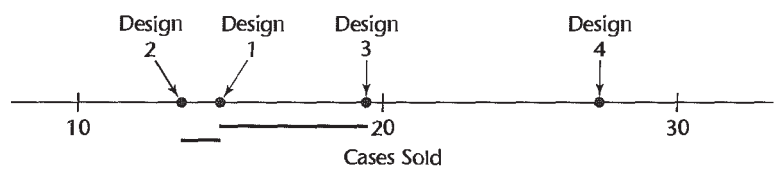
Hence, the confidence limits are $1.2 \pm 2.50(2.05)$ and the confidence interval for $\mu_1 - \mu_2$ is:

$$-3.9 \leq \mu_1 - \mu_2 \leq 6.3$$

In the same way, we obtain the other five confidence intervals:

$$-.6 = (19.5 - 14.6) - 2.50(2.18) \leq \mu_3 - \mu_1 \leq (19.5 - 14.6) + 2.50(2.18) = 10.4$$
$$7.5 = (27.2 - 14.6) - 2.50(2.05) \leq \mu_4 - \mu_1 \leq (27.2 - 14.6) + 2.50(2.05) = 17.7$$
$$.7 = (19.5 - 13.4) - 2.50(2.18) \leq \mu_3 - \mu_2 \leq (19.5 - 13.4) + 2.50(2.18) = 11.6$$
$$8.7 = (27.2 - 13.4) - 2.50(2.05) \leq \mu_4 - \mu_2 \leq (27.2 - 13.4) + 2.50(2.05) = 18.9$$
$$2.3 = (27.2 - 19.5) - 2.50(2.18) \leq \mu_4 - \mu_3 \leq (27.2 - 19.5) + 2.50(2.18) = 13.2$$

We summarize the comparative performance by a line plot, indicating each nonsignificant difference by a rule.



We can conclude with at least 90 percent family confidence that design 4 is clearly the most effective design. However, the small-scale study does not permit a complete ordering among the other three designs. Design 3 is more effective than design 2 but may not be more effective than design 1, which in turn may not be more effective than design 2.

Often, the results of the family of pairwise tests are summarized by setting up groups of factor levels whose means do not differ according to the single degree of freedom tests. For the Kenton Food Company example, there are three such groups:

Group 1		Group 2		Group 3	
Design 4	$\bar{Y}_{4\cdot} = 27.2$	Design 3	$\bar{Y}_{3\cdot} = 19.5$	Design 1	$\bar{Y}_{1\cdot} = 14.6$
		Design 1	$\bar{Y}_{1\cdot} = 14.6$	Design 2	$\bar{Y}_{2\cdot} = 13.4$

Comments

1. When the Tukey procedure is used with unequal sample sizes, it is sometimes called the *Tukey-Kramer procedure.*

2. When not all pairwise comparisons are of interest, the confidence coefficient for the family of comparisons under consideration will be greater than the specification $1 - \alpha$ used in setting up the Tukey intervals. Similarly, the family significance level for simultaneous testing will be less than α.

3. The Tukey procedure can be used for data snooping as long as the effects to be studied on the basis of preliminary data analysis are pairwise comparisons.

4. The Tukey procedure can be modified to handle general contrasts of factor level means. We do not discuss this modification since the Scheffé method (to be discussed next) is to be preferred for this situation.

5. To derive the Tukey simultaneous confidence intervals for the case when all sample sizes are equal, i.e., when $n_i \equiv n$ so that $n_T = rn$, consider the deviations:

$$(\bar{Y}_{1\cdot} - \mu_1), \ldots, (\bar{Y}_{r\cdot} - \mu_r) \tag{17.34}$$

and assume that ANOVA model (17.1) applies. The deviations in (17.34) are then independent variables (because the error terms are independent), they are normally distributed (because the error terms are independent normal variables), they have the same expectation zero (because μ_i is subtracted from $\bar{Y}_{i\cdot}$), and they have the same variance σ^2/n. Further, MSE/n is an estimator of σ^2/n that is independent of the deviations $(\bar{Y}_{i\cdot} - \mu_i)$ per theorem (17.6). Thus, it follows from the definition of the studentized range q in (17.29) that:

$$\frac{\max(\bar{Y}_{i\cdot} - \mu_i) - \min(\bar{Y}_{i\cdot} - \mu_i)}{\sqrt{\dfrac{MSE}{n}}} \sim q(r, n_T - r) \tag{17.35}$$

where $n_T - r$ is the number of degrees of freedom associated with MSE, $\max(\bar{Y}_{i\cdot} - \mu_i)$ is the largest deviation, and $\min(\bar{Y}_{i\cdot} - \mu_i)$ is the smallest deviation.

In view of (17.35), we can write the following probability statement:

$$P\left\{\frac{\max(\bar{Y}_{i\cdot} - \mu_i) - \min(\bar{Y}_{i\cdot} - \mu_i)}{\sqrt{\dfrac{MSE}{n}}} \leq q(1-\alpha; r, n_T - r)\right\} = 1 - \alpha \tag{17.36}$$

Note now that the following inequality holds for *all* pairs of factor levels i and i':

$$|(\bar{Y}_{i\cdot} - \mu_i) - (\bar{Y}_{i'\cdot} - \mu_{i'})| \leq \max(\bar{Y}_{i\cdot} - \mu_i) - \min(\bar{Y}_{i\cdot} - \mu_i) \tag{17.37}$$

The absolute value at the left is needed since the factor levels i and i' are not ordered so that we may be subtracting the larger deviation from the smaller. To put this another way, we are merely concerned here with the difference between the two factor level deviations regardless of direction.

Since inequality (17.37) holds for all pairs of factor levels i and i', it follows from (17.36) that the probability:

$$P\left\{\left|\frac{(\bar{Y}_{i\cdot} - \mu_i) - (\bar{Y}_{i'\cdot} - \mu_{i'})}{\sqrt{\dfrac{MSE}{n}}}\right| \leq q(1-\alpha; r, n_T - r)\right\} = 1 - \alpha \tag{17.38}$$

holds for all $r(r-1)/2$ pairwise comparisons among the r factor levels. By rearranging the inequality in (17.38), using the definitions of $s^2\{\hat{D}\}$ in (17.30b) and of T in (17.30c), and noting that for the equal sample size case $s^2\{\hat{D}\}$ becomes:

$$s^2\{\hat{D}\} = MSE\left(\frac{1}{n} + \frac{1}{n}\right) = \frac{2MSE}{n} \qquad \text{when } n_i \equiv n$$

we obtain the Tukey multiple comparison confidence limits in (17.30).

6. When the Tukey multiple comparison procedure is used for testing pairwise differences as in (17.31), the tests are sometimes called *honestly significant difference tests*.

7. The pairwise comparison plot can be used as an approximate plot when the sample sizes are not equal, provided that the sample sizes do not differ greatly. For this case, the comparison limits

should be obtained as follows:

$$\bar{Y}_{i\cdot} \pm \frac{1}{2} q(1-\alpha; r, n_T - r) s\{\bar{Y}_{i\cdot}\} \tag{17.39}$$

The limits in (17.39) are identical to those in (17.33) when the sample sizes are equal. ■

17.6 Scheffé Multiple Comparison Procedure

The Scheffé multiple comparison procedure was encountered previously for regression models. It is also applicable for analysis of variance models. It applies for analysis of variance models when:

The family of interest is the set of all possible contrasts among the factor level means:

$$L = \sum c_i \mu_i \qquad \text{where } \sum c_i = 0 \tag{17.40}$$

In other words, the family consists of estimates of all possible contrasts L or of tests concerning all possible contrasts of the form:

$$H_0\colon L = 0$$
$$H_a\colon L \neq 0$$

Thus, infinitely many statements belong to this family. The family confidence level for the Scheffé procedure is exactly $1-\alpha$, and the family significance level is exactly α, whether the factor level sample sizes are equal or unequal.

Simultaneous Estimation

We noted earlier that an unbiased estimator of L is:

$$\hat{L} = \sum c_i \bar{Y}_{i\cdot} \tag{17.41}$$

for which the estimated variance is:

$$s^2\{\hat{L}\} = MSE \sum \frac{c_i^2}{n_i} \tag{17.42}$$

The Scheffé confidence intervals for the family of contrasts L are of the form:

$$\hat{L} \pm S s\{\hat{L}\} \tag{17.43}$$

where:

$$S^2 = (r-1) F(1-\alpha; r-1, n_T - r) \tag{17.43a}$$

and $\hat{L}$ and $s\{\hat{L}\}$ are given by (17.41) and (17.42), respectively. If we were to calculate the confidence intervals in (17.43) for all conceivable contrasts, then in $(1-\alpha)100$ percent of repetitions of the experiment, the entire set of confidence intervals in the family would be correct.

Note that the simultaneous confidence limits in (17.43) differ from those for a single confidence limit in (17.24) only with respect to the multiple of the estimated standard deviation.

Simultaneous Testing

Tests involving contrasts of the form:

$$H_0: L = 0$$
$$H_a: L \neq 0 \tag{17.44}$$

can be carried out by examination of the corresponding Scheffé confidence intervals based on (17.43). H_0 is concluded at the α family level of significance if the confidence interval includes zero; otherwise H_a is concluded. An equivalent direct testing procedure for the alternatives in (17.44) uses the test statistic:

$$F^* = \frac{\hat{L}^2}{(r-1)s^2\{\hat{L}\}} \tag{17.45}$$

Conclusion H_0 in (17.44) is reached at the α family significance level if $F^* \leq F(1-\alpha; r-1, n_T - r)$; otherwise, H_a is concluded.

Example

In the Kenton Food Company example, interest centered on estimating the following four contrasts with family confidence coefficient .90:

Comparison of 3-color and 5-color designs:

$$L_1 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

Comparison of designs with and without cartoons:

$$L_2 = \frac{\mu_1 + \mu_3}{2} - \frac{\mu_2 + \mu_4}{2}$$

Comparison of the two 3-color designs:

$$L_3 = \mu_1 - \mu_2$$

Comparison of the two 5-color designs:

$$L_4 = \mu_3 - \mu_4$$

Consider first the estimation of L_1. Earlier, we found:

$$\hat{L}_1 = -9.35$$
$$s\{\hat{L}_1\} = 1.50$$

Since $r - 1 = 3$ and $n_T - r = 15$ (Table 17.1), we have:

$$S^2 = (r-1)F(1-\alpha; r-1, n_T - r) = 3F(.90; 3, 15) = 3(2.49) = 7.47$$

so that $S = 2.73$. Hence, the 90 percent confidence limits for L_1 by the Scheffé multiple comparison procedure are $-9.35 \pm 2.73(1.50)$ and the desired confidence interval is:

$$-13.4 \leq L_1 \leq -5.3$$

In similar fashion, we obtain the other desired confidence intervals, and the entire set is:

$$-13.4 \leq L_1 \leq -5.3$$
$$-7.3 \leq L_2 \leq .8$$
$$-4.4 \leq L_3 \leq 6.8$$
$$-13.7 \leq L_4 \leq -1.7$$

Note that the confidence interval for L_1 does not include zero. Hence, if we wished to test H_0: $L_1 = 0$ versus H_a: $L_1 \neq 0$, we would conclude H_a, that the mean sales for 3-color and 5-color designs differ. The confidence interval provides additional information, however; namely, that mean sales for 5-color designs exceed mean sales for 3-color designs, by somewhere between 5.3 and 13.4 cases per store.

Any chain of conclusions derived from the set of confidence intervals has associated with it family confidence coefficient .90. The principal conclusions drawn by the sales manager were as follows: 5-color designs lead to higher mean sales than 3-color designs, the increase being somewhere between 5 and 13 cases per store. No overall effect of cartoons in the package design is indicated, although the use of a cartoon in 5-color designs leads to lower mean sales than when no cartoon is used.

Comments

1. If in the Kenton Food Company example we had wished to estimate a single contrast with statement confidence coefficient .90, the required t value would have been $t(.95; 15) = 1.753$. This t value is smaller than the Scheffé multiple $S = 2.73$, so that the single confidence interval would be somewhat narrower. The increased width of the interval with the Scheffé procedure is the price paid for a known confidence coefficient for a family of statements and a chain of conclusions drawn from them, and for the possibility of making comparisons not specified in advance of the data analysis.

2. Since applications of the Scheffé procedure never involve all conceivable contrasts, the confidence coefficient for the finite family of statements actually considered will be greater than $1 - \alpha$ so that $1 - \alpha$ serves as a guaranteed lower bound. Similarly, the significance level for the finite family of tests considered will be less than α. For this reason, it has been suggested that lower confidence levels and higher significance levels be used with the Scheffé procedure than would ordinarily be employed. Confidence coefficients of 90 percent and 95 percent and significance levels of $\alpha = .10$ and $\alpha = .05$ with the Scheffé procedure are frequently mentioned.

3. The Scheffé procedure can be used for a wide variety of data snooping since the family of statements contains all possible contrasts. ■

Comparison of Scheffé and Tukey Procedures

1. If only pairwise comparisons are to be made, the Tukey procedure gives narrower confidence limits and is therefore the preferred method.

2. The Scheffé procedure has the property that if the F test of factor level equality indicates that the factor level means μ_i are not equal, the corresponding Scheffé multiple comparison procedure will find at least one contrast (out of all possible contrasts) that differs significantly from zero (the confidence interval does not cover zero). It may be, though, that this contrast is not one of those that has been estimated.

17.7 Bonferroni Multiple Comparison Procedure

The Bonferroni multiple comparison procedure was encountered earlier for regression models. It is also applicable for analysis of variance models when:

> The family of interest is a particular set of pairwise comparisons, contrasts, or linear combinations that is specified by the user in advance of the data analysis.

The Bonferroni procedure is applicable whether the factor level sample sizes are equal or unequal and whether inferences center on pairwise comparisons, contrasts, linear combinations, or a mixture of these.

Simultaneous Estimation

We shall denote the number of statements in the family by g and treat them all as linear combinations since pairwise comparisons and contrasts are special cases of linear combinations. The Bonferroni inequality (4.4) then implies that the confidence coefficient is at least $1 - \alpha$ that the following confidence limits for the g linear combinations L are all correct:

$$\hat{L} \pm Bs\{\hat{L}\} \tag{17.46}$$

where:

$$B = t(1 - \alpha/2g; n_T - r) \tag{17.46a}$$

Simultaneous Testing

When we wish to conduct a series of tests of the form:

$$H_0\colon L = 0$$
$$H_a\colon L \neq 0$$

we can use either the confidence intervals based on (17.46) or the test statistics:

$$t^* = \frac{\hat{L}}{s\{\hat{L}\}} \tag{17.47}$$

If $|t^*| \leq t(1 - \alpha/2g; n_T - r)$, we conclude H_0; otherwise, H_a is concluded.

Example

The sales manager of the Kenton Food Company is interested in estimating the following two contrasts with family confidence coefficient .975:

Comparison of 3-color and 5-color designs:

$$L_1 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

Comparison of designs with and without cartoons:

$$L_2 = \frac{\mu_1 + \mu_3}{2} - \frac{\mu_2 + \mu_4}{2}$$

Earlier we found:

$$\hat{L}_1 = -9.35 \qquad s\{\hat{L}_1\} = 1.50$$
$$\hat{L}_2 = -3.25 \qquad s\{\hat{L}_2\} = 1.50$$

For a 97.5 percent family confidence coefficient with the Bonferroni method, we require:

$$B = t[1 - .025/2(2); 15] = t(.99375; 15) = 2.84$$

We can now complete the confidence intervals for the two contrasts. For L_1, we have confidence limits $-9.35 \pm 2.84(1.50)$, which lead to the confidence interval:

$$-13.6 \leq L_1 \leq -5.1$$

Similarly, we obtain the other confidence interval:

$$-7.5 \leq L_2 \leq 1.0$$

These confidence intervals have a guaranteed family confidence coefficient of 97.5 percent, which means that in at least 97.5 percent of repetitions of the experiment, both intervals will be correct.

Again, we would conclude from this family of estimates that mean sales for 5-color designs are higher than those for 3-color designs (by somewhere between 5 and 14 cases per store), and that no overall effect of cartoons in the package design is indicated.

The Scheffé multiple for a 97.5 percent family confidence coefficient in this case would have been:

$$S^2 = 3F(.975; 3, 15) = 3(4.15) = 12.45$$

or $S = 3.53$, as compared to the Bonferroni multiple $B = 2.84$. Thus, the Scheffé procedure here would have led to wider confidence intervals than the Bonferroni procedure.

Comment

It is not necessary that all comparisons be estimated with statement confidence coefficients $1 - \alpha/g$ for the Bonferroni family confidence coefficient to be $1 - \alpha$. Different statement confidence coefficients may be used, depending upon the importance of each statement, provided that $\alpha_1 + \alpha_2 + \cdots + \alpha_g = \alpha$. ■

Comparison of Bonferroni Procedure with Scheffé and Tukey Procedures

1. If all pairwise comparisons are of interest, the Tukey procedure is superior to the Bonferroni procedure, leading to narrower confidence intervals. If not all pairwise comparisons are to be considered, the Bonferroni procedure may be the better one at times.

2. The Bonferroni procedure will be better than the Scheffé procedure when the number of contrasts of interest is about the same as the number of factor levels, or less. Indeed, the number of contrasts of interest must exceed the number of factor levels by a considerable amount before the Scheffé procedure becomes better.

3. All three procedures are of the form "estimator $\pm$ multiplier $\times$ SE." The only difference among the three procedures is the multiplier. In any given problem, one may compute the Bonferroni multiple as well as the Scheffé multiple and, when appropriate, the Tukey multiple, and select the one that is smallest. This choice is proper since it does not depend on the observed data.

4. The Bonferroni multiple comparison procedure does not lend itself to data snooping unless one can specify in advance the family of inferences in which one may be interested

and provided this family is not large. On the other hand, the Tukey and Scheffé procedures involve families of inferences that lend themselves naturally to data snooping.

5. Other specialized multiple comparison procedures have been developed. For example, Dunnett's procedure (Ref. 17.2) performs pairwise comparisons of each treatment against a control treatment only whereas Hsu's procedure (Ref. 17.3) selects the "best" treatment and identifies those treatments that are worse than the "best."

Analysis of Means

One use of the Bonferroni simultaneous testing procedure is in the analysis of means (ANOM), introduced by Ott (Ref. 17.4). ANOM is an alternative to the standard F test for the equality of treatment means. It is conducted by testing H_0: $\tau_1 = 0$ versus H_a: $\tau_1 \neq 0$, H_0: $\tau_2 = 0$ versus H_a: $\tau_2 \neq 0$, and so on for all treatment effects τ_i. The statistics employed are the r estimated treatment effects defined in (16.75b):

$$\hat{\tau}_i = \bar{Y}_{i\cdot} - \hat{\mu}_{\cdot} \qquad i = 1, \ldots, r \tag{17.48}$$

where $\hat{\mu}_{\cdot}$ is the least squares mean given in (16.75a):

$$\hat{\mu}_{\cdot} = \frac{\sum \bar{Y}_{i\cdot}}{r} \tag{17.48a}$$

The estimated variance of $\hat{\tau}_i$ is obtained by (17.22) since $\hat{\tau}_i$ is a contrast of the estimated treatment means $\bar{Y}_{i\cdot}$:

$$s^2\{\hat{\tau}_i\} = \frac{MSE}{n_i}\left(\frac{r-1}{r}\right)^2 + \frac{MSE}{r^2}\sum_{h \neq i}\frac{1}{n_h} \tag{17.49}$$

Simultaneous testing by the Bonferroni procedure can be carried out by setting up for each treatment effect the confidence interval using (17.46) and noting whether or not the interval contains zero. The results are sometimes summarized in an *analysis of means plot*. It is easy to show that a contrast $\hat{\tau}_i = \bar{Y}_{i\cdot} - \hat{\mu}_{\cdot}$ is inside (outside) one of the Bonferroni contrast intervals whenever the cell mean $\bar{Y}_{i\cdot}$ is inside (outside) the limits $\hat{\mu}_{\cdot} \pm t(1 - \alpha/2r; n_T - r)s\{\hat{\tau}_i\}$. In an analysis of means plot, the cell means are plotted along with the indicated limits and the least squares mean $\hat{\mu}_{\cdot}$ in (17.48a). If any of the cell means fall above (below) these limits, the conclusion is drawn that the cell mean is larger (smaller) than the overall mean.

ANOM is similar to ANOVA for detecting the differences between cell means. However, an important difference between ANOVA and ANOM is that the former tests whether the cell means are different from each other, whereas the latter tests whether the cell means are different from the overall mean. Various enhancements for the analysis of means have been provided, including those in References 17.5 and 17.6.

Example

In Figure 17.5 we present a MINITAB ANOM plot for the Kenton Food Company example using $\alpha = .05$. We conclude that the mean of sales for design 4 is greater than the overall unweighted mean (16.63), while the mean of sales for both design 1 and design 2 are less than the overall unweighted mean. Note that MINITAB bases its ANOM procedure on the weighted mean $\hat{\mu}_{\cdot} = \bar{Y}_{\cdot\cdot}$, rather than the least squares mean in (17.48a).

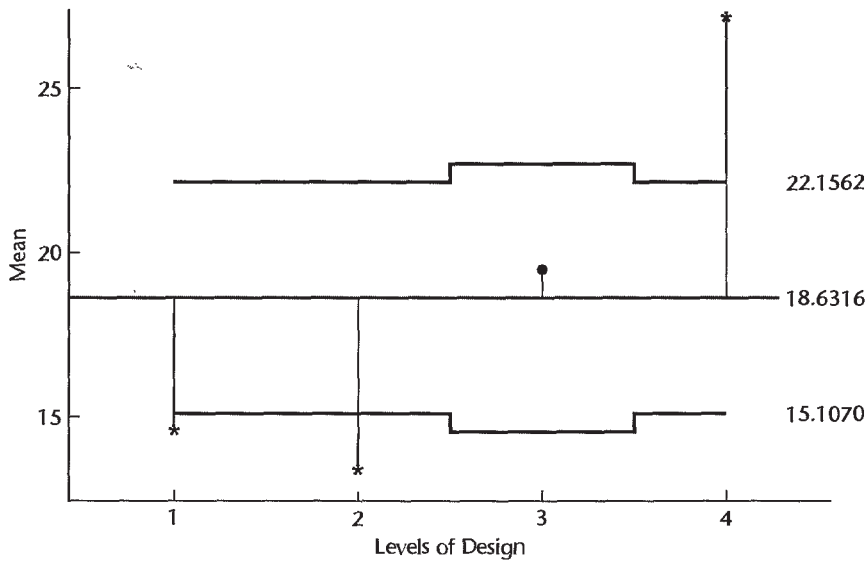


FIGURE 17.5 Analysis of Means Plot—Kenton Food Company Example.

17.8 Planning of Sample Sizes with Estimation Approach

In Section 16.10 we considered the planning of sample sizes using the power approach. We now take up another approach, the estimation approach to planning sample sizes, which may be used either in conjunction with the control of Type I and Type II errors or by itself. The essence of the approach is to specify the major comparisons of interest and to determine the expected widths of the confidence intervals for various sample sizes, given an advance planning value for the standard deviation σ. The approach is iterative, starting with an initial judgment of needed sample sizes. This initial judgment may be based on the needed sample sizes to control the risks of Type I and Type II errors when these have been obtained previously. If the anticipated widths of the confidence intervals based on the initial sample sizes are satisfactory, the iteration process is terminated. If one or more widths are too great, larger sample sizes need to be tried next. If the widths are narrower than they need be, smaller sample sizes should be tried next. This process is continued until those sample sizes are found that yield satisfactory anticipated widths for the important confidence intervals. We proceed to illustrate the estimation approach to planning sample sizes with two examples.

Example 1—Equal Sample Sizes

We are to plan sample sizes for the snow tires example discussed in Section 16.10 by means of the estimation approach; the sample sizes for each tire brand are to be equal, that is, $n_i \equiv n$. Management wishes three types of estimates:

1. A comparison of the mean tread lives for each pair of brands:

$$\mu_i - \mu_{i'}$$

2. A comparison of the mean tread lives for the two high-priced brands (1 and 4) and the two low-priced brands (2 and 3):

$$\frac{\mu_1 + \mu_4}{2} - \frac{\mu_2 + \mu_3}{2}$$

3. A comparison of the mean tread lives for the national brands (1, 2, and 4) and the local brand (3):

$$\frac{\mu_1 + \mu_2 + \mu_4}{3} - \mu_3$$

Management further has indicated that it wishes a family confidence coefficient of .95 for the entire set of comparisons.

We first need a planning value for the standard deviation of the tread lives of tires. Suppose that from past experience we judge the standard deviation to be approximately $\sigma = 2$ (thousand miles). Next, we require an initial judgment of needed sample sizes and shall consider $n = 10$ as a starting point.

We know from (17.21) that the variance of an estimated contrast $\hat{L}$ when $n_i \equiv n$ is:

$$\sigma^2\{\hat{L}\} = \frac{\sigma^2}{n} \sum c_i^2 \qquad \text{when } n_i \equiv n$$

Hence, given $\sigma = 2$ and $n = 10$, the anticipated values of the standard deviations of the required estimators are:

Contrast	Anticipated Variance	Anticipated Standard Deviation
Pairwise comparisons	$\frac{(2)^2}{10}[(1)^2 + (-1)^2] = .80$	.89
High- and low-priced brands	$\frac{(2)^2}{10}\left[\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2\right] = .40$	.63
National and local brands	$\frac{(2)^2}{10}\left[\left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + \left(\frac{1}{3}\right)^2 + (-1)^2\right] = .53$	.73

We shall employ the Scheffé multiple comparison procedure and therefore require the Scheffé multiple S in (17.43a) for $r = 4$, $n_T = 10(4) = 40$, and $1 - \alpha = .95$:

$$S^2 = (r - 1)F(1 - \alpha; r - 1, n_T - r) = 3F(.95; 3, 36) = 3(2.87) = 8.61$$

or $S = 2.93$. Hence, the anticipated widths of the confidence intervals are:

Contrast	Anticipated Width of Confidence Interval $= \pm S\sigma\{\hat{L}\}$
Pairwise comparisons	$\pm 2.93(.89) = \pm 2.61$ (thousand miles)
High- and low-priced brands	$\pm 2.93(.63) = \pm 1.85$ (thousand miles)
National and local brands	$\pm 2.93(.73) = \pm 2.14$ (thousand miles)

Management was satisfied with these anticipated widths. However, it was decided to increase the sample sizes from 10 to 15 in case the actual standard deviation of the tread lives of tires is somewhat greater than the anticipated value $\sigma = 2$ (thousand miles).

Example 2—Unequal Sample Sizes

In the snow tires example, suppose that tire brand 4 is the snow tire presently used and is to serve as the basis of comparison for the other brands. The comparisons of interest therefore are $\mu_1 - \mu_4$, $\mu_2 - \mu_4$, and $\mu_3 - \mu_4$. The sample size for brand 4 is to be twice as large as for the other brands in order to improve the precision of the three pairwise comparisons. The desired precision, with a family confidence coefficient of .90, is to be ± 1 (thousand miles). The Bonferroni procedure will be used to provide assurance as to the family confidence level.

We know from (17.13) that the variance of an estimated difference $\hat{L}_i = \bar{Y}_{i\cdot} - \bar{Y}_{4\cdot}$ (the difference is now denoted more generally by $\hat{L}$) is for $i = 1, 2, 3$:

$$\sigma^2\{\hat{L}_i\} = \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_4}\right)$$

We shall denote the sample sizes for brands 1, 2, and 3 by n and for brand 4 by $2n$. Hence, the variance of $\hat{L}_i$ becomes:

$$\sigma^2\{\hat{L}_i\} = \sigma^2\left(\frac{1}{n} + \frac{1}{2n}\right) = \frac{3\sigma^2}{2n}$$

Using again the planning value $\sigma = 2$ and an initial sample size $n = 10$, we find $\sigma^2\{\hat{L}_i\} = .60$ and $\sigma\{\hat{L}_i\} = .77$. For $\alpha = .10$ and $g = 3$ comparisons, the Bonferroni multiple is $B = t(.9833; 46) = 2.19$. Note that $n_T = 3(10) + 20 = 50$ for the first iteration; hence $n_T - r = 50 - 4 = 46$. The anticipated width of the confidence intervals therefore is $2.19(.77) = \pm 1.69$. This is larger than the specified width ± 1.0, so a larger sample size needs to be tried next.

We shall try $n = 30$ next. We find that $\sigma\{\hat{L}_i\} = .45$ now, and the Bonferroni multiple will be $B = t(.9833; 146) = 2.15$. Hence, the anticipated width of the confidence intervals for $n = 30$ is $2.15(.45) = \pm .97$. This is slightly smaller than the specified width ± 1.0. However, since the planning value for σ may not be entirely accurate, management may decide to use 30 tires for each of the new brands and 60 tires for brand 4, the presently used snow tires.

Comment

Since one cannot be certain that the planning value for the standard deviation is correct, it is advisable to study a range of values for the standard deviation before making a final decision on sample size. ■

17.9 Analysis of Factor Effects when Factor Is Quantitative

When the factor under investigation is quantitative, the analysis of factor effects can be carried beyond the point of multiple comparisons to include a study of the nature of the response function. Consider an experimental study undertaken to investigate the effect on sales of the price of a product. Five different price levels are investigated (78 cents, 79 cents, 85 cents, 88 cents, and 89 cents), and the experimental unit is a store. After a preliminary test of whether mean sales differ for the five price levels studied, the analyst might use multiple comparisons to examine whether "odd pricing" at 79 cents actually leads to higher sales than "even pricing" at 78 cents, as well as other questions of interest. In addition, the analyst may wish to study whether mean sales are a specified function of price, in the range of prices studied in the experiment. Further, once the relation has been established, the analyst may wish to use it for estimating sales volumes at various price levels not studied.

The methods of regression analysis discussed earlier are, of course, appropriate for the analysis of the response function. Since the single-factor studies discussed in this chapter almost always involve replications at the different factor levels, the lack of fit of a specified response function can be tested. For this purpose, the analysis of variance error sum of squares in (16.29) serves as the pure error sum of squares in (3.16), the two being identical. We illustrate this relation in the following example.

Example

In a study to reduce raw material costs in a glassworks firm, an operations analyst collected the experimental data in Table 17.4 on the number of acceptable units produced from equal amounts of raw material by 28 entry-level piecework employees who had received special training as part of the experiment. Four training levels were used (6, 8, 10, and 12 hours), with seven of the employees being assigned at random to each level. The higher the number of acceptable pieces, the more efficient is the employee in utilizing the raw material. This study is a single-factor completely randomized design with four factor levels.

Preliminary Analysis. The analyst first tested whether or not the mean number of acceptable pieces is the same for the four training levels. ANOVA model (17.1) was employed:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{17.50}$$

The alternative conclusions and appropriate test statistic are:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$

$$H_a: \text{not all } \mu_i \text{ are equal}$$

$$F^* = \frac{MSTR}{MSE}$$

TABLE 17.4 Data—Piecework Trainees Example.

Treatment (hours of training) i		Employee (j) 1	2	3	4	5	6	7
1	6 hours	40	39	39	36	42	43	41
2	8 hours	53	48	49	50	51	50	48
3	10 hours	53	58	56	59	53	59	58
4	12 hours	63	62	59	61	62	62	61

The SPSS^X output for single-factor ANOVA is shown in Figure 17.6. Residual analysis (to be discussed in Chapter 18) showed ANOVA model (17.50) to be apt. Therefore, the analyst proceeded with the test, using $\alpha = .05$. The decision rule is:

$$\text{If } F^* \leq F(.95; 3, 24) = 3.01, \text{ conclude } H_0$$
$$\text{If } F^* > 3.01, \text{ conclude } H_a$$

FIGURE 17.6 SPSS^X Computer Output—Piecework Trainees Example.

	GROUP	COUNT (n_i)	MEAN ($\bar{Y}_{i\cdot}$)	STANDARD DEVIATION
	GRP01	7	40.0000	2.3094
Treatment →	GRP02	7	49.8571	1.7728
	GRP03	7	56.5714	2.6367
	GRP04	7	61.4286	1.2724
	TOTAL	28	51.9643	8.4129

ANALYSIS OF VARIANCE

SOURCE	D F	SUM OF SQUARES	MEAN SQUARES
BETWEEN GROUPS	3	*SSTR* → 1808.6778	602.8926 ← *MSTR*
WITHIN GROUPS	24	*SSE* → 102.2856	4.2619 ← *MSE*
TOTAL	27	*SSTO* → 1910.9634	

F RATIO	F PROB.
141.461	0.0000
↑ *F**	↑ *P-value*

MULTIPLE RANGE TEST

TUKEY-HSD PROCEDURE
RANGES FOR THE 0.050 LEVEL -

3.90 ← ***q*(.95; 4, 24)**

HOMOGENEOUS SUBSETS

SUBSET 1		SUBSET 3	
GROUP	GRP01	GROUP	GRP03
MEAN	40.0000	MEAN	56.5714

SUBSET 2		SUBSET 4	
GROUP	GRP02	GROUP	GRP04
MEAN	49.8571	MEAN	61.4286

From Figure 17.6, we have:

$$F^* = \frac{MSTR}{MSE} = \frac{602.8926}{4.2619} = 141.5$$

Since $F^* = 141.5 > 3.01$, the analyst concluded H_a, that training level effects differed and that further analysis of them is warranted. The P-value for the test statistic is 0+, as shown in Figure 17.6.

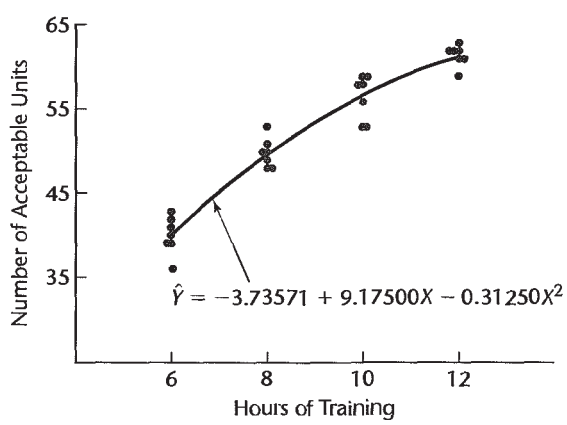
Investigation of Treatment Effects. The analyst's interest next centered on multiple comparisons of all pairs of treatment means. A Tukey multiple comparison option in the SPSS^X computer package was used. It gave the output shown in the lower portion of Figure 17.6. This output presents the results of single-degree-of-freedom tests conducted by means of the Tukey multiple comparison procedure for all pairwise comparisons. (The confidence intervals for the pairwise comparisons are not shown in the output.) All factor levels for which the test concludes that the pairwise means are equal are placed in the same group. This form of summary of single-degree-of-freedom tests was illustrated earlier for the Kenton Food Company example. When a group contains only one factor level, as is the case for all groups in the output of Figure 17.6, the implication is that all single-degree-of-freedom tests involving this factor level and each of the other factor levels lead to conclusion H_a, that the two factor level means being compared are not equal.

Two points should be noted in particular from the results in Figure 17.6: (1) All pairwise factor level differences are statistically significant. (2) There is some indication that differences between the means for adjoining factor levels diminish as the number of hours of training increases; that is, diminishing returns appear to set in as the length of training is increased.

Estimation of Response Function. These findings were in accord with the analyst's expectations that the treatment means μ_i would most likely follow a quadratic response function with respect to training level. The scatter plot in Figure 17.7 supports this expectation. The analyst now wished to investigate this point further by fitting a quadratic regression model. The model to be fitted and tested is:

$$Y_{ij} = \beta_0 + \beta_1 x_i + \beta_{11} x_i^2 + \varepsilon_{ij} \tag{17.51}$$

FIGURE 17.7 Scatter Plot and Fitted Quadratic Response Function—Piecework Trainees Example.



where Y_{ij} and ε_{ij} are defined as earlier, the βs are regression parameters, and x_i denotes the number of hours of training in the ith training level (X_i) centered around $\bar{X} = 9$, i.e., $x_i = X_i - 9$.

A portion of the data for the regression analysis is given in Table 17.5. Regressing Y on x and x^2 yielded the estimated regression function:

$$\hat{Y} = 53.52679 + 3.55000x - .31250x^2 \tag{17.52}$$

The analysis of variance for regression model (17.51) is shown in Table 17.6a. For completeness, we repeat in Table 17.6b the analysis of variance for ANOVA model (17.50).

TABLE 17.5 Illustration of Data for Regression Analysis—Piecework Trainees Example.

i	j	Y_{ij}	x_i	x_i^2
1	1	40	$6-9=-3$	9
1	2	39	$6-9=-3$	9
...	...	...	...	...
2	1	53	$8-9=-1$	1
2	2	48	$8-9=-1$	1
...	...	...	...	...
4	6	62	$12-9=3$	9
4	7	61	$12-9=3$	9

TABLE 17.6 Analyses of Variance—Piecework Trainees Example.

(a) Regression Model (17.51)

Source of Variation	*SS*	*df*	*MS*
Regression	1,808.100	2	904.05
Error	102.864	25	4.11
Total	1,910.964	27	

(b) Analysis of Variance Model (17.50)

Source of Variation	*SS*	*df*	*MS*
Treatments	1,808.678	3	602.89
Error	102.286	24	4.26
Total	1,910.964	27	

(c) ANOVA for Lack of Fit Test

Source of Variation	*SS*	*df*	*MS*
Regression	1,808.100	2	904.05
Error	102.864	25	4.11
Lack of fit	.578	1	.58
Pure error	102.286	24	4.26
Total	1,910.964	27	

Since the data contain replicates, the analyst could test regression model (17.51) for lack of fit, utilizing the fact that the ANOVA error sum of squares in (16.29) is identical to the regression pure error sum of squares in (3.16). Both measure variation around the mean of the Y observations at any given level of X (i.e., around the estimated treatment mean $\bar{Y}_{i\cdot}$). Hence, the lack of fit sum of squares can be readily obtained from previous results:

$$SSLF = \underset{\text{(Table 17.6a)}}{SSE} - \underset{\text{(Table 17.6b)}}{SSPE} = 102.864 - 102.286 = .578 \tag{17.53}$$

Since there are $c = r = 4$ levels of X here and $p = 3$ parameters in the regression model, $SSLF$ has associated with it $c - p = 4 - 3 = 1$ degree of freedom. Hence, we obtain $MSLF = .578/1 = .578$. Table 17.6c contains the analysis of variance for the regression model, with the error sum of squares and degrees of freedom broken down into lack of fit and pure error components.

The alternative conclusions (6.68a) for the test of lack of fit here are:

$$H_0\colon E\{Y\} = \beta_0 + \beta_1 x + \beta_{11} x^2$$
$$H_a\colon E\{Y\} \neq \beta_0 + \beta_1 x + \beta_{11} x^2$$

and test statistic (6.68b) is:

$$F^* = \frac{MSLF}{MSPE}$$

For $\alpha = .05$, decision rule (6.68c) becomes:

$$\text{If } F^* \leq F(.95; 1, 24) = 4.26, \text{ conclude } H_0$$
$$\text{If } F^* > 4.26, \text{ conclude } H_a$$

We calculate the test statistic from Table 17.6c:

$$F^* = \frac{.58}{4.26} = .136$$

Since $F^* = .136 \leq 4.26$, the analyst concluded that the quadratic response function is a good fit. Consequently, the fitted regression function in (17.52) was used in further evaluation of the relation between mean number of acceptable pieces produced and level of training, after expressing the fitted response function in the original predictor variable X (number of hours of training):

$$\hat{Y} = -3.73571 + 9.17500X - .31250X^2$$

Figure 17.7 displays this fitted response function.

Cited References

17.1. Cochran, W. G., and G. M. Cox. *Experimental Designs*. 2nd ed. New York: John Wiley & Sons, 1957, p. 74.

17.2. Dunnett, C. W. "A Multiple Comparison Procedure for Comparing Several Treatments with a Control." *Journal of the American Statistical Association* 50 (1955), pp. 1096–1121.

17.3. Hsu, J. C. *Multiple Comparisons: Theory and Methods*. London: Chapman & Hall, 1996.

17.4. Ott, E. R. "Analysis of Means—A Graphical Procedure." *Industrial Quality Control* 24 (1967), pp. 101–109.

17.5. Nelson, L. S. "Exact Critical Values for Use with the Analysis of Means," *Journal of Quality Technology* 15 (1983), pp. 40–44.

17.6. Nelson, P. R. "Additional Uses for the Analysis of Means and Extended Tables of Critical Values," *Technometrics* 35 (1993), pp. 61–71.

oblems

17.1. Refer to **Premium distribution** Problem 16.12. A student, asked to give a class demonstration of the use of a confidence interval for comparing two treatment means, proposed to construct a 99 percent confidence interval for the pairwise comparison $D = \mu_5 - \mu_3$. The student selected this particular comparison because the estimated treatment means $\bar{Y}_{5\cdot}$ and $\bar{Y}_{3\cdot}$ are the largest and smallest, respectively, and stated: "This confidence interval is particularly useful. If it does not straddle zero, it indicates, with significance level $\alpha = .01$, that the factor level means are not equal."

a. Explain why the student's assertion is not correct.

b. How should the confidence interval be constructed so that the assertion can be made with significance level $\alpha = .01$?

17.2. A trainee examined a set of experimental data to find comparisons that "look promising" and calculated a family of Bonferroni confidence intervals for these comparisons with a 90 percent family confidence coefficient. Upon being informed that the Bonferroni procedure is not applicable in this case because the comparisons had been suggested by the data, the trainee stated: "This makes no difference. I would use the same formulas for the point estimates and the estimated standard errors even if the comparisons were not suggested by the data." Respond.

17.3. Consider the following linear combinations of interest in a single-factor study involving four factor levels:

$$\text{(i)} \quad \mu_1 + 3\mu_2 - 4\mu_3$$

$$\text{(ii)} \quad .3\mu_1 + .5\mu_2 + .1\mu_3 + .1\mu_4$$

$$\text{(iii)} \quad \frac{\mu_1 + \mu_2 + \mu_3}{3} - \mu_4$$

a. Which of the linear combinations are contrasts? State the coefficients for each of the contrasts.

b. Give an unbiased estimator for each of the linear combinations. Also give the estimated variance of each estimator assuming that $n_i \equiv n$.

17.4. A single-factor ANOVA study consists of $r = 6$ treatments with sample sizes $n_i \equiv 10$.

a. Assuming that pairwise comparisons of the treatment means are to be made with a 90 percent family confidence coefficient, find the T, S, and B multiples for the following numbers of pairwise comparisons in the family: $g = 2, 5, 15$. What generalization is suggested by your results?

b. Assuming that contrasts of the treatment means are to be estimated with a 90 percent family confidence coefficient, find the S and B multiples for the following numbers of contrasts in the family: $g = 2, 5, 15$. What generalization is suggested by your results?

17.5. Consider a single-factor study with $r = 5$ treatments and sample sizes $n_i \equiv 5$.

a. Find the T, S, and B multiples if $g = 2, 5$, and 10 pairwise comparisons are to be made with a 95 percent family confidence coefficient. What generalization is suggested by your results?

b. What would be the T, S, and B multiples for sample sizes $n_i \equiv 20$? Does the generali obtained in part (a) still hold?

17.6. In making multiple comparisons, why is it appropriate to use the multiple comparison p dure that leads to the tightest confidence intervals for the sample data obtained? Discuss.

17.7. For a single-factor study with $r = 2$ treatments and sample sizes $n_i \equiv 10$, find the T, S, B multiples for $g = 1$ pairwise comparison with a 99 percent family confidence coeffici What generalization is suggested by your results?

*17.8. Refer to **Productivity improvement** Problem 16.7.

a. Prepare a line plot of the estimated factor level means $\bar{Y}_{i.}$. What does this plot suggest reg ing the effect of the level of research and development expenditures on mean productivi improvement?

b. Estimate the mean productivity improvement for firms with high research and developme expenditures levels; use a 95 percent confidence interval.

c. Obtain a 95 percent confidence interval for $D = \mu_2 - \mu_1$. Interpret your interval estimat

d. Obtain confidence intervals for all pairwise comparisons of the treatment means; use th Tukey procedure and a 90 percent family confidence coefficient. State your findings an prepare a graphic summary by underlining nonsignificant comparisons in your line plot in part (a).

e. Is the Tukey procedure employed in part (d) the most efficient one that could be used here? Explain.

17.9. Refer to **Questionnaire color** Problem 16.8.

a. Prepare a bar-interval graph of the estimated factor level means $\bar{Y}_{i.}$, where the interval correspond to the confidence limits in (17.7) with $\alpha = .05$. What does this plot sugges about the effect of color on the response rate? Is your conclusion in accord with the test result in Problem 16.8c?

b. Estimate the mean response rate for blue questionnaires; use a 90 percent confidence interval.

c. Test whether or not $D = \mu_3 - \mu_2 = 0$; use $\alpha = .10$. State the alternatives, decision rule, an conclusion. In light of the result for the ANOVA test in Problem 16.8e, is your conclusion surprising? Explain.

17.10. Refer to **Rehabilitation therapy** Problem 16.9.

a. Prepare a line plot of the estimated factor level means $\bar{Y}_{i.}$. What does this plot suggest about the effect of prior physical fitness on the mean time required in therapy?

b. Estimate with a 99 percent confidence interval the mean number of days required in therapy for persons of average physical fitness.

c. Obtain confidence intervals for $D_1 = \mu_2 - \mu_3$ and $D_2 = \mu_1 - \mu_2$; use the Bonferroni procedure with a 95 percent family confidence coefficient. Interpret your results.

d. Would the Tukey procedure have been more efficient to use in part (c)? Explain.

e. If the researcher also wished to estimate $D_3 = \mu_1 - \mu_3$, still with a 95 percent family confidence coefficient, would the B multiple in part (c) need to be modified? Would this also be the case if the Tukey procedure had been employed?

f. Test for all pairs of factor level means whether or not they differ; use the Tukey procedure with $\alpha = .05$. Set up groups of factor levels whose means do not differ.

*17.11. Refer to **Cash offers** Problem 16.10.

a. Prepare a main effects plot of the estimated factor level means $\bar{Y}_{i.}$. What does this plot suggest regarding the effect of the owner's age on the mean cash offer?

b. Estimate the mean cash offer for young owners; use a 99 percent confidence interval.

c. Construct a 99 percent confidence interval for $D = \mu_3 - \mu_1$. Interpret your interval estimate.

d. Test whether or not $\mu_2 - \mu_1 = \mu_3 - \mu_2$; control the α risk at .01. State the alternatives, decision rule, and conclusion.

e. Obtain confidence intervals for all pairwise comparisons between the treatment means; use the Tukey procedure and a 90 percent family confidence coefficient. Interpret your results and provide a graphic summary by preparing a paired comparison plot. Are your conclusions in accord with those in part (a)?

f. Would the Bonferroni procedure have been more efficient to use in part (e) than the Tukey procedure? Explain.

*17.12. Refer to **Filling machines** Problem 16.11.

a. Prepare a main effects plot of the estimated factor level means $\bar{Y}_{i\cdot}$. What does this plot suggest regarding the variation in the mean fills for the six machines?

b. Construct a 95 percent confidence interval for the mean fill for machine 1.

c. Obtain a 95 percent confidence interval for $D = \mu_2 - \mu_1$. Interpret your interval estimate.

d. Prepare a paired comparison plot and interpret it.

e. The consultant is particularly interested in comparing the mean fills for machines 1, 4, and 5. Use the Bonferroni testing procedure for all pairwise comparisons among these three treatment means with family level of significance $\alpha = .10$. Interpret your results and provide a graphic summary by preparing a line plot of the estimated factor level means with nonsignificant differences underlined. Do your conclusions agree with those in part (a)?

f. Would the Tukey testing procedure have been more efficient to use in part (e) than the Bonferroni testing procedure? Explain.

17.13. Refer to **Premium distribution** Problem 16.12.

a. Prepare an interval plot of the estimated factor level means $\bar{Y}_{i\cdot}$, where the intervals correspond to the confidence limits in (17.7) with $\alpha = .10$. What does this plot suggest about the variation in the mean time lapses for the five agents?

b. Test for all pairs of factor level means whether or not they differ; use the Tukey procedure with $\alpha = .10$. Set up groups of factor levels whose means do not differ. Use a paired comparison plot to summarize the results.

c. Construct a 90 percent confidence interval for the mean time lapse for agent 1.

d. Obtain a 90 percent confidence interval for $D = \mu_2 - \mu_1$. Interpret your interval estimate.

e. The marketing director wishes to compare the mean time lapses for agents 1, 3, and 5. Obtain confidence intervals for all pairwise comparisons among these three treatment means; use the Bonferroni procedure with a 90 percent family confidence coefficient. Interpret your results and present a graphic summary by preparing a line plot of the estimated factor level means with nonsignificant differences underlined. Do your conclusions agree with those in part (a)?

f. Would the Tukey procedure have been more efficient to use in part (e) than the Bonferroni procedure? Explain.

*17.14. Refer to **Productivity improvement** Problem 16.7.

a. Estimate the difference in mean productivity improvement between firms with low or moderate research and development expenditures and firms with high expenditures; use a 95 percent confidence interval. Employ an unweighted mean for the low and moderate expenditures groups. Interpret your interval estimate.

b. The sample sizes for the three factor levels are proportional to the population sizes. The economist wishes to estimate the mean productivity gain last year for all firms in the

population. Estimate this overall mean productivity improvement with a 95 percent confidence interval.

c. Using the Scheffé procedure, obtain confidence intervals for the following comparisons with 90 percent family confidence coefficient:

$$D_1 = \mu_3 - \mu_2 \qquad D_3 = \mu_2 - \mu_1$$
$$D_2 = \mu_3 - \mu_1 \qquad L_1 = \frac{\mu_1 + \mu_2}{2} \quad \mu_3$$

Interpret your results and describe your findings.

17.15. Refer to **Rehabilitation therapy** Problem 16.9.

a. Estimate the contrast $L = (\mu_1 - \mu_2) - (\mu_2 - \mu_3)$ with a 99 percent confidence interval. Interpret your interval estimate.

b. Estimate the following comparisons using the Bonferroni procedure with a 95 percent family confidence coefficient:

$$D_1 = \mu_1 - \mu_2 \qquad D_3 = \mu_2 - \mu_3$$
$$D_2 = \mu_1 - \mu_3 \qquad L_1 = D_1 - D_3$$

Interpret your results and describe your findings.

c. Would the Scheffé procedure have been more efficient to use in part (b) than the Bonferroni procedure? Explain.

*17.16. Refer to **Cash offers** Problem 16.10.

a. Estimate the contrast $L = (\mu_3 - \mu_2) - (\mu_2 - \mu_1)$ with a 99 percent confidence interval. Interpret your interval estimate.

b. Estimate the following comparisons with a 90 percent family confidence coefficient; employ the most efficient multiple comparison procedure:

$$D_1 = \mu_2 - \mu_1 \qquad D_3 = \mu_3 - \mu_1$$
$$D_2 = \mu_3 - \mu_2 \qquad L_1 = D_2 - D_1$$

Interpret your results.

*17.17. Refer to **Filling machines** Problem 16.11. Machines 1 and 2 were purchased new five years ago, machines 3 and 4 were purchased in a reconditioned state five years ago, and machines 5 and 6 were purchased new last year.

a. Estimate the contrast:

$$L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

with a 95 percent confidence interval. Interpret your interval estimate.

b. Estimate the following comparisons with a 90 percent family confidence coefficient; use the most efficient multiple comparison procedure:

$$D_1 = \mu_1 - \mu_2 \qquad L_1 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$
$$D_2 = \mu_3 - \mu_4 \qquad L_2 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_5 + \mu_6}{2}$$
$$D_3 = \mu_5 - \mu_6 \qquad L_3 = \frac{\mu_1 + \mu_2 + \mu_5 + \mu_6}{4} - \frac{\mu_3 + \mu_4}{2}$$
$$L_4 = \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4}{4} - \frac{\mu_5 + \mu_6}{2}$$

Interpret your results. What can the consultant learn from these results about the differences between the six filling machines?

17.18. Refer to **Premium distribution** Problem 16.12. Agents 1 and 2 distribute merchandise only, agents 3 and 4 distribute cash-value coupons only, and agent 5 distributes both merchandise and coupons.

a. Estimate the contrast:

$$L = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

with a 90 percent confidence interval. Interpret your interval estimate.

b. Estimate the following comparisons with 90 percent family confidence coefficient; use the Scheffé procedure:

$$D_1 = \mu_1 - \mu_2 \qquad L_1 = \frac{\mu_1 + \mu_2}{2} - \mu_5$$

$$D_2 = \mu_3 - \mu_4 \qquad L_2 = \frac{\mu_3 + \mu_4}{2} - \mu_5$$

$$L_3 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

Interpret your results.

c. Of all premium distributions, 25 percent are handled by agent 1, 20 percent by agent 2, 20 percent by agent 3, 20 percent by agent 4, and 15 percent by agent 5. Estimate the overall mean time lapse for premium distributions with a 90 percent confidence interval.

*17.19. Refer to **Filling machines** Problem 16.11.

a. Use the analysis of means procedure to test for equality of treatment effects, with family significance level .05. Which treatments have the strongest effects?

b. Using the results in part (a), obtain the analysis of means plot. What additional information does this plot provide in comparison with the main effects plot in Problem 17.12a?

17.20. Refer to **Premium distribution** Problem 16.12.

a. Use the analysis of means procedure to test for equality of treatment effects, with family significance level .10. Which treatments have the strongest effects?

b. Using the results in part (a), obtain the analysis of means plot. What additional information does this plot provide in comparison with the interval plot in Problem 17.13a?

17.21. Refer to **Solution concentration** Problem 3.15. Suppose the chemist initially wishes to employ ANOVA model (16.2) to determine whether or not the concentration of the solution is affected by the amount of time that has elapsed since preparation.

a. State the analysis of variance model.

b. Prepare a main effects plot of the estimated factor level means $\bar{Y}_{i.}$. What does this plot suggest about the relation between the solution concentration and time?

c. Obtain the analysis of variance table.

d. Test whether or not the factor level means are equal; use $\alpha = .025$. State the alternatives, decision rule, and conclusion.

e. Make pairwise comparisons of factor level means between all adjacent lengths of time; use the Bonferroni procedure with a 95 percent family confidence coefficient. Are your conclusions in accord with those in part (b)? Do your results suggest that the regression relation is not linear?

17.22. A market researcher stated in a seminar: "The power approach to determining sample sizes for analysis of variance problems is not meaningful; only the estimation approach should be used. We never conduct a study where all treatment means are expected to be equal, so we are always interested in a variety of estimates." Discuss.

17.23. Refer to **Questionnaire color** Problem 16.8. Suppose estimates of all pairwise comparisons are of primary importance. What would be the required sample sizes if the precision of all pairwise comparisons is to be ± 3.0, using the Tukey procedure with a 95 percent family confidence coefficient?

17.24. Refer to **Rehabilitation therapy** Problem 16.9. Suppose primary interest is in estimating the two pairwise comparisons:

$$L_1 = \mu_1 - \mu_2 \qquad L_2 = \mu_3 - \mu_2$$

What would be the required sample sizes if the precision of each comparison is to be ± 3.0 days, using the most efficient multiple comparison procedure with a 95 percent family confidence coefficient?

*17.25. Refer to **Filling machines** Problem 16.11. Suppose primary interest is in estimating the following comparisons:

$$L_1 = \mu_1 - \mu_2 \qquad L_3 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

$$L_2 = \mu_3 - \mu_4 \qquad L_4 = \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4}{4} - \frac{\mu_5 + \mu_6}{2}$$

What would be the required sample sizes if the precision of each of these comparisons is not to exceed $\pm .08$ ounce, using the best multiple comparison procedure with a 95 percent family confidence coefficient?

17.26. Refer to **Premium distribution** Problem 16.12. Suppose primary interest is in estimating the following comparisons:

$$L_1 = \mu_1 - \mu_2 \qquad L_3 = \frac{\mu_1 + \mu_2}{2} - \mu_5$$

$$L_2 = \mu_3 - \mu_4 \qquad L_4 = \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4}{2}$$

What would be the required sample sizes if the precision of each of the estimated comparisons is not to exceed ± 1.0 day, using the most efficient multiple comparison procedure with a 90 percent family confidence coefficient?

17.27. Refer to **Rehabilitation therapy** Problem 16.9. Suppose that primary interest is in comparing the below-average and above-average physical fitness groups, respectively, with the average physical fitness group. Thus, two comparisons are of interest:

$$L_1 = \mu_1 - \mu_2 \qquad L_2 = \mu_3 - \mu_2$$

Assume that a reasonable planning value for the error standard deviation is $\sigma = 4.5$ days.

a. It has been decided to use equal sample sizes (n) for the below-average and above-average groups. If twice this sample size ($2n$) were to be used for the average physical fitness group, what would be the required sample sizes if the precision of each pairwise comparison is to be ± 2.5 days, using the Bonferroni procedure and a 90 percent family confidence coefficient?

b. Repeat the calculations in part (a) if the sample size for the average physical fitness group is to be: (1) n and (2) $3n$, all other specifications remaining the same.

c. Compare your results in parts (a) and (b). Which design leads to the smallest total sample size here?

17.28. Refer to **Rehabilitation therapy** Problem 16.9. A biometrician has developed a scale for physical fitness status, as follows:

Physical Fitness Status	Scale Value
Below average	83
Average	100
Above average	121

a. Using this physical fitness status scale, fit first-order regression model (1.1) for regressing number of days required for therapy (Y) on physical fitness status (X).

b. Obtain the residuals and plot them against X. Does a linear regression model appear to fit the data?

c. Perform an F test to determine whether or not there is lack of fit of a linear regression function; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

d. Could you test for lack of fit of a quadratic regression function here? Explain.

*17.29. Refer to **Filling machines** Problem 16.11. A maintenance engineer has suggested that the differences in mean fills for the six machines are largely related to the length of time since a machine last received major servicing. Service records indicate these lengths of time to be as follows (in months):

Filling Machine	Number of Months	Filling Machine	Number of Months
1	.4	4	5.3
2	3.7	5	1.4
3	6.1	6	2.1

a. Fit second-order polynomial regression model (8.2) for regressing amount of fill (Y) on number of months since major servicing (X).

b. Obtain the residuals and plot them against X. Does a quadratic regression function appear to fit the data?

c. Perform an F test to determine whether or not there is lack of fit of a quadratic regression function; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

d. Test whether or not the quadratic term in the response function can be dropped from the model; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

Exercises

17.30. Show that when $r = 2$ and $n_i \equiv n$, q defined in (17.35) is equivalent to $\sqrt{2}|t^*|$, where t^* is defined in (A.65) in Appendix A.

17.31. Starting with (17.38), complete the derivation of (17.30).

17.32. Show that when $r = 2$, S^2 defined in (17.43a) is equivalent to $[t(1 - \alpha/2; n_T - r)]^2$.

17.33. Show that the estimated variance of $\hat{\tau}_i$ in (17.48) is given by (17.49).

17.34. (Calculus needed.) Refer to **Rehabilitation therapy** Problem 16.9. The sample sizes for the below-average, average, and above-average physical fitness groups are to be n, kn, and n, respectively. Assuming that ANOVA model (16.2) is appropriate, find the optimal value of k to minimize the variances of $\hat{L}_1 = \bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}$ and $\hat{L}_2 = \bar{Y}_{3\cdot} - \bar{Y}_{2\cdot}$ for a given total sample size n_T.

Projects

17.35. Refer to the **SENIC** data set in Appendix C.1 and Project 16.42. Obtain confidence intervals for all pairwise comparisons between the four regions; use the Tukey procedure and a 90 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means and underline all nonsignificant comparisons.

17.36. Refer to the **CDI** data set in Appendix C.2 and Project 16.44. Obtain confidence intervals for all pairwise comparisons between the four regions; use the Tukey procedure and a 90 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means and underline all nonsignificant comparisons.

17.37. Refer to the **Market share** data set in Appendix C.3 and Project 16.45. Obtain confidence intervals for all pairwise comparisons among the four factor levels; use the Tukey procedure and a 95 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means, underscoring all nonsignificant comparisons.

17.38. Refer to Project 16.46e.

a. For each replication, construct confidence intervals for all pairwise comparisons among the three treatment means; use the Tukey procedure with a 95 percent family confidence coefficient. Then determine whether all confidence intervals for the replication are correct, given that $\mu_1 = 80$, $\mu_2 = 60$, and $\mu_3 = 160$.

b. For what proportion of the 100 replications are all confidence intervals correct? Is this proportion close to theoretical expectations? Discuss.

Case Studies

17.39. Refer to the **Prostate cancer** data set in Appendix C.5 and Case Study 16.49. Obtain confidence intervals for all pairwise comparisons among the three Gleason score levels; use the Tukey procedure and a 95 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means, underscoring all nonsignificant comparisons.

17.40. Refer to the **Real estate sales** data set in Appendix C.7 and Case Study 16.50. Obtain confidence intervals for all pairwise comparisons among the four number-of-bedroom categories; use the Tukey procedure and a 90 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means, underscoring all nonsignificant comparisons.

17.41. Refer to the **Ischemic heart disease** data set in Appendix C.9 and Case Study 16.51. Obtain confidence intervals for all pairwise comparisons among the six number-of-intervention categories; use the Tukey procedure and a 90 percent family confidence coefficient. Interpret your results and state your findings. Prepare a line plot of the estimated factor level means, underscoring all nonsignificant comparisons.

Chapter 18

ANOVA Diagnostics and Remedial Measures

When discussing regression analysis, we emphasized the importance of examining the appropriateness of the regression model under consideration, and noted the effectiveness of residual plots and other diagnostics for spotting major departures from the tentative model. Examination of the appropriateness of analysis of variance models is no less important.

In this chapter, we take up the use of residual plots for diagnosing the appropriateness of analysis of variance models, as well as formal tests for the constancy of the error variance. We also discuss the use of transformations of the response variable as a remedial measure to improve the appropriateness of the analysis of variance model for estimation and test inferences.

For pedagogic reasons, as in regression analysis, we have discussed inference procedures before diagnostics and remedial measures. The actual sequence of developing and using any statistical model is, of course, the reverse:

1. Examine whether the proposed model is appropriate for the set of data at hand.
2. If the proposed model is not appropriate, consider remedial measures, such as transformation of the data or modification of the model.
3. After review of the appropriateness of the model and completion of any necessary remedial measures and an evaluation of their effectiveness, inferences based on the model can be undertaken.

It is not necessary, nor is it usually possible, that an ANOVA model fit the data perfectly. As will be noted later, ANOVA models are reasonably robust against certain types of departures from the model, such as the error terms not being exactly normally distributed. The major purpose of the examination of the appropriateness of the model is therefore to detect serious departures from the conditions assumed by the model.

18.1 Residual Analysis

Residual analysis for ANOVA models corresponds closely to that for regression models. We therefore discuss only briefly some key issues in the use of residual analysis for ANOVA models.

Residuals

The residuals e_{ij} for the ANOVA cell means model (16.2) were defined in (16.20):

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i\cdot} \tag{18.1}$$

As in regression, semistudentized residuals, studentized residuals, and studentized deleted residuals are often helpful for diagnosing ANOVA model departures. The definitions of these residuals for regression in Chapters 3 and 10 are still applicable for ANOVA models. However, in view of the simple nature of the **X** matrix for ANOVA models, the regression formulas often simplify here. The semistudentized residuals e_{ij}^* in (3.5) for regression remain unchanged:

$$e_{ij}^* = \frac{e_{ij}}{\sqrt{MSE}} \tag{18.2}$$

The studentized residuals r_{ij} in (10.20) become here:

$$r_{ij} = \frac{e_{ij}}{s\{e_{ij}\}} \tag{18.3}$$

where:

$$s\{e_{ij}\} = \sqrt{\frac{MSE(n_i - 1)}{n_i}} \tag{18.3a}$$

Finally, the studentized deleted residuals t_{ij} in (10.26) become here:

$$t_{ij} = e_{ij}\left[\frac{n_T - r - 1}{SSE\left(1 - \frac{1}{n_i}\right) - e_{ij}^2}\right]^{1/2} \tag{18.4}$$

Comment

For ANOVA model (16.2), it can be shown that the leverage of Y_{ij}, defined in (10.18), is given by:

$$h_{ij,ij} = \frac{1}{n_i} \tag{18.5}$$

Hence, the variance of the residual e_{ij} for ANOVA model (16.2) can be obtained by substituting (18.5) into (10.14):

$$\sigma^2\{e_{ij}\} = \frac{\sigma^2(n_i - 1)}{n_i} \tag{18.6}$$

Replacing σ^2 by the unbiased estimator MSE and taking the square root lead to the estimated standard deviation $s\{e_{ij}\}$ in (18.3a).

When the treatment sample sizes n_i are the same, the leverages of all the observations Y_{ij} are the same. As a result, the estimated standard deviations of the residuals, $s\{e_{ij}\}$, are all the same so that the semistudentized residuals e_{ij}^* and the studentized residuals r_{ij} provide essentially the same information, differing only by a constant factor near 1 unless the treatment sample size is very small. ■

Residual Plots

Residual plots useful for analysis of variance models include: (1) plots against the fitted values, (2) time or other sequence plots, (3) dot plots, and (4) normal probability plots. All of these plots have been encountered previously. We therefore proceed directly to an

example to illustrate the use of residual plots for evaluating the appropriateness of analysis of variance models.

xample

Table 18.1 contains a portion of the residuals for the rust inhibitor example of Chapter 17. For ease of presentation, the treatments are shown in the columns of the table. The residuals were obtained from the data in Table 17.2a. For instance, the residual for the first experimental unit treated with brand A rust inhibitor is:

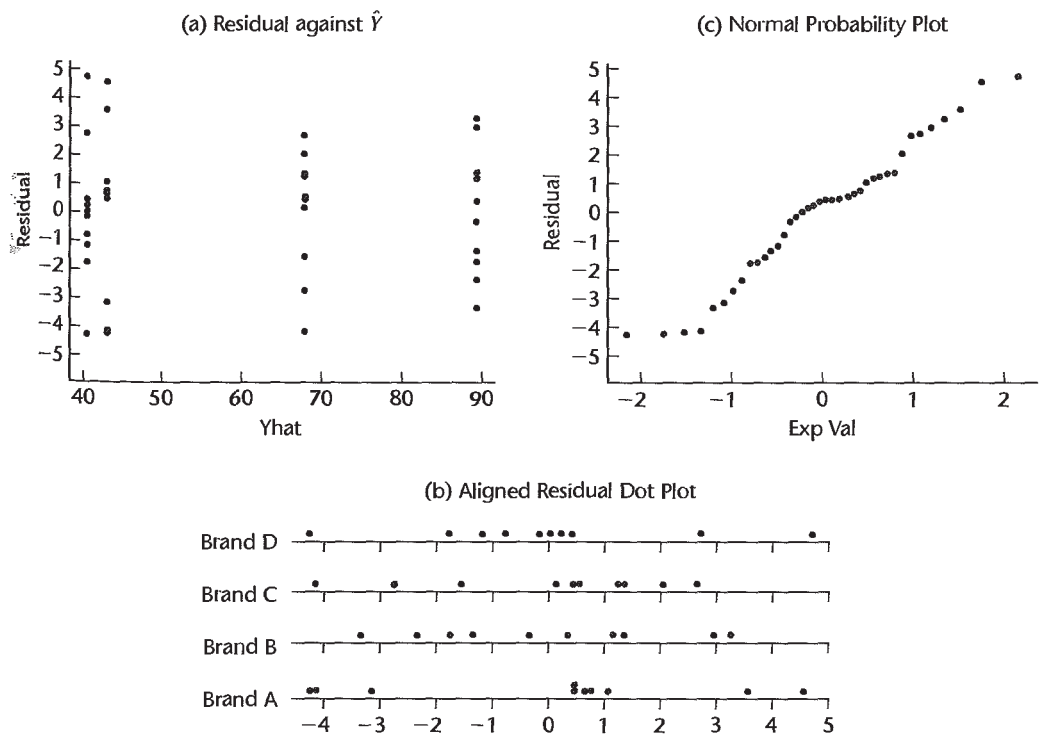
$$e_{11} = Y_{11} - \hat{Y}_{11} = Y_{11} - \bar{Y}_{1\cdot} = 43.9 - 43.14 = .76$$

Figure 18.1 presents three MINITAB diagnostic residual plots. Figure 18.1a contains a *residual plot against the fitted values*. This plot differs in appearance from similar plots for

TABLE 18.1
esiduals—
Rust Inhibitor
xample.

	Brand			
	A	B	C	D
j	$i = 1$	$i = 2$	$i = 3$	$i = 4$
1	.76	.36	.45	−4.27
2	−4.14	−2.34	1.35	4.73
3	3.56	3.26	.55	.23
...	...	...	...	...
8	−4.24	−1.34	−2.75	−1.77
9	.46	1.36	−4.15	.43
10	−3.14	−.34	1.25	−.77

FIGURE 18.1 MINITAB Diagnostic Residual Plots—Rust Inhibitor Example.



regression analysis because the fitted values $\hat{Y}_{ij}$ here are the same for all observations for a given factor level. Recall from (16.17) that $\hat{Y}_{ij} = \bar{Y}_{i.}$.

Figure 18.1b contains *aligned dot plots* of the residuals for each factor level. These plots are similar to the residual plot against the fitted values in Figure 18.1a, except here the residual scale is the horizontal one. An advantage of the plot in Figure 18.1a is that it facilitates an assessment of the relation between the magnitudes of the error variances and the factor level means. A disadvantage is that some of the estimated factor level means may be far apart, making a comparison of the factor levels more difficult. This difficulty is remedied in Figure 18.1b since dot plots can be placed close together to facilitate comparisons between factor levels.

Figure 18.1c contains a *normal probability plot* of the residuals. This plot is exactly the same as for regression models.

No sequence plot of the residuals is presented here because the data for the rust inhibitor example were not ordered according to time or in some other logical sequence.

All of the plots in Figure 18.1, as we shall see, suggest that ANOVA model (16.2) is appropriate for the rust inhibitor data.

Diagnosis of Departures from ANOVA Model

We consider now how residual plots can be helpful in diagnosing the following departures from ANOVA model (16.2):

1. Nonconstancy of error variance
2. Nonindependence of error terms
3. Outliers
4. Omission of important explanatory variables
5. Nonnormality of error terms

Nonconstancy of Error Variance. ANOVA model (16.2) requires that the error terms ε_{ij} have constant variance for all factor levels. When the sample sizes are not large and do not differ greatly, the appropriateness of this assumption can be studied by using the residuals, semistudentized residuals, or studentized residuals. *Plots of residuals against fitted values* or *dot plots of residuals* are helpful. When the sample sizes differ greatly, studentized residuals should be used in these plots. Constancy of the error variance is shown in these plots by the plots having about the same extent of scatter of the residuals around zero for each factor level. This is the case for the rust inhibitor example in Figures 18.1a and 18.1b.

Figure 18.2 is a prototype residual plot against the fitted values when the error variances are not constant. This plot portrays the case where the error terms for factor level 3 have a larger variance than those for the other two factor levels.

When the sample sizes for the different factor levels are large, *histograms* or boxplots of the residuals for each treatment—arranged vertically and using the same scale, like the dot plots in Figure 18.1b—are an effective means for examining the constancy of the error variance, as well as for assessing whether the error terms are normally distributed.

A number of statistical tests have been developed for formally examining the equality of the r factor level variances; two of these tests will be discussed in Section 18.2.

Nonindependence of Error Terms. Whenever data are obtained in a time sequence, a *residual sequence plot* should be prepared to examine if the error terms are serially

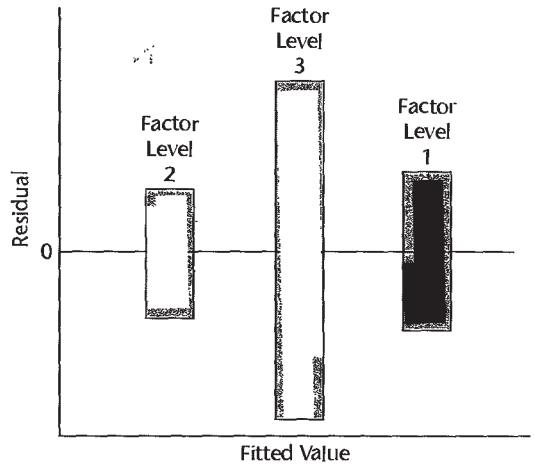


FIGURE 18.2 Prototype Plot of Residuals Against Fitted Values When a Factor Term Variance Is Not Constant for Different Factor Levels.

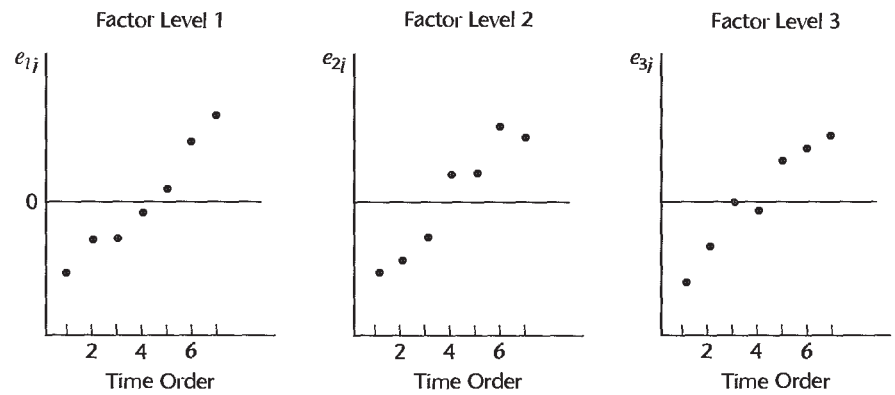


FIGURE 18.3 Residual Sequence Plots for Group Interaction Study Illustrating Time-Related Effect.

correlated. Figure 18.3 contains the residuals for an experiment on group interactions. Three different treatments were applied, and the group interactions were recorded on videotapes. Seven replications were made for each treatment. Afterward, the experimenter measured the number of interactions by viewing the tapes in randomized order. Figure 18.3 strongly suggests that the experimenter discerned a larger number of interactions as more experience in viewing the tapes was gained. As a result, the residuals in Figure 18.3 appear to be serially correlated. In this instance, an inclusion in the model of a linear term for the time effect might be sufficient to assure independence of the error terms in the revised model.

Time-related effects may also lead to increases or decreases in the error variance over time. For instance, an experimenter may make more precise measurements over time. Figure 18.4 portrays residual sequence plots where the error variance decreases over time.

When the data are ordered in some other logical sequence, such as in a geographic sequence, a plot of the residuals against this ordering is helpful for ascertaining whether the error terms are serially correlated according to this ordering.

Outliers. The detection of outliers is facilitated by various plots of the studentized deleted residuals. *Residual plots against fitted values, residual dot plots, box plots,* and *stem-and-*

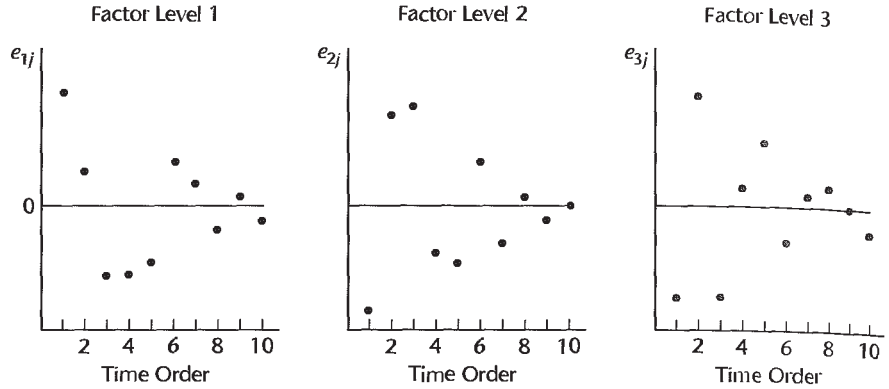


FIGURE 18.4 Residual Sequence Plots Illustrating Decreasing Error Variance over Time.

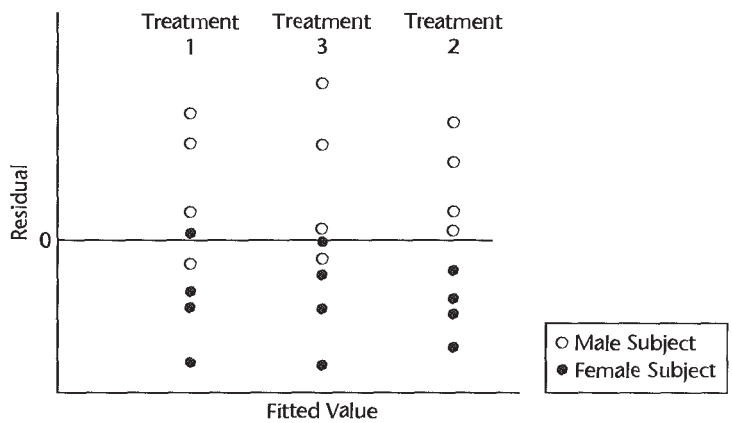


FIGURE 18.5 Residual Plot against Fitted Values Illustrating Omission of Important Explanatory Variable.

leaf plots are particularly helpful. These plots easily reveal outlying observations, that is, observations that differ from the fitted value by far more than do other observations. As noted in Chapter 3, it is wise practice to discard outlying observations only if they can be identified as being due to such specific causes as instrumentation malfunctioning, observer measurement blunder, or recording error.

The test for outliers in regression discussed in Chapter 10 is applicable to analysis of variance as well. The appropriate Bonferroni critical value here is $t(1-\alpha/2n_T; n_T-r-1)$. If the largest absolute studentized deleted residual exceeds this critical value, that case should be considered an outlier. Note that the implicit family of tests here consists of the tests on all n_T residuals for the study since we do not know in advance which case will have the largest absolute studentized deleted residual.

Occasionally, a test for an outlier is suggested in advance of the analysis, as when a substitute operator is used for one of the production runs in a manufacturing experiment. Concern about the validity of this response observation might lead to an outlier test. In this case, the Bonferroni critical value would be $t(1-\alpha/2; n_T-r-1)$.

Omission of Important Explanatory Variables. Residual analysis may also be used to study whether or not the single-factor ANOVA model is an adequate model. In a learning experiment involving three motivational treatments, the residuals shown in Figure 18.5 were obtained. The residual plot against the fitted values in Figure 18.5 shows no unusual

overall pattern. The experimenter wondered, however, whether the treatment effects differ according to the gender of the subject. In Figure 18.5 the residuals for male subjects are shown by open circles, and those for females by dots. The results in Figure 18.5 suggest strongly that for each of the motivational treatments studied, the treatment effects do differ according to gender. Here, an analysis of covariance model, recognizing both motivational treatment and gender of subject as explanatory variables as mentioned in Chapter 15, might be more useful. Analysis of covariance models will be discussed in Chapter 22.

Note that residual analysis here does not invalidate the original single-factor model. Rather, the residual analysis points out that the original model overlooks differences in treatment effects that may be important to recognize. Since there are usually many explanatory variables that have some effect on the response, the analyst needs to identify for residual analysis those explanatory variables that most likely have an important effect on the response.

Nonnormality of Error Terms. The normality of the error terms can be studied from *histograms, dot plots, box plots,* and *normal probability plots* of the residuals. In addition, comparisons can be made of observed frequencies with expected frequencies if normality holds, and formal chi-square goodness of fit or related tests can be utilized. The discussion in Chapter 3 about these methods for assessing the normality of the error terms for regression is entirely applicable to ANOVA models.

When the factor level sample sizes are large, the study of normality can be made separately for each treatment. When the factor level sample sizes are small, one can combine the residuals e_{ij} for all treatments into one group, provided that the evidence suggests that there are no major differences in the error variances for the treatments studied. This combining was done in the rust inhibitor example in Figure 18.1c. This figure does not indicate any serious departures from normality. The pattern of the points is reasonably linear except possibly in the tails. The coefficient of correlation between the ordered residuals and their expected values under normality is .987, which also supports the reasonableness of the normality assumption.

When unequal variances of the error terms for the different factor levels are indicated and normality must be examined for the combined data, studentized residuals (18.3) should be used, with MSE replaced by the sample variance s_i^2 in (16.39) for observations from the ith treatment. If ordinary residuals were used, nonnormality might be indicated solely because of the failure of the error terms to have equal variances.

Comment

As for regression models, the ANOVA residuals e_{ij} are not independent random variables. For ANOVA model (16.2), they are subject to the restrictions in (16.21). Consequently, statistical tests that require independent observations are not exactly appropriate for ANOVA residuals. If, however, the number of residuals for each factor level is not small, the effect of the correlations will only be modest. It has been noted that graphic plots of residuals are less subject to the effects of correlation than are statistical tests because graphic plots contain the individual residuals and not simply functions of them. ■

18.2 Tests for Constancy of Error Variance

Several formal tests are available for studying the constancy of the error variance, as required by the ANOVA model. We shall consider two of these, the Hartley test (Ref. 18.1) and the Brown-Forsythe test (Ref. 18.2). Both tests assume that independent random samples are

obtained from each population. The Hartley test is simple to carry out, but is applicable only if the sample sizes are equal and if the error terms are normally distributed. The test is designed to be sensitive to substantial differences between the largest and the smallest factor level variances. The Brown-Forsythe test, discussed in Chapter 3, is slightly more difficult to compute but is more generally applicable. The test has been shown to be robust to departures from normality, and sample sizes need not be equal.

Both the Hartley test and the Brown-Forsythe test are often conducted at low α levels when used for testing the constancy of the error variance in the analysis of variance. The reason is that, as we shall note in Section 18.6, the F test for equality of factor level means is robust against nonconstancy of the error variance when the factor level sample sizes are approximately equal, as long as the differences in the variances are not extremely large. Hence, the purpose of using the Hartley or Brown-Forsythe tests in ANOVA is often to determine whether extremely large differences in the error variances exist. For this purpose, a low α level may be employed since only large differences in the error variances need to be detected.

Hartley Test

We shall describe the Hartley test in general terms. The test considers r normal populations; the variance of the ith population is denoted by σ_i^2. Independent samples of equal size are selected from the r populations; the sample variance for the ith population is denoted by s_i^2 and the common number of degrees of freedom associated with each sample variance is denoted by df. The alternatives to be tested are:

$$\begin{aligned} &H_0\colon \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_r^2 \\ &H_a\colon \text{not all } \sigma_i^2 \text{ are equal} \end{aligned} \tag{18.7}$$

The Hartley test statistic, denoted by H^*, is based solely on the largest sample variance, denoted by $\max(s_i^2)$, and the smallest sample variance, denoted by $\min(s_i^2)$:

$$H^* = \frac{\max\left(s_i^2\right)}{\min\left(s_i^2\right)} \tag{18.8}$$

Values of H^* near 1 support H_0, and large values of H^* support H_a. The distribution of H^* when H_0 holds has been tabulated, and selected percentiles are presented in Table B.10. The distribution of H^* depends on the number of populations r and the common number of degrees of freedom df.

The appropriate decision rule for controlling the risk of making a Type I error at α is:

$$\begin{aligned} &\text{If } H^* \le H(1-\alpha; r, df), \text{ conclude } H_0 \\ &\text{If } H^* > H(1-\alpha; r, df), \text{ conclude } H_a \end{aligned} \tag{18.9}$$

where $H(1-\alpha; r, df)$ is the $(1-\alpha)100$ percentile of the distribution of H^* when H_0 holds, for r populations and df degrees of freedom for each sample variance.

When the Hartley test is used for the single-factor ANOVA model (16.2) with equal sample sizes, $n_i \equiv n$, we have $df = n - 1$. The r normal populations are the normal probability distributions of the Y observations for the r factor levels. The sample variance

s_i^2 is the variance of the n_i observations Y_{ij} for the ith factor level or equivalently the variance of the n_i residuals e_{ij}, defined in (16.39); for $n_i \equiv n$, s_i^2 becomes:

$$s_i^2 = \frac{\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_{i\cdot})^2}{n-1} = \frac{\sum_{j=1}^{n} e_{ij}^2}{n-1} \qquad \text{when } n_i \equiv n \tag{18.10}$$

Example

The ABT Electronics Corporation performed an experiment to evaluate five types of flux for use in soldering printed circuit boards. A major concern of the firm's reliability engineers was the strength of the soldered joints. To test the five types of flux, 40 printed circuit boards were selected at random. Each of the five flux types was randomly assigned to 8 of the 40 circuit boards and an electronic switch was soldered to each board using the designated flux type. Following a four-week storage period, the 40 circuit boards were tested by an hydraulically operated testing machine which exerted increasing pulling force on each switch. The force (in pounds) required to break a joint, termed the pull strength, is the response of interest. This design is a completely randomized design, with eight replicates of the five treatments corresponding to the five levels of the categorical factor, flux type.

A portion of the observed pull strengths in the experiment is shown in Table 18.2, along with the estimated treatment means $\bar{Y}_{i\cdot}$ and sample variances s_i^2. A dot plot of these data is presented in Figure 18.6. Notice that the variability in pull strengths for the third solder type appears to be larger than for the others.

Since approximate normality is required by the Hartley test, normal probability plots of the residuals were first constructed for each treatment (not shown). The approximate normality of the residuals for each treatment was supported by the plots and by the correlation test (the correlations in the five plots are .982, .981, .977, .958, and .939; the critical value for $\alpha = .05$ from Table B.6 is .906).

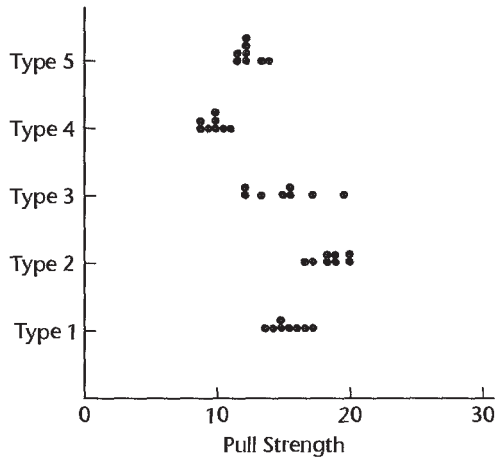
The alternatives for the Hartley test here are:

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_5^2$$
$$H_a: \text{not all } \sigma_i^2 \text{ are equal}$$

TABLE 18.2 Solder Joint Pull Strengths—ABT Electronics Example.

Joint	Flux Type (i)				
j	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
1	14.87	18.43	16.95	8.59	11.55
2	16.81	18.76	12.28	10.90	13.36
...	...	...	...	...	...
7	17.40	17.16	19.35	9.41	12.05
8	14.62	16.40	15.52	10.04	11.95
$\bar{Y}_{i\cdot}$	15.420	18.528	15.004	9.741	12.340
$\tilde{Y}_i$	15.170	18.595	15.255	10.010	12.105
s_i^2	1.531	1.570	6.183	.667	.592

FIGURE 18.6 Dot Plots of Pull Strengths—ABT Electronics Example.



For level of significance $\alpha = .05$, $r = 5$, and $df = 8 - 1 = 7$, we require $H(.95; 5, 7) = 9.70$. Hence the appropriate decision rule is:

$$\text{If } H^* \leq 9.70, \text{ conclude } H_0$$
$$\text{If } H^* > 9.70, \text{ conclude } H_a$$

From Table 18.2 we see that $\max(s_i^2) = 6.183$ and $\min(s_i^2) = .592$. Hence the test statistic is:

$$H^* = \frac{6.183}{.592} = 10.44$$

Since $H^* = 10.44 > 9.70$, we conclude H_a, that the five treatment variances are not equal.

Comments

1. The Hartley test strictly requires equal sample sizes. If the sample sizes are unequal but do not differ greatly, the Hartley test may still be used as an approximate test. For this purpose, the average number of degrees of freedom would be used for entering Table B.10.

2. The Hartley test is quite sensitive to departures from the assumption of normal populations and should not be used when substantial departures from normality exist. ■

Brown-Forsythe Test

The Brown-Forsythe test for constancy of the error variance in regression was discussed in Chapter 3. The test was originally developed for use in ANOVA applications and is more general than its use for regression described in Chapter 3. The Brown-Forsythe test, just like the Hartley test, can be used to study the equality of r population variances. Unlike the Hartley test, the Brown-Forsythe test is robust against departures from normality, which often occur together with unequal variances. Also, the Brown-Forsythe test does not require equal sample sizes.

To test the alternatives in (18.7) using the Brown-Forsythe test, we first compute the absolute deviations of the Y_{ij} observations about their respective factor level medians $\tilde{Y}_i$:

$$d_{ij} = |Y_{ij} - \tilde{Y}_i| \tag{18.11}$$

The Brown-Forsythe test then determines whether or not the expected values of the absolute deviations for the r treatments are equal. If the r error variances σ_i^2 are equal, so will the expected values of the absolute deviations be equal. Unequal error variances imply differing expected values of the absolute deviations. The Brown-Forsythe test statistic is simply the ordinary F^* statistic in (16.55) for testing differences in the treatment means, but now based on the absolute deviations d_{ij} in (18.11):

$$F_{BF}^* = \frac{MSTR}{MSE} \tag{18.12}$$

where:

$$MSTR = \frac{\sum n_i(\bar{d}_{i\cdot} - \bar{d}_{\cdot\cdot})^2}{r-1} \tag{18.12a}$$

$$MSE = \frac{\sum\sum(d_{ij} - \bar{d}_{i\cdot})^2}{n_T - r} \tag{18.12b}$$

$$\bar{d}_{i\cdot} = \frac{\sum_j d_{ij}}{n_i} \tag{18.12c}$$

$$\bar{d}_{\cdot\cdot} = \frac{\sum\sum d_{ij}}{n_T} \tag{18.12d}$$

If the error terms have constant variance and the factor level sample sizes are not extremely small, F_{BF}^* follows approximately an F distribution with $r-1$ and $n_T - r$ degrees of freedom. Large F_{BF}^* values indicate that the error terms do not have constant variance.

Example

Table 18.2 for the ABT Electronics Corporation example provides the sample medians $\tilde{Y}_i$ for the five treatments. The absolute deviations d_{ij} in (18.11) are shown in Table 18.3. We illustrate their calculation for d_{11}:

$$d_{11} = |Y_{11} - \tilde{Y}_1| = |14.87 - 15.170| = .300$$

The F_{BF}^* test statistic (18.12) based on the absolute deviations is obtained in the usual manner; it is $F_{BF}^* = 2.94$. For $\alpha = .05$, we require $F(.95; 4, 35) = 2.64$. Since $F_{BF}^* = 2.94 > 2.64$, we conclude H_a, that the error terms do not have constant variance. The P-value for this test is .034.

TABLE 18.3 Absolute Deviations of Responses from Treatment Medians—ABT Electronics Example.

Joint	Flux Type (i)				
j	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
1	.300	.165	1.695	1.420	.555
2	1.640	.165	2.975	.890	1.255
...	...	...	...	...	...
7	2.230	1.435	4.095	.600	.055
8	.550	2.195	.265	.030	.155

18.3 Overview of Remedial Measures

In the remainder of this chapter, we consider three remedial measures for two common departures from ANOVA model (16.2)—nonconstancy of the error variance and nonnormality of the distribution of the error terms.

1. If the error terms are normally distributed but the variance of the error terms is not constant, a standard remedial measure is to use weighted least squares. We have already considered weighted least squares for nonconstancy of the error variance in regression models. These weighted least squares procedures for regression carry over directly to analysis of variance models.

2. Often, nonconstancy of the error variance is accompanied by nonnormality of the error term distribution. A standard remedial measure here is to transform the response variable Y. We shall present two approaches to finding an appropriate transformation to make the error distribution more nearly normal and to help stabilize the variance of the error terms—some simple guides and the Box-Cox procedure. The latter was considered in Chapter 3 for regression models and is directly applicable to analysis of variance models.

3. When there are major departures from ANOVA model (16.2) and transformations are not successful in stabilizing the error variance and bringing the error distribution close to normal, a nonparametric test for the equality of the factor level means may be used instead of the standard F test. We shall consider a nonparametric test that is based on the ranks of the Y observations.

We begin our discussion of remedial measures with weighted least squares.

18.4 Weighted Least Squares

When the errors ε_{ij} are normally distributed but their variances are not the same for the different factor levels, cell means model (16.2) becomes:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{18.13}$$

where ε_{ij} are independent $N(0, \sigma_i^2)$.

Weighted least squares is a standard remedial measure here, just as for the comparable situation in regression. In fact, we shall use the regression approach to the analysis of variance for implementing weighted least squares. All of the earlier discussion on weighted least squares for regression is applicable to the analysis of variance.

Since the factor level variances σ_i^2 are usually unknown, they must be estimated. This is ordinarily done by means of the sample variances s_i^2 in (16.39), in which case the weight w_{ij} for the jth case of the ith factor level is:

$$w_{ij} = \frac{1}{s_i^2} \tag{18.14}$$

The test for the equality of the factor level means in (16.54) is now conducted by the general linear test approach described in Chapter 2. The full model is fitted, using the weights in (18.14), and the error sum of squares is obtained, now denoted by $SSE_w(F)$. Next, the reduced model under H_0 is fitted and the error sum of squares $SSE_w(R)$ is obtained. Test

statistic (2.70) is employed, as usual. We shall see that $df_F = n_T - r$ and $df_R = n_T - 1$. Hence, the general linear test statistic here is:

$$F_w^* = \frac{SSE_w(R) - SSE_w(F)}{r-1} \div \frac{SSE_w(F)}{n_T - r} \tag{18.15}$$

Since the weights are based on the estimated variances s_i^2, the distribution of F_w^* under H_0 is only approximately an F distribution with $r-1$ and $n_T - r$ degrees of freedom. When the factor level sample sizes are reasonably large, the approximation generally is satisfactory. As explained in Chapter 11, bootstrapping can be employed to examine the effect of using estimated weights.

Example

Recall in the ABT Electronics example that the normality assumption appears to be reasonably well supported by the data, but the error variance is not constant. Weighted least squares will now be used to test the alternatives:

$$\begin{aligned} &H_0\colon \mu_1 = \mu_2 = \cdots = \mu_5 \\ &H_a\colon \text{not all } \mu_i \text{ are equal} \end{aligned} \tag{18.16}$$

The weights will be based on the sample variances in Table 18.2:

$$w_{1j} = \frac{1}{1.531} = .653 \qquad w_{2j} = \frac{1}{1.570} = .637 \qquad w_{3j} = \frac{1}{6.183} = .162$$

$$w_{4j} = \frac{1}{.667} = 1.499 \qquad w_{5j} = \frac{1}{.592} = 1.689$$

We shall use regression model (16.85) to represent cell means model (18.13):

$$Y_{ij} = \mu_1 X_{ij1} + \mu_2 X_{ij2} + \cdots + \mu_5 X_{ij5} + \varepsilon_{ij} \qquad \text{Full model} \tag{18.17}$$

where:

$$X_1 = \begin{cases} 1 & \text{if case from factor level 1} \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$X_5 = \begin{cases} 1 & \text{if case from factor level 5} \\ 0 & \text{otherwise} \end{cases}$$

Note that the factor level means μ_i play the role of regression coefficients and that the regression model has no intercept.

Table 18.4 repeats from Table 18.2 a portion of the experimental data in column 1 and contains the coding of the indicator variables in columns 2–6 and the weights in column 7. Note, for instance, that the coding for cases from the first treatment is $X_1 = 1$, $X_2 = 0$, $X_3 = 0$, $X_4 = 0$, and $X_5 = 0$, and similarly for cases from the other treatments.

Figure 18.7a contains the MINITAB output when Y in column 1 of Table 18.4 is regressed on X_1, X_2, X_3, X_4, and X_5 in columns 2–6, using the weights in column 7 and specifying no intercept. We see that $SSE_w(F) = 35.0$.

The reduced model under H_0 is given by (16.86):

$$Y_{ij} = \mu_c + \varepsilon_{ij} \qquad \text{Reduced model} \tag{18.18}$$

TABLE 18.4 Data for Weighted Least Squares Regression—ABT Electronics Examp

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			Full Model					Weights	Reduced M
i	j	Y_{ij}	X_{ij1}	X_{ij2}	X_{ij3}	X_{ij4}	X_{ij5}	w_{ij}	X_{ij}
1	1	14.87	1	0	0	0	0	.653	1
1	2	16.81	1	0	0	0	0	.653	1
...	...	...	...	...	...	...	...	...	...
1	7	17.40	1	0	0	0	0	.653	1
1	8	14.62	1	0	0	0	0	.653	1
2	1	18.43	0	1	0	0	0	.637	1
2	2	18.76	0	1	0	0	0	.637	1
...	...	...	...	...	...	...	...	...	...
5	7	12.05	0	0	0	0	1	1.689	1
5	8	11.95	0	0	0	0	1	1.689	1

FIGURE 18.7 MINITAB Weighted Regression Output for Full and Reduced Models—ABT Electronics Example.

(a) Full Model

The regression equation is
Y = 15.4 X1 + 18.5 X2 + 15.0 X3 + 9.74 X4 + 12.3 X5

Predictor	Coef	Stdev	t-ratio	p
Noconstant				
X1	15.4200	0.4375	35.24	0.000
X2	18.5275	0.4430	41.82	0.000
X3	15.0037	0.8785	17.08	0.000
X4	9.7413	0.2888	33.73	0.000
X5	12.3400	0.2721	45.36	0.000

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	5	6478.7	1295.7	1295.56	0.000
Error	35	35.0	1.0		
Total	40	6513.7			

(b) Reduced Model

The regression equation is
Y = 12.9 X

Predictor	Coef	Stdev	t-ratio	p
Noconstant				
X	12.8764	0.4981	25.85	0.000

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	6154.5	6154.5	668.28	0.000
Error	39	359.2	9.2		
Total	40	6513.7			

where μ_c is the common mean response under H_0. The corresponding regression model is:

$$Y_{ij} = \mu_c X_{ij} + \varepsilon_{ij} \tag{18.19}$$

where $X_{ij} \equiv 1$. Note that regression model (18.19) has no intercept.

The new X variable is shown in Table 18.4, column 8. Regressing Y in column 1 on X in column 8, using the weights in column 7 and specifying no intercept, leads to the MINITAB output in Figure 18.7b. We see that $SSE_w(R) = 359.2$. We have $n_T - 1 = 40 - 1 = 39$ and $n_T - r = 40 - 5 = 35$. Hence, test statistic (18.15) is:

$$F_w^* = \frac{359.2 - 35.0}{39 - 35} \div \frac{35.0}{35} = 81.05$$

For $\alpha = .01$, we require $F(.99; 4, 35) = 3.908$. Since $F^* = 81.05 > 3.908$, the approximate F test leads to conclusion H_a, that the factor level means differ. The approximate P-value of the test is 0+.

Comments

1. The weighted least squares estimates of the factor level means μ_i are always the estimated factor level means $\bar{Y}_{i\cdot}$, as may be seen by comparing the estimated regression coefficients in Figure 18.7a with the estimated factor level means in Table 18.2. Hence, for ANOVA model (18.13), the weighted and ordinary least squares estimates of the factor level means μ_i are the same.

2. When the sample variances s_i^2 are used as weights, the error sum of squares for the fit of full model (18.17) will always be $SSE_w(F) = n_T - r$. Note that in our example $SSE_w(F) = 35.0$ and $n_T - r = 40 - 5 = 35$.

3. Some analysis of variance computer packages have an option for weighted least squares, with the user specifying the weights. ■

18.5 Transformations of Response Variable

When both the model assumptions of constancy of the error variance and normality of the error distributions are violated, a transformation of the response variable is often useful. We describe now two approaches to finding a useful transformation—some simple guides and the Box-Cox procedure.

Simple Guides to Finding a Transformation

The following are four simple guides to finding a useful transformation. The guides were developed from theoretical considerations to stabilize the error variances, but these transformations often also are helpful in bringing the distribution of the error terms more closely to a normal distribution.

Variance Proportional to μ_i. When the variance of the error terms for each factor level (denoted by σ_i^2) is proportional to the factor level mean μ_i, a square root transformation is helpful:

$$\text{If } \sigma_i^2 \text{ proportional to } \mu_i: \qquad Y' = \sqrt{Y} \qquad \text{or} \qquad Y' = \sqrt{Y} + \sqrt{Y+1} \tag{18.20}$$

This type of situation is often found when the observed variable Y is a count, such as the number of attempts by a subject before the correct solution is found.

Standard Deviation Proportional to μ_i. When the standard deviation of the error terms for each factor level is proportional to the factor level mean, a helpful transformation is the logarithmic transformation:

$$\text{If } \sigma_i \text{ proportional to } \mu_i: \qquad Y' = \log Y \tag{18.21}$$

Standard Deviation Proportional to μ_i^2. When the error term standard deviation is proportional to the square of the factor level mean for the different factor levels, an appropriate transformation is the reciprocal transformation:

$$\text{If } \sigma_i \text{ proportional to } \mu_i^2: \qquad Y' = \frac{1}{Y} \tag{18.22}$$

Response Is a Proportion. At times, the observed variable Y_{ij} is a proportion p_{ij}. For instance, the treatments may be different training procedures, the unit of observation is a company training class, and the observed variable Y_{ij} is the proportion of employees in the jth class for the ith training procedure who benefited substantially by the training. Note that n_i here refers to the number of classes receiving the ith training procedure, not to the number of students.

It is well known that for the binomial distribution the variance of the sample proportion depends on the true proportion. When the number of cases on which each sample proportion is based is the same, this variance is:

$$\sigma^2\{p_{ij}\} = \frac{\pi_i(1-\pi_i)}{m} \tag{18.23}$$

Here π_i denotes the population proportion for the ith treatment and m is the common number of cases on which each sample proportion is based. Since $\sigma^2\{p_{ij}\}$ depends on the treatment proportion π_i, the variances of the error terms will not be stable if the treatment proportions π_i differ. An appropriate transformation for this case is the arcsine transformation:

$$\text{If response is a proportion:} \qquad Y' = 2 \arcsin \sqrt{Y} \tag{18.24}$$

When the proportions p_{ij} are based on different numbers of cases (for instance, in our earlier illustration there may be different numbers of employees in each training class), transformation (18.24) should be employed together with a weighted least squares analysis as described in Section 18.4. The use of the arcsin transformation when the response is a proportion can be an effective, yet simple, remedial measure. A more rigorous approach would involve the use of logistic regression as discussed in Chapter 14.

Use of Simple Guides. To examine whether one of the simple transformation guides is applicable, the statistics $s_i^2/\bar{Y}_{i\cdot}$, $s_i/\bar{Y}_{i\cdot}$, and $s_i/\bar{Y}_{i\cdot}^2$ should be calculated for each factor level, where s_i^2 is the sample variance of the Y observations for the ith factor level, defined in (16.39). Approximate constancy of one of the three statistics over all factor levels would suggest the corresponding transformation as useful for stabilizing the error variance and making the error distributions more nearly normal.

xample

Servo-Data, Inc., operates mainframe computers at three different locations. The computers are identical as to make and model, but are subject to different degrees of voltage fluctuation

LE 18.5
e between
puter
ures at
tions (in
)—
o-Data
ple.

Failure Interval	Location (i)					
	1		2		3	
j	Y_{1j}	R_{1j}	Y_{2j}	R_{2j}	Y_{3j}	R_{3j}
1	4.41	2	8.24	4	106.19	14
2	100.65	13	81.16	11	33.83	7
3	14.45	6	7.35	3	78.88	10
4	47.13	9	12.29	5	342.81	15
5	85.21	12	1.61	1	44.33	8

i	$\bar{Y}_{i\cdot}$	s_i^2	i	$\bar{R}_{i\cdot}$	s_i^2
1	50.4	1,789	1	8.4	20.3
2	22.1	1,103	2	4.8	14.2
3	121.2	16,167	3	10.8	12.7
	$\bar{Y}_{\cdot\cdot} = 64.6$			$\bar{R}_{\cdot\cdot} = 8.00$	

in the power lines serving the respective installations. Table 18.5 contains the lengths of time between computer failures Y_{ij} for the three locations, for five failure intervals each. The table also contains the ranks R_{ij} (from 1 to 15) for Y_{ij}, which we shall use in Section 18.7 for nonparametric analysis. Even though the sample sizes are small, the data suggest highly skewed distributions having nonconstant error variance. This is an observational study because no randomization of treatments to experimental units occurred.

To study whether one of the simple guides is helpful here, we have calculated the following statistics based on the results in Table 18.5.

i	$\dfrac{s_i^2}{\bar{Y}_{i\cdot}}$	$\dfrac{s_i}{\bar{Y}_{i\cdot}}$	$\dfrac{s_i}{\bar{Y}_{i\cdot}^2}$
1	35.5	.84	.017
2	49.9	1.50	.068
3	133.4	1.05	.009

The relation $s_i/\bar{Y}_{i\cdot}$ is the most stable, hence the logarithmic transformation (18.21) may be helpful here. We shall continue this example after discussing the use of the Box-Cox procedure for finding an appropriate transformation in the analysis of variance.

Box-Cox Procedure

The Box-Cox transformation procedure was described in Chapter 3 for regression. As noted there, the Box-Cox procedure identifies a power transformation of the type Y^λ to correct for both lack of normality and nonconstancy of the error variance. The procedure is entirely applicable to the analysis of variance. As for regression, the numerical search procedure for ANOVA models considers different values of the parameter λ. For each value of λ, the Y observations are transformed according to (3.36) and ANOVA model (16.2) is fitted and the

error sum of squares *SSE* is obtained. The value of λ that minimizes *SSE* is the maximum likelihood estimate of λ. As we saw in regression, *SSE* as a function of λ is often flat in the neighborhood of the maximum likelihood estimate $\hat{\lambda}$, so that a meaningful value of λ in the neighborhood may be chosen for the transformation in preference to the maximum likelihood value.

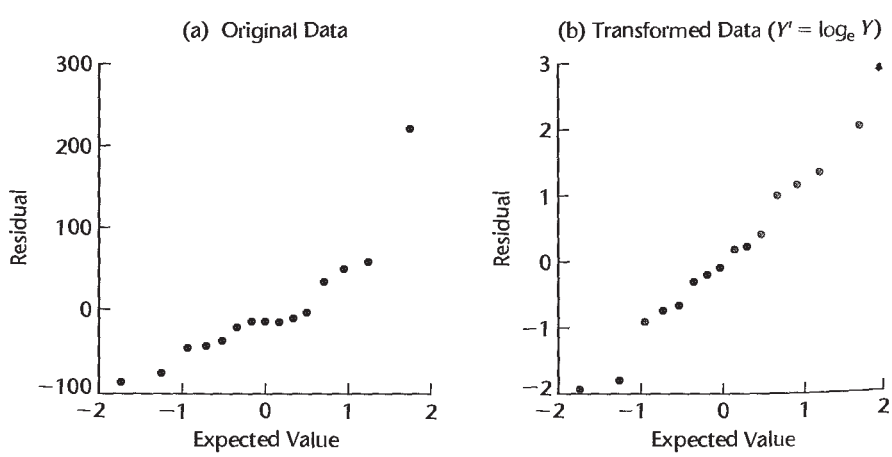
Example

The Box-Cox procedure was applied in the Servo-Data example of Table 18.5 by using 21 equally spaced values of λ between -1 and 1. For each value of λ, the Y observations were transformed according to (3.36) and *SSE* for ANOVA model (16.2) was calculated. A portion of the results is shown in Table 18.6. The smallest *SSE* is obtained with $\lambda = .1$. However, note that *SSE* does not change much between $-.10$ and .20. Hence, the parameter $\lambda = 0$ may be preferred because it leads to the meaningful logarithmic transformation. This is also the transformation selected according to the simple guides. Normal probability plots of the residuals for the original and transformed data ($Y' = \log_e Y$) are shown in Figure 18.8. The normality assumption appears to be much more reasonable for the transformed data ($r = .991$). Also, the variances of the transformed data are much more stable now ($s_1^2 = 1.742$, $s_2^2 = 1.974$, $s_3^2 = .817$) as compared to the variances for the original data in Table 18.5.

TABLE 18.6 **Calculations for Box-Cox Procedure—Servo-Data Example.**

λ	*SSE* (in thousands)	λ	*SSE* (in thousands)
−1.0	203.7	.10	15.3
−.80	95.1	.20	15.6
−.60	48.7	.40	18.7
−.40	28.3	.60	26.4
−.20	19.2	.80	42.6
−.10	17.0	1.0	76.2
.00	15.7		

FIGURE 18.8 **Normal Probability Plots for Original and Transformed Data—Servo-Data Example.**



A single factor ANOVA was performed on Y', the logarithm of the Y observations. The resulting F test for equality of treatment means was:

$$F^* = \frac{MSTR}{MSE} = \frac{5.7264}{1.5112} = 3.789$$

For $\alpha = .10$, we require $F(.90; 2, 12) = 2.81$. Since $F^* = 3.789 > 2.81$, we conclude H_a, that the three means are not equal. The P-value of the test is .053. The transformed means for the three groups are 3.413, 2.797, and 4.437, respectively. The Bonferroni pairwise comparison procedure was then conducted at the .10 level, with $s^2\{\hat{D}\} = .6045$, $s\{\hat{D}\} = .7775$, $B = t(.9833; 12) = 2.402$, and $Bs\{\hat{D}\} = 1.868$. The resulting 90 percent Bonferroni pairwise confidence intervals are:

$$-2.984 \leq \mu_2 - \mu_1 \leq .752$$
$$-.884 \leq \mu_3 - \mu_1 \leq 2.892$$
$$.272 \leq \mu_3 - \mu_2 \leq 4.008$$

Therefore, we conclude that location 3 has longer average time computer failures than location 2.

Comments

1. It is wise policy, as mentioned for regression, to check the residuals after a transformation has been applied to make sure that the transformation has been effective in both stabilizing the variances and making the distribution of the error terms reasonably normal.

2. When a transformation of the observations is required, one can work completely with the transformed data for testing the equality of factor level means. On the other hand, it is often desirable when making estimates of factor level effects to change a confidence interval based on the transformed variable back to an interval in the original variable for easier understanding of the significance of the results.

3. The variance stabilizing transformations (18.20), (18.21), (18.22), and (18.24) are obtained by using a Taylor series expansion for the variance of Y. An explanation of the approach may be found in Reference 18.3. ■

18.6 Effects of Departures from Model

In preceding sections, we considered how residual analysis and other diagnostic techniques can be helpful in assessing the appropriateness of the ANOVA model for the data at hand. We also discussed the use of transformations for both stabilizing the variance and obtaining an error distribution more nearly normal. The question now arises: what are the effects of any remaining departures from the model on the inferences made? A thorough review of the many studies investigating these effects has been made by Scheffé (Ref. 18.4). Here, we summarize the findings.

Nonnormality

For the fixed ANOVA model I, lack of normality is not an important matter, provided the departure from normality is not extreme. It may be noted in this connection that *kurtosis* of the error distribution (either more or less peaked than a normal distribution) is more important than skewness of the distribution in terms of the effects on inferences.

The point estimators of factor level means and contrasts are unbiased whether or not the populations are normal. The F test for the equality of factor level means is but little affected by lack of normality, either in terms of the level of significance or power of the test. Hence, the F test is a *robust* test against departures from normality. For instance, while the specified level of significance might be .05, the actual level for a nonnormal error distribution might be .04 or .065. Typically, the achieved level of significance in the presence of nonnormality is slightly higher than the specified one, and the achieved power of the test is slightly less than the calculated one. Single interval estimates of factor level means and contrasts and the Scheffé multiple comparison procedure also are not much affected by lack of normality, provided that the sample sizes are not extremely small.

For the random ANOVA model II (to be discussed in Chapter 25), lack of normality has more serious implications. The estimators of the variance components are still unbiased, but the actual confidence coefficient for interval estimates may be substantially different from the specified one.

Unequal Error Variances

When the error variances are unequal, the F test for the equality of means with the fixed ANOVA model is only slightly affected if all factor level sample sizes are equal or do not differ greatly. Specifically, unequal error variances then raise the actual level of significance slightly higher than the specified level. Similarly, the Scheffé multiple comparison procedure based on the F distribution is not affected to any substantial extent by unequal variances when the sample sizes are equal or are approximately the same. Thus, the F test and related analyses are robust against unequal variances when the sample sizes are approximately equal. Single comparisons between factor level means, on the other hand, can be substantially affected by unequal variances, so that the actual and specified confidence coefficients may differ markedly in these cases.

The use of equal sample sizes for all factor levels not only tends to minimize the effects of unequal variances on inferences with the F distribution but also simplifies calculational procedures. Thus, here at least, simplicity and robustness go hand in hand.

For the random ANOVA model II, unequal error variances can have pronounced effects on inferences about the variance components, even with equal sample sizes.

Nonindependence of Error Terms

Lack of independence of the error terms can have serious effects on inferences in the analysis of variance, for both fixed and random ANOVA models. Since this defect is often difficult to correct, it is important to prevent it in the first place whenever feasible. The use of randomization in those stages of a study that are likely to lead to correlated error terms can be a most important insurance policy. In the case of observational data, however, randomization may not be possible. Here, in the presence of correlated error terms, it may be possible to modify the model. For instance, in the earlier discussion based on Figure 18.3, we noted that inclusion in the model of a linear term for the learning effect of the analyst might remove the correlation of the error terms.

Modification of the model because of correlated error terms may also be necessary in experimental studies. In one case, the experimenter asked each of 10 subjects to give ratings to four new flavors of a fruit syrup and to the standard flavor, on a scale from 0 to 100. When the single-factor analysis of variance model was applied, the experimenter found

high degrees of correlation in the residuals for each subject. The experimenter thereupon modified the model to a repeated measures design model (Chapter 27). As described in Chapter 15, this latter type of model is intended for situations where the same subject is given each of the different treatments and differences between subjects are expected.

8.7 Nonparametric Rank F Test

When transformations are not successful in bringing the distributions of the error terms close enough to normality to meet the robustness properties of the standard inference procedures, a nonparametric inference procedure can be useful. Nonparametric procedures do not depend on the distribution of the error terms; often the only requirement is that the distribution is continuous. The nonparametric procedure considered here assumes that the r populations under study are continuous distributions that differ only with respect to location. Thus it provides a test for differences in population means or medians, assuming that the shapes of the populations (i.e., variances, skewness, kurtosis, etc.) are identical.

The test procedure is very simple. All n_T observations are ranked from 1 to n_T in ascending order. Then, the usual F^* test statistic in (16.55) is calculated, but now based on the ranks, and the F test is carried out in the ordinary manner.

Test Procedure

The Y_{ij} observations first need to be ranked in ascending order from 1 to n_T. We shall let R_{ij} denote the rank of Y_{ij}. In the case of ties among some observations, each of the tied observations is given the mean of the ranks involved. For instance, if two observations are tied for what would otherwise have been the third- and fourth-ranked positions, each would be given the mean value 3.5.

To test whether the treatment means are equal, the usual F^* test statistic is obtained based on the ranks R_{ij}. This test statistic is now denoted by F_R^*:

$$F_R^* = \frac{MSTR}{MSE} \tag{18.25}$$

where:

$$MSTR = \frac{\sum n_i(\bar{R}_{i\cdot} - \bar{R}_{\cdot\cdot})^2}{r-1} \tag{18.25a}$$

$$MSE = \frac{\sum\sum(R_{ij} - \bar{R}_{i\cdot})^2}{n_T - r} \tag{18.25b}$$

$$\bar{R}_{i\cdot} = \frac{\sum_j R_{ij}}{n_i} \tag{18.25c}$$

$$\bar{R}_{\cdot\cdot} = \frac{\sum\sum R_{ij}}{n_T} = \frac{(n_T+1)}{2} \tag{18.25d}$$

Note that $\bar{R}_{\cdot\cdot}$, the overall mean of the ranks, is a constant for any given total number of cases n_T.

When the treatment means are the same, test statistic F_R^* follows approximately the $F(r-1, n_T - r)$ distribution provided that the sample sizes n_i are not very small. To test

the alternatives:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_r$$
$$H_a: \text{not all } \mu_i \text{ are equal} \tag{18.26a}$$

the appropriate decision rule to control the Type I error at α is:

$$\text{If } F_R^* \leq F(1-\alpha; r-1, n_T - r), \text{ conclude } H_0$$
$$\text{If } F_R^* > F(1-\alpha; r-1, n_T - r), \text{ conclude } H_a \tag{18.26b}$$

Example

In the Servo-Data example of Table 18.5, we noted earlier that the logarithmic transformation of Y improves considerably the appropriateness of the assumptions of normality and constancy of the error variance. If the search for a transformation of Y had not been successful, or as an alternative to the transformation approach, we could use the nonparametric rank F test. To use this test, we first rank the data in Table 18.5 from 1 to 15. The ranks are shown in Table 18.5. Note, incidentally, from Table 18.5 that the rank transformation has helped to stabilize the variances of the transformed observations (i.e., the ranks) for the three treatments. We now calculate $SSTR$ and SSE as follows:

$$SSTR = 5[(8.4 - 8.0)^2 + (4.8 - 8.0)^2 + (10.8 - 8.0)^2] = 91.20$$
$$SSE = (2 - 8.4)^2 + (13 - 8.4)^2 + \cdots + (8 - 10.8)^2 = 188.80$$

Note that the overall mean $\bar{R}_{..}$ here is $(n_T + 1)/2 = (15 + 1)/2 = 8.0$. The F_R^* test statistic is therefore:

$$F_R^* = \frac{91.20}{3-1} \div \frac{188.8}{15-3} = 2.90$$

For $\alpha = .10$, we require $F(.90; 2, 12) = 2.81$. Since $F_R^* = 2.90 > 2.81$, we conclude H_a. The P-value of the test is .094.

Recall that when we conducted the standard F test based on the logarithmic transformation of Y, which was suggested both by the simple guides and the Box-Cox procedure, we found that it leads to the same conclusion here; but its P-value—.053—is considerably smaller. Thus, both tests show that the mean time between computer failures differs for the three locations.

Comment

The *Kruskal-Wallis test* (Ref. 18.5), a widely used nonparametric test for testing the equality of treatment means, is based on a test statistic that is equivalent to the rank F test statistic. The Kruskal-Wallis test statistic, denoted by X^2_{KW}, is also based on the ranks R_{ij} from 1 to n_T and is defined as follows:

$$X^2_{KW} = \frac{SSTR}{\frac{SSTO}{n_T - 1}} \tag{18.27}$$

where:

$$SSTO = \sum\sum(R_{ij} - \bar{R}_{..})^2 \tag{18.27a}$$

Instead of using the F distribution approximation, the Kruskal-Wallis test uses a chi-square distribution approximation. If the n_i are reasonably large (five or more is the usual advice), X^2_{KW} is approximately a χ^2 random variable with $r - 1$ degrees of freedom when all treatment means are equal. The decision

rule therefore is:

$$\begin{aligned} &\text{If } X^2_{KW} \le \chi^2(1-\alpha; r-1), \text{ conclude } H_0 \\ &\text{If } X^2_{KW} > \chi^2(1-\alpha; r-1), \text{ conclude } H_a \end{aligned} \tag{18.28}$$

The F^*_R and X^2_{KW} test statistics are equivalent, being related as follows:

$$F^*_R = \frac{(n_T - r)X^2_{KW}}{(r-1)(n_T - 1 - X^2_{KW})} \tag{18.29}$$

■

ultiple Pairwise Testing Procedure

If the rank F test (or the Kruskal-Wallis test) leads to the conclusion that the factor level means μ_i are not equal, it is frequently desired to obtain information about the comparative magnitudes of these means based on the ranked data. A large-sample testing analogue of the Bonferroni pairwise comparison procedure discussed in Section 17.7, based on the ranks of the observations, may be employed for this purpose, provided that the sample sizes are not too small. Testing limits for all $g = r(r-1)/2$ pairwise tests using the mean ranks $\bar{R}_{i\cdot}$ are set up as follows for family level of significance α:

$$(\bar{R}_{i\cdot} - \bar{R}_{i'\cdot}) \pm B\left[\frac{n_T(n_T+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\right]^{1/2} \tag{18.30}$$

where:

$$B = z(1 - \alpha/2g) \tag{18.30a}$$

$$g = \frac{r(r-1)}{2} \tag{18.30b}$$

If the testing limits include zero, we conclude that the corresponding treatment means μ_i and $\mu_{i'}$ do not differ. If the testing limits do not include zero, we conclude that the two corresponding treatment means differ. On the basis of all pairwise tests, we then set up groups of treatment means whose members do not differ according to the simultaneous testing procedure. In this way, we obtain information about the comparative magnitudes of the treatment means μ_i.

Example

For the Servo-Data example in Table 18.5, we wish to ascertain, if possible, which location has the longest mean time between computer failures based on the rank data. For a family significance level of $\alpha = .10$ and $g = r(r-1)/2 = 3(2)/2 = 3$ pairwise tests, we require $B = z(.9833) = 2.13$. Since all treatment sample sizes are equal, we need to calculate the right term in (18.30) only once:

$$B\left[\frac{n_T(n_T+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\right]^{1/2} = 2.13\left[\frac{15(16)}{12}\left(\frac{1}{5} + \frac{1}{5}\right)\right]^{1/2} = 6.02$$

Hence, the testing limits for the three pairwise tests are:

Locations 1 and 2:	$(8.4 - 4.8) \pm 6.02$	or	-2.4 and 9.6
Locations 3 and 2:	$(10.8 - 4.8) \pm 6.02$	or	$-.02$ and 12.0
Locations 3 and 1:	$(10.8 - 8.4) \pm 6.02$	or	-3.6 and 8.4

Since no test shows a significant difference, we obtain only one grouping:

Group 1
Location 1
Location 2
Location 3

Note that zero is just inside the lower boundary of the testing limits for locations 2 and 3.

Recall that when the Bonferroni pairwise comparison procedure was conducted on the logarithm of the responses, we concluded that a significant difference existed between the means of locations 2 and 3. Thus here, and in general for small sample sizes, the simple transformations discussed in Section 18.5 are often preferred to the rank transformation because the resulting ANOVA tests are less conservative and tend to have greater statistical power than those associated with the rank transformation.

18.8 Case Example—Heart Transplant

In heart transplant surgery, the similarity of the donor's tissue type and that of the recipient is of importance because large differences may increase the probability that the transplanted heart is rejected. Table 18.7 shows a portion of the survival times (in days) obtained from an observational study of 39 patients following heart transplant surgery. The data are grouped into three categories, according to the degree of mismatch between the donor tissue and the recipient tissue. Investigators would like to determine if the mean survival time changes with the degree of mismatch. The alternatives to be tested are:

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$$H_a: \text{not all } \mu_i \text{ are equal}$$

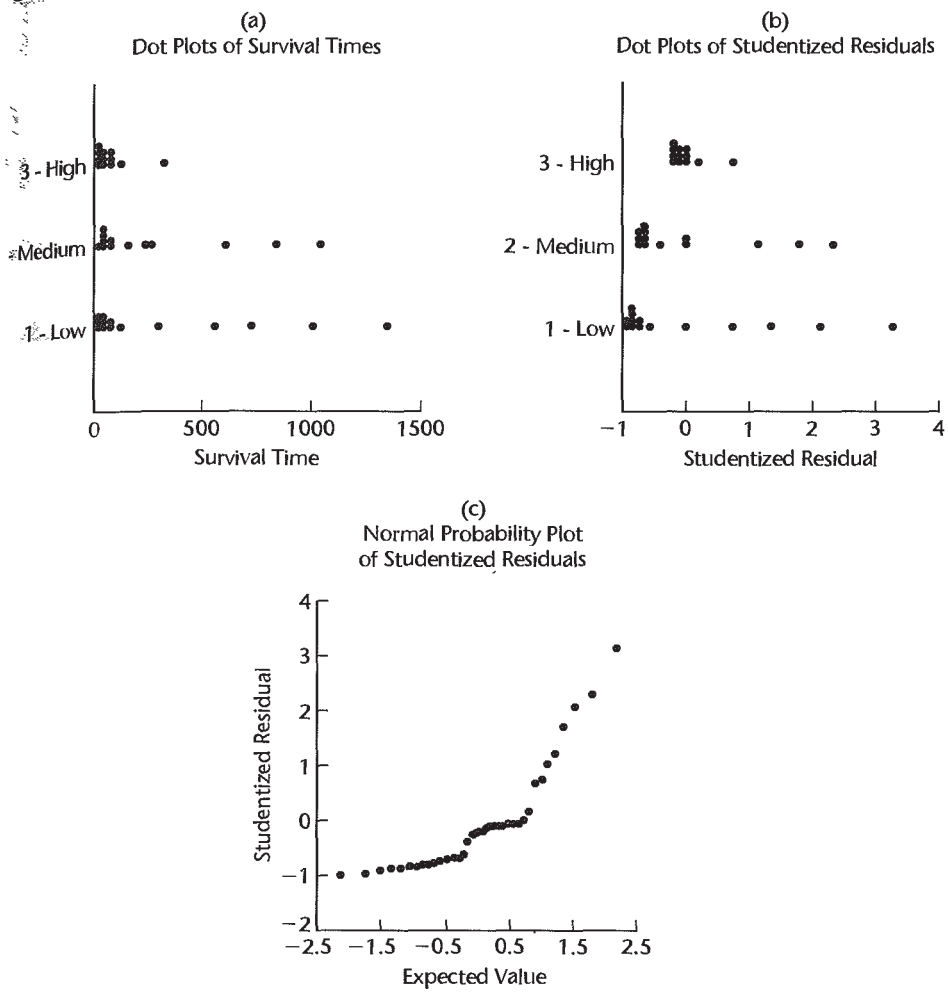
A SYSTAT dot plot of the data by mismatch category is provided in Figure 18.9a. The plot suggests that average survival time may decrease with higher degree of mismatch. An initial fit of analysis of variance model (16.2) was made and the studentized residuals were

TABLE 18.7 Survival Times of Patients Following Heart Transplant Surgery—Heart Transplant Example.

	Degree of Tissue Mismatch (i)		
Case j	Low $i = 1$	Medium $i = 2$	High $i = 3$
1	44	15	3
2	551	280	136
3	127	1,024	65
...	...	...	...
12	47	836	48
13	994	51	
14	26		

Source: M. L. Puri and P. K. Sen, *Nonparametric Methods in General Linear Models* (New York: John Wiley & Sons, 1985).

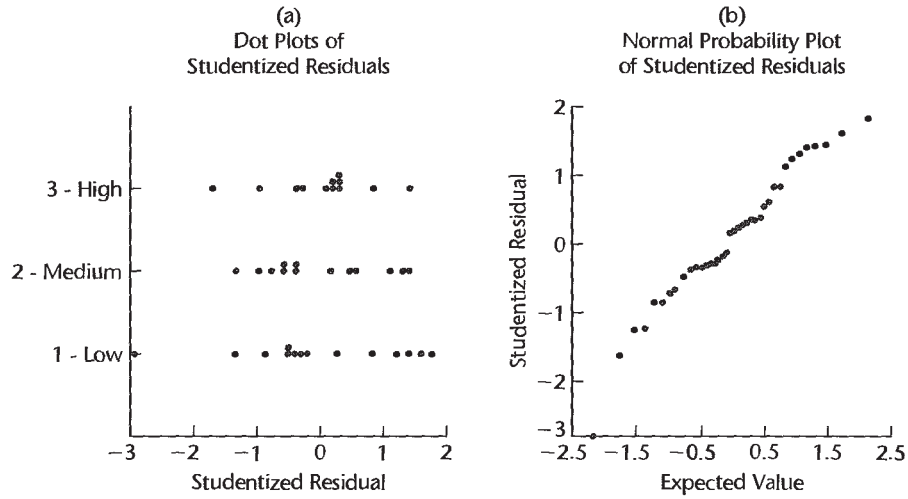
FIGURE 18.9 SYSTAT Diagnostic Plots—Heart Transplant Example.



obtained for diagnostic purposes. Two residual plots are presented in Figures 18.9b and 18.9c. The dot plot of the studentized residuals in Figure 18.9b shows that the distribution of the residuals is positively skewed. It also suggests that the error variance may be smaller in the high mismatch group. The Brown-Forsythe test in (18.12) was conducted to examine the constancy of the error variance. The Brown-Forsythe test statistic is $F^*_{BF} = 1.91$ and the P-value is .163, supporting constancy of the error variance. On the other hand, the positive skewness of the residuals is confirmed by the upward-curving shape of the normal probability plot in Figure 18.9c and the correlation test for normality ($r = .895$; for $\alpha = .05$, the interpolated critical value in Table B.6 is .971).

A transformation of the response variable was therefore investigated. The Box-Cox procedure led to the maximum likelihood estimate $\hat{\lambda} = .06$, which suggested the logarithmic transformation ($\lambda = 0$). The new response variable $Y' = \log_e Y$ was therefore obtained

FIGURE 18.10 Diagnostic Plots and ANOVA Table for Transformed Data—Heart Transplant Example.



(c)
ANOVA Table

DEP VAR: LOGTIME N: 39 MULTIPLE R: 0.282 SQUARED MULTIPLE R: 0.080

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
CATEGORY	7.611	2	3.806	1.560	0.224
ERROR	87.834	36	2.440		

and ANOVA model (16.2) was fitted to this transformed variable. Two plots of studentized residuals are shown in Figure 18.10. A dot plot of the studentized residuals is presented in Figure 18.10a. Notice that the distribution of the residuals now appears to be symmetric, with constant variance. The normality of the distribution of the error terms is supported by the normal probability plot in Figure 18.10b and the correlation test for normality ($r = .982 > .971$).

The residual dot plot in Figure 18.10a shows the possible presence of an outlier in the low tissue mismatch category (studentized residual $= -2.99$). For this case the studentized deleted residual is -3.40. The Bonferroni critical value for the outlier test is $t(1-.05/2(39); 36) = t(.999359; 36) = 3.49$. Since $|-3.40| = 3.40 \leq 3.49$, we conclude that this case is not an outlier.

It therefore appears that the logarithmic transformation was successful so that ANOVA model (16.2) is appropriate for the transformed survival times. The ANOVA table for the transformed data is shown in Figure 18.10c. We see that $F^* = 1.56$ and that the P-value for the test is .224. For $\alpha = .10$, we therefore conclude H_0, that the mean survival time for heart transplant patients with the characteristics of those included in the study does not depend on the degree of tissue mismatch.

ed eferences

18.1. Hartley, H. O. "Testing the Homogeneity of a Set of Variances," *Biometrika* 31 (1940), pp. 249–255.

18.2. Brown, M. B., and A. B. Forsythe. "Robust Tests for Equality of Variances," *Journal of the American Statistical Association* 69 (1974), pp. 364–67.

18.3. Snedecor, G. W., and W. G. Cochran. *Statistical Methods.* 8th ed. Ames, Iowa: Iowa State University Press, 1989.

18.4. Scheffé, H. *The Analysis of Variance.* New York: John Wiley & Sons, 1959.

18.5. Kruskal, W. H., and W. A. Wallis. "Use of Ranks on One-Criterion Variance Analysis," *Journal of the American Statistical Association* 47 (1952), pp. 583–621 (corrections appear in Vol. 48, pp. 907–11).

oblems

18.1. Refer to Figures 18.3 and 18.4. What feature of the residual sequence plots enables you to diagnose that in one case the error variance changes over time whereas in the other case the effect is of a different nature? Could you make a diagnosis about time effects from a residual dot plot?

18.2. A student proposed in class that deviations of the observations Y_{ij} around the estimated overall mean $\bar{Y}_{..}$ be plotted to assist in evaluating the appropriateness of ANOVA model (16.2). Would these deviations be helpful in studying the independence of the error terms? The constancy of the variance of the error terms? The normality of the error terms? Discuss.

18.3. A consultant discussing ANOVA applications in a seminar stated: "Sometimes I find that treatment effects in an experiment do not show up through differences in the treatment means. Hence, it is important to compare the residual plots for the treatments." A member of the audience asked: "I don't think I understood your point regarding differences in treatment means being explored using residual plots." Discuss.

*18.4. Refer to **Productivity improvement** Problem 16.7.

a. Prepare aligned residual dot plots by factor level. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

d. The economist wishes to investigate whether location of the firm's home office is related to productivity improvement. The home office locations are as follows (U: U.S.; E: Europe):

	j											
i	1	2	3	4	5	6	7	8	9	10	11	12
1	U	E	E	E	E	U	U	U	U			
2	E	E	E	E	U	U	U	U	U	E	E	E
3	E	U	E	U	U	E						

Prepare aligned residual dot plots by factor level in which the location of the home office is identified. Does it appear that ANOVA model (16.2) could be improved by adding location of home office as a second factor? Explain.

18.5. Refer to **Questionnaire color** Problem 16.8.

a. Prepare aligned residual dot plots by color. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. The observations within each factor level are in geographic sequence. Prepare residual sequence plots. What can be studied from these plots? What are your findings?

d. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .025$. State the alternatives, decision rule, and conclusion.

18.6. Refer to **Rehabilitation therapy** Problem 16.9.

a. Obtain the residuals and prepare aligned residual dot plots by factor level. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. The observations within each factor level are in time order. Prepare residual sequence plots and analyze them. What are your findings?

d. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

*18.7. Refer to **Cash offers** Problem 16.10.

a. Obtain the residuals and prepare aligned residual dot plots by factor level. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. The observations within each factor level are in time order. Prepare residual sequence plots and interpret them. What are your findings?

d. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .025$. State the alternatives, decision rule, and conclusion.

e. An executive in the consumer organization has been told that used-car dealers in the region tend to make lower cash offers during weekends (Friday evening through Sunday) than at other times. The times when offers were obtained are as follows (W: weekend; O: other time):

	j											
i	1	2	3	4	5	6	7	8	9	10	11	12
1	O	O	W	O	W	O	W	O	W	O	W	W
2	O	W	W	O	W	O	W	O	O	W	W	O
3	O	W	O	W	O	O	O	W	W	W	O	W

Prepare aligned residual dot plots by factor level in which the time of the offer is identified. Does it appear that ANOVA model (16.2) could be improved by adding time of offer as a second factor? Explain.

*18.8. Refer to **Filling machines** Problem 16.11.

a. Obtain the residuals and prepare aligned residual dot plots by machine. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. The observations within each factor level are in time order. Prepare residual sequence plots and interpret them. What are your findings?

d. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

18.9. Refer to **Premium distribution** Problem 16.12.

a. Obtain the residuals and prepare aligned residual dot plots by agent. What departures from ANOVA model (16.2) can be studied from these plots? What are your findings?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

c. The observations within each factor level are in time order. Prepare residual sequence plots and interpret them. What are your findings?

d. Obtain the studentized deleted residuals and conduct the Bonferroni outlier test; use $\alpha = .025$. State the alternatives, decision rule, and conclusion.

18.10. **Computerized game.** Four teams competed in 20 trials of a computerized business game. Each trial involved a new game, the objective for each team being to maximize profits in the given trial. A researcher fitted ANOVA model (16.2) to determine whether or not the mean profits for the four teams are the same and obtained the following residuals:

	j						
i	1	2	3	...	18	19	20
1	.10	.28	.10	...	.10	.28	.28
2	−1.44	−1.44	−1.12	...	1.02	1.18	1.51
3	−.93	−.70	−.81	...	.54	.43	.65
4	−.15	.11	.25	...	.11	.25	.38

The residuals for each team are given in time order. Construct appropriate residual plots to study whether the error terms are independent from trial to trial for each team. What are your findings?

*18.11. Refer to **Productivity improvement** Problem 16.7. Examine by means of the Brown-Forsythe test whether or not the treatment error variances are equal; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

18.12. Refer to **Rehabilitation therapy** Problem 16.9. Examine by means of the Brown-Forsythe test whether or not the treatment error variances are equal; use $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

*18.13. Refer to **Cash offers** Problem 16.10. Assume that the error terms are approximately normally distributed.

a. Examine by means of the Hartley test whether or not the treatment error variances are equal; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Would you reach the same conclusion as in part (a) with the Brown-Forsythe test?

*18.14. Refer to **Filling machines** Problem 16.11. Assume that the error terms are approximately normally distributed.

a. Examine by means of the Hartley test whether or not the treatment error variances are equal; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Would you reach the same conclusion as in part (a) with the Brown-Forsythe test statistic?

18.15. **Helicopter service.** An operations analyst in a sheriff's department studied how frequently their emergency helicopter was used during the past year, by time of day (shift 1: 2 A.M.–8 A.M.; shift 2: 8 A.M.–2 P.M.; shift 3: 2 P.M.–8 P.M.; shift 4: 8 P.M.–2 A.M.). Random samples of size 20 for each shift were obtained. The data follow (in time order):

	j						
i	1	2	3	...	18	19	20
1	4	3	5	...	4	1	6
2	0	2	0	..	2	2	0
3	2	1	0		0	2	4
4	5	2	4		5	2	3

Since the data are counts, the analyst was concerned about the normality and equal variances assumptions of ANOVA model (16.2).

a. Obtain the fitted values and residuals for ANOVA model (16.2).

b. Prepare suitable residual plots to study whether or not the error variances are equal for the four shifts. What are your findings?

c. Test by means of the Brown-Forsythe test whether or not the treatment error variances are equal; use $\alpha = .10$. What is the P-value of the test? Are your results consistent with the diagnosis in part (b)?

d. For each shift, calculate $\bar{Y}_i$, and s_i. Examine the three relations found in the table on page 791 and determine the transformation that is most appropriate here. What do you conclude?

e. Use the Box-Cox procedure to find an appropriate power transformation of Y, first adding the constant 1 to each Y observation. Evaluate SSE for the values of λ given in Table 18.6. Does $\lambda = .5$, a square-root transformation, appear to be reasonable, based on the Box-Cox procedure?

18.16. Refer to **Helicopter service** Problem 18.15. The analyst decided to apply the square root transformation $Y' = \sqrt{Y}$ and examine its effectiveness.

a. Obtain the transformed response data, fit ANOVA model (16.2), and obtain the residuals.

b. Prepare suitable plots of the residuals to study the equality of the error variances of the transformed response variable for the four shifts. Also obtain a normal probability plot and the coefficient of correlation between the ordered residuals and their expected values under normality. What are your findings? Does the transformation appear to have been effective?

c. Test by means of the Brown-Forsythe test whether or not the treatment error variances for the transformed response variable are equal; use $\alpha = .10$. State the alternatives,

decision rule, and conclusion. Are your findings in part (b) consistent with your conclusion here?

*18.17. **Winding speeds.** In a completely randomized design to study the effect of the speed of winding thread (1: slow; 2: normal; 3: fast; 4: maximum) onto 75-yard spools, 16 runs of 10,000 spools each were made at each of the four winding speeds. The response variable is the number of thread breaks during the production run. The results (in time order) are as follows:

	j						
i	1	2	3	...	14	15	16
1	4	3	2	...	2	3	4
2	7	6	4	...	4	7	6
3	12	6	14	...	13	10	14
4	17	15	7	...	19	9	23

Since the responses are counts, the researcher was concerned about the normality and equal variances assumptions of ANOVA model (16.2).

a. Obtain the fitted values and residuals for ANOVA model (16.2).

b. Prepare suitable residual plots to study whether or not the error variances are equal for the four winding speeds. What are your findings?

c. Test by means of the Brown-Forsythe test whether or not the treatment error variances are equal; use $\alpha = .05$. What is the P-value of the test? Are your results consistent with the diagnosis in part (b)?

d. For each winding speed, calculate $\bar{Y}_{i\cdot}$ and s_i. Examine the three relations found in the table on page 791 and determine the transformation that is most appropriate here. What do you conclude?

e. Use the Box-Cox procedure to find an appropriate power transformation of Y. Evaluate SSE for the values of λ given in Table 18.6. Does $\lambda = 0$, a logarithmic transformation, appear to be reasonable, based on the Box-Cox procedure?

*18.18. Refer to **Winding speeds** Problem 18.17. The researcher decided to apply the logarithmic transformation $Y' = \log_{10} Y$ and investigate its effectiveness.

a. Obtain the transformed response data, fit ANOVA model (16.2), and obtain the residuals.

b. Prepare suitable plots of the residuals to study the equality of the error variances of the transformed response variable for the four winding speeds. Also obtain a normal probability plot and the coefficient of correlation between the ordered residuals and their expected values under normality. What are your findings about the effectiveness of the transformation?

c. Test by means of the Brown-Forsythe test whether or not the treatment error variances for the transformed response variable are equal; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. Are your findings in part (b) consistent with your conclusion here?

18.19. Refer to **Helicopter service** Problem 18.15. Assume that ANOVA model (18.13) is appropriate. Use weighted least squares with the untransformed data to test for the equality of the shift means; control the α risk at .05. State the alternatives, full and reduced regression models, decision rule, and conclusion.

*18.20. Refer to **Winding speeds** Problem 18.17. Assume that ANOVA model (18.13) is appropriate. Use weighted least squares with the untransformed data to test for the equality of the winding

thread speed means; use $\alpha = .01$. State the alternatives, full and reduced regression models, decision rule, and conclusion.

18.21. Why is the nonparametric rank F test a nonparametric test?

18.22. Explain why the limits in (18.30) are testing limits and not confidence limits.

*18.23. Refer to **Productivity improvement** Problem 16.7.

a. Conduct the nonparametric rank F test; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

b. What is the P-value of the test in part (a)?

c. Does the conclusion in part (a) differ from the one in Problem 16.7e?

d. Do the data suggest that a nonparametric test is needed here?

e. Conduct multiple pairwise tests based on the ranked data to group the three types of firms according to mean productivity improvement. Use family level of significance $\alpha = .10$. Describe your findings.

*18.24. Refer to **Cash offers** Problem 16.10.

a. Conduct the nonparametric rank F test; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

b. What is the P-value of the test in part (a)?

c. Does the conclusion in part (a) differ from the one in Problem 16.10e?

d. Do the data suggest that a nonparametric test is needed here?

e. Conduct multiple pairwise tests based on the ranked data to group the three age categories according to mean cash offer. Use family level of significance $\alpha = .10$. Describe your findings.

18.25. **Telephone communications.** A management consultant was engaged by a firm to improve the cost-effectiveness of its communications. As part of the study, the consultant selected 10 home-office executives at random from each of the (1) sales, (2) production, and (3) research and development divisions, and studied the communications of these executives during the past 10 weeks in great detail. Among other data, the consultant obtained the following information on weekly dollar costs of long-distance telephone calls to branch offices by the executives:

	j									
i	1	2	3	4	5	6	7	8	9	10
1	666	920	495	602	1,499	960	796	343	894	813
2	488	362	156	546	216	542	345	291	516	126
3	391	450	609	910	705	472	645	496	763	1,309

The consultant decided to employ a nonparametric approach to test whether or not the mean telephone expenses for the three divisions are equal.

a. What feature of the data may have suggested the use of a nonparametric test?

b. Conduct the nonparametric rank F test, controlling the risk of Type I error at $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Conduct multiple pairwise tests based on the ranked data to group the three divisions according to mean telephone expenditures; use family level of significance $\alpha = .05$. Describe your findings.

Exercises

18.26. Refer to Figure 18.3. Modify ANOVA model (16.2) to include a linear trend term for the time effect. Is this modified model still an ANOVA model? A linear model?

18.27. Show that $n_T(n_T+1)/12$ in (18.30) is the sample variance of the consecutive integers 1 to n_T.

18.28. Show that test statistics (18.25) and (18.27) are related according to (18.29).

Projects

18.29. Refer to the **SENIC** data set in Appendix C.1 and Project 16.42.

a. Obtain the residuals and prepare aligned residual dot plots by region. Are any serious departures from ANOVA model (16.2) suggested by your plots?

b. Obtain a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Is the normality assumption reasonable here?

c. Examine by means of the Brown-Forsythe test whether or not the geographic region error variances are equal; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

18.30. Refer to the **SENIC** data set in Appendix C.1. A test of whether or not mean length of stay (variable 2) is the same in the four geographic regions (variable 9) is desired, but concern exists about the normality and equal variances assumptions of ANOVA model (16.2).

a. Obtain the residuals and plot them against the fitted values to study whether or not the error variances are equal for the four geographic regions. What are your findings?

b. For each geographic region, calculate $\bar{Y}_{i\cdot}$ and s_i. Examine the three relations found in the table on page 791 and determine the transformation that is the most appropriate one here. What do you conclude?

c. Use the Box-Cox procedure to find an appropriate power transformation of Y. Evaluate SSE for the values of λ given in Table 18.6. Does $\lambda = -1$, a reciprocal transformation, appear to be reasonable, based on the Box-Cox procedure?

d. Use the reciprocal transformation $Y' = 1/Y$ to obtain transformed response data.

e. Fit ANOVA model (16.2) to the transformed data and obtain the residuals. Plot these residuals against the fitted values to study the equality of the error variances of the transformed response variable for the four regions. Also obtain a normal probability plot of the residuals and the coefficient of correlation between the ordered residuals and their expected values under normality. What are your findings?

f. Examine by means of the Brown-Forsythe test whether or not the geographic region variances for the transformed response variable are equal; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

g. Assume that ANOVA model (16.2) is appropriate for the transformed response variable. Test whether or not the mean length of stay in the transformed units is the same in the four geographic regions. Control the α risk at .01. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

18.31. Refer to the **CDI** data set in Appendix C.2 and Project 16.44.

a. Obtain the residuals and prepare aligned residual dot plots by region. Are any serious departures from ANOVA model (16.2) suggested by your plots?

b. Obtain a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Is the normality assumption reasonable here?

c. Examine by means of the Brown-Forsythe test whether or not the geographic region error variances are equal; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

18.32. Refer to the **Market share** data set in Appendix C.3 and Project 16.45.

a. Obtain the residuals and prepare aligned residual dot plots by factor-level combinations. Are any serious departures from ANOVA model (16.2) suggested by your plots?

b. Obtain a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Is the normality assumption reasonable here?

c. Examine by means of the Brown-Forsythe test whether or not the factor level error variances are equal; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

18.33. Refer to the **SENIC** data set in Appendix C.1 and Project 16.42.

a. Use the nonparametric rank F test to determine whether or not the mean infection risk is the same in the four regions; control the level of significance at $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Is your conclusion in part (a) the same as that obtained in Project 16.42? Is the nonparametric test more reasonable here?

c. Use the multiple pairwise testing procedure (18.30) to group the regions; employ family significance level $\alpha = .10$. What are your findings?

18.34. Refer to the **CDI** data set in Appendix C.2 and Project 16.44.

a. Use the nonparametric rank F test to determine whether or not the mean crime rate is the same in the four regions; control the level of significance at $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Is your conclusion in part (a) the same as that obtained in Project 16.44? Is the nonparametric test more reasonable here?

c. Use the multiple pairwise testing procedure (18.30) to group the regions; employ family significance level $\alpha = .05$. What are your findings?

18.35. Refer to the **Market share** data set in Appendix C.3 and Project 16.45.

a. Use the nonparametric rank F test to determine whether or not the mean average monthly share is the same for the four factor combinations; control the level of significance at $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Is your conclusion in part (a) the same as that obtained in Project 16.45? Is the nonparametric test more reasonable here?

c. Use the multiple pairwise testing procedure (18.30) to group the factor combinations; employ family significance level $\alpha = .05$. What are your findings?

18.36. Obtain the exact sampling distribution of the nonparametric rank F_R^* test statistic in (18.25) when H_0 holds, for the case $r = 2$ and $n_i \equiv 2$. [*Hint:* What does the equality of the treatment means imply about the arrangement of the ranks 1, 2, 3, 4?]

18.37. Three populations are being studied; each is uniform between 300 and 800.

a. Generate 10 random observations from each of the three uniform populations and calculate the F_R^* test statistic (18.25).

b. Repeat part (a) 500 times.

c. Calculate the mean and standard deviation of the 500 test statistics. How do these values compare with the characteristics of the relevant F distribution?

d. What proportion of the 500 test statistics obtained in part (b) is less than $F(.90; 2, 27)$? What proportion is less than $F(.99; 2, 27)$? How do these proportions agree with theoretical expectations?

Case Studies

18.38. Refer to the **Prostate cancer** data set in Appendix C.5 and Case Study 16.49. Check to see whether concern exists about the assumption of normality and equal variances for the ANOVA model that you decided upon in Case Study 16.49. Document the steps taken in your assessment of these concerns. Is a transformation indicated here? If yes, what transformation is recommended? Why?

18.39. Refer to the **Real estate sales** data set in Appendix C.7 and Case Study 16.50. Check to see whether concern exists about the assumption of normality and equal variances for the ANOVA model that you decided upon in Case Study 16.50. Document the steps taken in your assessment of these concerns. Is a transformation indicated here? If yes, what transformation is recommended? Why?

18.40. Refer to the **Ischemic heart disease** data set in Appendix C.9 and Case Study 16.51. Check to see whether concern exists about the assumption of normality and equal variances for the ANOVA model that you decided upon in Case Study 16.51. Document the steps taken in your assessment of these concerns. Is a transformation indicated here? If yes, what transformation is recommended? Why?

Part V

ulti-Factor
tudies

Chapter

Two-Factor Studies with Equal Sample Sizes

In Part IV, we considered the design and analysis of experimental and observational studies in which the effects of one factor are investigated. Now we are concerned with investigations of the simultaneous effects of two or more factors. In this chapter, we take up the analysis of variance for two-factor studies where the factors are crossed and all sample sizes are equal. In Chapters 20, 21, 22, and 23, we continue the discussion of two-factor studies by taking up the analysis of factor effects with one case per cell, randomized complete block designs, the analysis of covariance, and two-factor studies with unequal sample sizes. In Chapter 24, we extend the analysis of variance to studies with three or more factors. Finally, in Chapter 25, we take up random and mixed effects models.

19.1 Two-Factor Observational and Experimental Studies

Two-factor studies, like single-factor studies, can be based on experimental or observational data. We begin with three examples of two-factor studies: the first is an experimental study, the second is an observational study, and the third has aspects of both experimental and observational studies.

Examples of Two-Factor Experiments and Observational Studies

Example 1

A company investigated the effects of selling price and type of promotional campaign on sales of one of its products. Three selling prices (55 cents, 60 cents, 65 cents) were studied, as were two types of promotional campaigns (radio advertising, newspaper advertising). Let us consider selling price to be factor A and promotional campaign to be factor B. Factor A here was studied at three price levels; in general, we use the symbol a to denote the number of levels of factor A investigated. Factor B was here studied at two levels; we use the symbol b to denote the number of levels of factor B investigated. Each combination of price and promotional campaign was studied, as shown in the

table below:

Treatment	Description
1	55 price, radio advertising
2	60 price, radio advertising
3	65 price, radio advertising
4	55 price, newspaper advertising
5	60 price, newspaper advertising
6	65 price, newspaper advertising

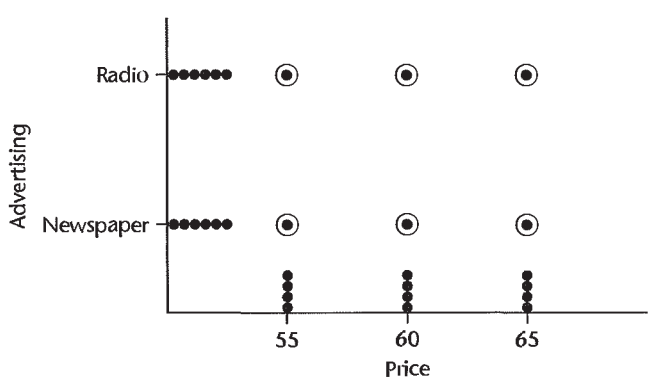
Each combination of a factor level of A and a factor level of B is a *treatment*. Thus, there are $3 \times 2 = 6$ treatments here altogether. In general, the total number of possible treatments in a two-factor study is ab.

Twelve communities throughout the United States, of approximately equal size and similar socioeconomic characteristics, were selected and the treatments were assigned to them at random, such that each treatment was given to two experimental units. The experiment can be represented by the graph in Figure 19.1. The two experimental units for each treatment combination are represented by the dot with circle circumscribed. Notice that four experimental units are assigned to each price level, as shown by the dot plot along the price (X) axis, and six experimental units are assigned to each mode of advertising, as shown by the dot plot along the advertising (Y) axis.

As before, we use n for the number of units receiving a given treatment when all treatment sample sizes are the same. For the $n = 2$ communities that were assigned treatment 1, for instance, the product price was fixed at 55 cents and radio advertising was employed, and so on for the other communities in the study.

This is an experimental study because control was exercised in assigning the factor A and factor B levels to the experimental units by means of random assignments of the treatments to the communities. The design used was a completely randomized design.

FIGURE 19.1 Experimental Layout—Example 1.



Example 2

An analyst studied the effects of family income (under \$15,000, \$15,000–\$29,999, \$30,000–\$49,999, \$50,000 and more) and stage in the life cycle of the family (stages 1, 2, 3, 4) on appliance purchases. Here, $4 \times 4 = 16$ treatments are defined. These are in part:

Treatment	Description
1	Under \$15,000 income, stage 1
2	Under \$15,000 income, stage 2
⋮	⋮
16	\$50,000 and more income, stage 4

The analyst selected 20 families with the required income and life-cycle characteristics for each of the "treatment" classes for this study, yielding 320 families for the entire study.

This study is an observational one because the data were obtained without assigning income and life-cycle stage to the families. Rather, the families were selected because they had the specified characteristics.

Example 3

A medical investigator studied the relationship between the response to three blood pressure lowering drug types for hypertensive males and females. Here, $3 \times 2 = 6$ treatments are defined. These are:

Treatment	Description
1	Drug type 1, males
2	Drug type 1, females
3	Drug type 2, males
4	Drug type 2, females
5	Drug type 3, males
6	Drug type 3, females

The investigator selected 30 adult males and 30 adult females and randomly assigned 10 males and 10 females to each of the three drug types, yielding 60 total subjects.

This study has one observational factor, gender, and one experimental factor, drug type. This design is referred to as a randomized complete block design where the gender factor is called a block. This design will be discussed in Chapter 21.

Comments

1. When we considered single-factor studies, we did not place any restrictions on the nature of the r factor levels under study. Formally, the ab treatments in a two-factor investigation could be considered as the r factor levels in a single-factor investigation and analyzed according to the methods discussed in Part IV. The reason why new methods of analysis are required is that we wish to analyze the ab treatments in special ways that recognize two factors are involved and enable us to obtain information about the main effects of each of the two factors as well as about any special joint effects.

2. When a completely randomized design is used in a multifactor study, the random assignments of treatments to the experimental units are made in the same manner as for a single-factor study. No new problems are encountered once the treatments are defined in terms of the factor levels of the various factors under study. ■

The One-Factor-at-a-Time (OFAAT) Approach to Experimentation

It is not uncommon for investigators to vary only one factor at a time, holding all others constant, when attempting to understand the effect of a given set of factors on a particular outcome. For example, to maximize sales in Example 1, we might be tempted to first fix price at a particular value such as 60 cents, and then determine which mode of advertising (radio or newspaper) is most effective. If this test reveals that newspaper advertising leads to higher sales, we would then run a second test in which the advertising mode is fixed at "newspaper," and the three price levels are tested. This *one-factor-at-a-time* (OFAAT) experimental approach is depicted in Figure 19.2.

We note a number of deficiencies of the OFAAT approach:

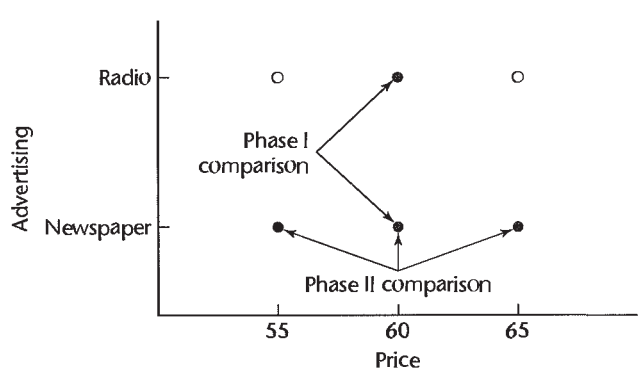
1. The OFAAT approach does not explore the entire space of treatment combinations, and important treatment combinations may therefore be missed. In Figure 19.2, we see that two treatment combinations—(radio, 55 cents) and (radio, 65 cents)—were omitted, or one-third of the total. The fraction of treatment combinations omitted can be much larger for studies involving larger numbers of factors and/or larger numbers of factor levels.

2. Interactions cannot be estimated. As we have seen in regression, an interaction between two predictors is present if the effect (slope) of one predictor changes with the level of the other predictor. With the OFAAT approach, this is impossible to determine, because the slope of one factor is obtained only for a fixed set of levels of the other factors.

3. A full randomization is not possible for the OFAAT approach, because the experiment must be fielded in stages. Thus if certain variables that are not under control of the experimenter change with the stages of the testing, the results may be adversely affected.

4. The OFAAT approach is often more difficult to field logistically, because of the sequence of stages. At each stage, the experimental apparatus is set up, responses are obtained, an analysis is carried out, and the next treatment combinations are determined. Setting up for each experimental phase can be difficult. For example, it may be necessary in an industrial experiment to reserve time on an assembly line or in a pilot plant well in advance. In a field study involving a survey, it may be necessary to preschedule subjects and interviewers. In addition, processing responses can be time-consuming—for example, if complicated laboratory analyses are required—and the subsequent phase of experimentation may be delayed significantly.

FIGURE 19.2 One-Factor-at-a-Time Approach—Example 1.



Advantages of Crossed, Multi-Factor Designs

Efficiency and Hidden Replication. Multi-factor studies are more efficient than the OFAAT experimental approach. Even though the OFAAT approach devotes all resources to studying the effect of only one factor, it does not yield any more precise information about that factor than a multi-factor experiment of the same size. With reference to Example 1 again, suppose that 12 communities were to be utilized in a traditional study, six assigned to radio advertising and the other six to newspaper advertising, and that the price would be kept constant at 60 cents. For this traditional study, the comparison between the two types of promotional campaigns would be based on two samples of six communities each. The same is true for the two-factor study in Example 1, since each promotional campaign occurs there in three treatments and each treatment has two communities assigned to it. Figure 19.1 reveals what is sometimes called *hidden replication* in a two-factor experiment. While there are only two replicates for each treatment combination, each level of advertising is repeated six times, and each level of price is repeated four times.

The increased efficiency due to hidden replication for main effect tests in multi-factor studies is only present when either unimportant interactions exist or when interaction effects are small relative to main effects. When important interactions are present, multiple comparisons of the individual cell means rather than comparisons of the main effects are usually conducted.

Assessment of Interactions. OFAAT studies provide no information about interactions. Specifically in our previous illustration, it does not provide any information about any special joint effects of price and promotional campaign. For instance, it might be that the price effects are not large when the promotional campaign is in newspapers but are large with radio advertising. Such interaction effects can be readily investigated from cross-classified multifactor studies.

Validity of Findings. In addition to being more efficient and readily providing information about interaction effects, multi-factor studies also can strengthen the validity of the findings. Suppose that in Example 1, management was principally interested in investigating the effects of price on sales. If the promotional campaign used in the price study had been newspaper advertising, doubts might exist as to whether or not the price effects differ for other promotional vehicles. By including type of promotional campaign as another factor in the study, management can get information about the persistence of the price effects with different promotional vehicles, without increasing the number of experimental units in the study. Thus, multifactor studies can include some factors of secondary importance to permit inferences about the primary factors with a greater range of validity.

Comments

1. Multi-factor studies permit a ready evaluation of interaction effects for observational data and economize on the number of cases required for the analysis, just as for experimental studies.

2. The advantages of multi-factor experiments just described should not lead one to think that inclusion of more factors necessarily results in a better study. Experiments involving many factors, each at numerous levels, become complex, costly, and time-consuming. It is often a better research strategy to begin with fewer factors and/or fewer levels for each factor, and then extend the investigation in accordance with the results obtained to date. In this way, resources can be devoted principally to the most promising avenues of investigation, and a better understanding of the effects of the factors can be obtained. ■

9.2 Meaning of ANOVA Model Elements

Before presenting a formal statement of the analysis of variance model for two-factor studies, we shall develop the model elements and discuss their meaning. This will not only be helpful in understanding the ANOVA model but will also provide insights into how the analysis of two-factor studies should proceed. *Throughout this section, we assume that all population means are known and are of equal importance when averages of these means are required.*

Illustration

To illustrate the meaning of the ANOVA model elements, we consider a simple two-factor study in which the effects of gender and age on learning of a task are of interest. For simplicity, the age factor has been defined in terms of only three factor levels (young, middle, old), as shown in Table 19.1a.

Treatment Means

The mean response for a given treatment in a two-factor study is denoted by μ_{ij}, where i refers to the level of factor A $(i = 1, \ldots, a)$ and j refers to the level of factor B $(j = 1, \ldots, b)$. Table 19.1a contains the true treatment means μ_{ij} for the learning example. Note, for instance, that $\mu_{11} = 9$, which indicates that the mean learning time for young males is 9 minutes. Similarly, we see that $\mu_{22} = 11$, so that the mean learning time for middle-aged females is 11 minutes.

The interpretation of a treatment mean μ_{ij} depends on whether the study is observational, experimental, or a mixture of the two. In an observational study, the treatment mean μ_{ij} corresponds to the population mean for the elements having the characteristics of the ith level of factor A and the jth level of factor B. For instance, in the learning example, the treatment mean μ_{11} is the mean learning time for the population of young males.

In an experimental study, the treatment mean μ_{ij} stands for the mean response that would be obtained if the treatment consisting of the ith level of factor A and the jth level of factor B were applied to all units in the population of experimental units about which

TABLE 19.1 Age Effect but No Gender Effect, with No Interactions—Learning Example.

(a) Mean Learning Times (in minutes)

	Factor B—Age			
Factor A—Gender	$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old	Row Average
$i = 1$ Male	9 (μ_{11})	11 (μ_{12})	16 (μ_{13})	12 ($\mu_{1\cdot}$)
$i = 2$ Female	9 (μ_{21})	11 (μ_{22})	16 (μ_{23})	12 ($\mu_{2\cdot}$)
Column average	9 ($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16 ($\mu_{\cdot 3}$)	12 ($\mu_{\cdot\cdot}$)

(b) Main Gender Effects (in minutes)

$$\alpha_1 = \mu_{1\cdot} - \mu_{\cdot\cdot} = 12 - 12 = 0$$
$$\alpha_2 = \mu_{2\cdot} - \mu_{\cdot\cdot} = 12 - 12 = 0$$

(c) Main Age Effects (in minutes)

$$\beta_1 = \mu_{\cdot 1} - \mu_{\cdot\cdot} = 9 - 12 = -3$$
$$\beta_2 = \mu_{\cdot 2} - \mu_{\cdot\cdot} = 11 - 12 = -1$$
$$\beta_3 = \mu_{\cdot 3} - \mu_{\cdot\cdot} = 16 - 12 = 4$$

inferences are to be drawn. For instance, in a study where factor A is type of training program (highly structured, partially structured, unstructured) and factor B is time of training (during work, after work), $6n$ employees are selected and n are assigned at random to each of the six treatments. The mean μ_{ij} here represents the mean response, say, mean gain in productivity, if the ith training program administered during the jth time were given to all employees in the population of experimental units.

Factor Level Means

The treatment means in Table 19.1a for the learning example indicate that the mean learning times for men and women are the same for each age group. On the other hand, the mean learning time increases with age for each gender. Thus, gender has no effect on mean learning time, but age does. This can also be seen quickly from the row averages and column averages shown in Table 19.1a, which in this case tell the complete story. The row averages are the gender factor level means, and the column averages are the age factor level means. We denote the column average for the first column by $\mu_{\cdot 1}$, which is the average of μ_{11} and μ_{21}. In general, the column average for the jth column is denoted by $\mu_{\cdot j}$:

$$\mu_{\cdot j} = \frac{\sum_{i=1}^{a} \mu_{ij}}{a} \tag{19.1}$$

and the row average for the ith row is denoted by $\mu_{i\cdot}$:

$$\mu_{i\cdot} = \frac{\sum_{j=1}^{b} \mu_{ij}}{b} \tag{19.2}$$

The overall mean learning time for all ages and both genders is denoted by $\mu_{\cdot\cdot}$, and is defined in the following equivalent fashions:

$$\mu_{\cdot\cdot} = \frac{\sum_i \sum_j \mu_{ij}}{ab} \tag{19.3a}$$

$$\mu_{\cdot\cdot} = \frac{\sum_i \mu_{i\cdot}}{a} \tag{19.3b}$$

$$\mu_{\cdot\cdot} = \frac{\sum_j \mu_{\cdot j}}{b} \tag{19.3c}$$

In Table 19.1a, the gender factor level means are $\mu_{1\cdot} = \mu_{2\cdot} = 12$ for the two genders, the age factor level means are $\mu_{\cdot 1} = 9$, $\mu_{\cdot 2} = 11$, and $\mu_{\cdot 3} = 16$ for the three age groups, and the overall mean learning time is $\mu_{\cdot\cdot} = 12$ minutes.

Main Effects

Main Age Effects. To summarize the main age effects, we shall consider the differences between each factor level mean and the overall mean. These differences are called main age effects. For instance, the main effect for young persons in Table 19.1a is the difference between $\mu_{\cdot 1}$, the mean learning time for young persons, and $\mu_{\cdot\cdot}$, the overall mean. This difference is denoted by β_1:

$$\beta_1 = \mu_{\cdot 1} - \mu_{\cdot\cdot} = 9 - 12 = -3$$

β_1 is called the *main effect* for factor B at the first level. This and the other main effects for factor B are shown in Table 19.1c.

Main Gender Effects. The main gender effects are defined in corresponding fashion, and denoted by α_i. For instance, we have:

$$\alpha_1 = \mu_{1\cdot} - \mu_{\cdot\cdot} = 12 - 12 = 0$$

α_1 is called the main effect for factor A at the first level. The main effects for factor A are shown in Table 19.1b. They are both zero, indicating that gender does not affect mean learning time.

General Definitions. In general, we define the main effect of factor A at the ith level as follows:

$$\alpha_i = \mu_{i\cdot} - \mu_{\cdot\cdot} \tag{19.4}$$

Similarly, the main effect of the jth level of factor B is defined:

$$\beta_j = \mu_{\cdot j} - \mu_{\cdot\cdot} \tag{19.5}$$

It follows from (19.3b) and (19.3c) that:

$$\sum_i \alpha_i = 0 \qquad \sum_j \beta_j = 0 \tag{19.6}$$

Thus, the sum of the main effects for each factor is zero.

Note again that a main effect indicates how much the factor level mean deviates from the overall mean. The greater the main effect, the more the factor level mean differs from the overall mean response averaged over the factor levels for both factors.

Additive Factor Effects

The factor effects in Table 19.1 have an interesting property. Each mean response μ_{ij} can be obtained by adding the respective gender and age main effects to the overall mean $\mu_{\cdot\cdot}$. For instance, we have:

$$\mu_{11} = \mu_{\cdot\cdot} + \alpha_1 + \beta_1 = 12 + 0 + (-3) = 9$$
$$\mu_{23} = \mu_{\cdot\cdot} + \alpha_2 + \beta_3 = 12 + 0 + 4 = 16$$

In general, we have for Table 19.1a:

$$\mu_{ij} = \mu_{\cdot\cdot} + \alpha_i + \beta_j \qquad \text{Additive factor effects} \tag{19.7}$$

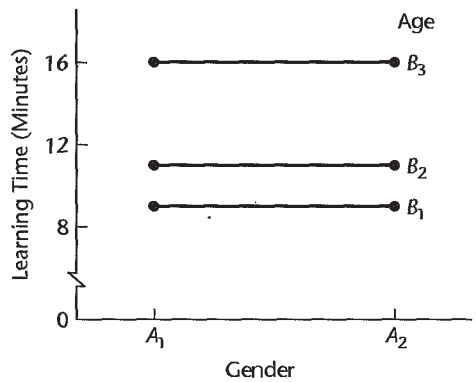
which can be expressed equivalently, using the definitions of α_i in (19.4) and of β_j in (19.5), as:

$$\mu_{ij} = \mu_{i\cdot} + \mu_{\cdot j} - \mu_{\cdot\cdot} \qquad \text{Additive factor effects} \tag{19.7a}$$

It can also be shown that each treatment mean μ_{ij} in Table 19.1a can be expressed in terms of three other treatment means:

$$\mu_{ij} = \mu_{ij'} + \mu_{i'j} - \mu_{i'j'} \qquad \text{Additive factor effects} \qquad i \neq i',\ j \neq j' \tag{19.7b}$$

FIGURE 19.3
Age Effect but No Gender Effect, with No Interactions—Learning Example.



For instance, we have:

$$\mu_{11} = \mu_{12} + \mu_{21} - \mu_{22} = 11 + 9 - 11 = 9$$

or:

$$\mu_{11} = \mu_{13} + \mu_{21} - \mu_{23} = 16 + 9 - 16 = 9$$

When all treatment means can be expressed in the form of (19.7), (19.7a), or (19.7b), we say that the *factors do not interact,* or that *no factor interactions are present,* or that the *factor effects are additive.* The significance of no factor interactions is that the effect of either factor does not depend on the level of the other factor. Consequently, the effects of the two factors can be described separately merely by analyzing the factor level means or the factor main effects. For instance, in the learning example in Table 19.1a, the two gender means signify that gender has no influence regardless of age, and the three age means portray the influence of age regardless of gender. The analysis of factor effects is therefore quite simple when there are no factor interactions.

Graphic Presentation. Figure 19.3 presents the mean learning times of Table 19.1a in the form of a *treatment means plot*—also known as an *interaction plot.* The X axis contains the gender factor levels (denoted by A_1 and A_2), and the Y axis contains learning time. Separate curves are drawn for each of the age factor levels (denoted by B_1, B_2, and B_3). The zero slope of each curve indicates that gender has no effect. The differences in the heights of the three curves show the age effects on learning time.

The points on each curve are conventionally connected by straight lines even though the variable on the X axis (gender, in our example) is not a continuous variable. When the variable on the X axis is qualitative, the slopes of the curves have no meaning, except when the slope is zero, which implies there are no factor level effects. If one of the two factors is a quantitative variable, it is ordinarily advisable to place that factor on the X scale.

Note that the treatment means plot in Figure 19.3 corresponds to a conditional effects plot in regression, such as the ones shown in Figure 8.7 on page 307. In each case, the effect of one variable is shown at different levels of the other variable.

A Second Example with Additive Factor Effects. Table 19.2a contains another illustration of factor effects that do not interact, for the same gender-age learning example as before. The situation here differs from that of Table 19.1a in that not only age but also

LE 19.2
e d nder Effects, No teractions— g ample.

(a) Mean Learning Times (in minutes)

	Factor *B*—Age			
Factor *A*—Gender	$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old	Row Average
$i = 1$ Male	11 (μ_{11})	13 (μ_{12})	18 (μ_{13})	14 ($\mu_{1\cdot}$)
$i = 2$ Female	7 (μ_{21})	9 (μ_{22})	14 (μ_{23})	10 ($\mu_{2\cdot}$)
Column average	9 ($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16 ($\mu_{\cdot 3}$)	12 ($\mu_{\cdot\cdot}$)

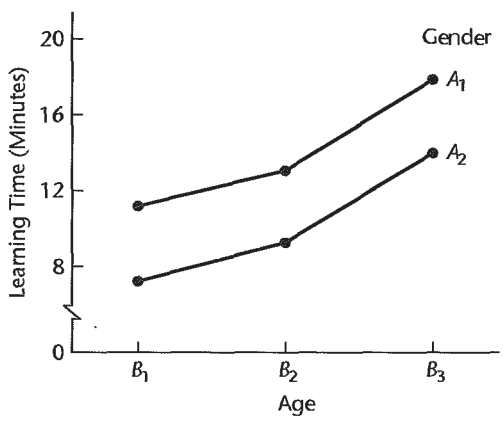
(b) Main Gender Effects (in minutes)

$$\alpha_1 = \mu_{1\cdot} - \mu_{\cdot\cdot} = 14 - 12 = 2$$
$$\alpha_2 = \mu_{2\cdot} - \mu_{\cdot\cdot} = 10 - 12 = -2$$

(c) Main Age Effects (in minutes)

$$\beta_1 = \mu_{\cdot 1} - \mu_{\cdot\cdot} = 9 - 12 = -3$$
$$\beta_2 = \mu_{\cdot 2} - \mu_{\cdot\cdot} = 11 - 12 = -1$$
$$\beta_3 = \mu_{\cdot 3} - \mu_{\cdot\cdot} = 16 - 12 = 4$$

FIGURE 19.4
Age and Gender Effects, with No Interactions— earning Example.



gender affects the learning time. This is evident from the fact that the mean learning times for men and women are not the same for any age group.

In Table 19.2a, as in Table 19.1a, every mean response can be decomposed according to (19.7):

$$\mu_{ij} = \mu_{\cdot\cdot} + \alpha_i + \beta_j$$

For instance:

$$\mu_{11} = \mu_{\cdot\cdot} + \alpha_1 + \beta_1 = 12 + 2 + (-3) = 11$$

Hence, the two factors do not interact, and the factor effects can be analyzed separately by examining the factor level means $\mu_{i\cdot}$ and $\mu_{\cdot j}$, respectively.

Figure 19.4 presents the data from Table 19.2a in the form of a treatment means plot. This time we have placed age on the X axis and used different curves for each gender. Note that the difference in the heights of the two curves reflects the gender difference and the departure from horizontal for each of the curves reflects the age effect. Furthermore, the two curves are parallel, which indicates that no two-factor interactions are present.

Equivalent Statements of Additive Factor Effects. We have said that two factors do not interact if *all* treatment means μ_{ij} can be expressed according to (19.7), (19.7a), or (19.7b). There are a number of other, equivalent, methods of recognizing when two factors do not interact. These are:

1. The difference between the mean responses for any two levels of factor B is the same for all levels of factor A. (For instance, in Table 19.2a, going from young to middle age leads to an increase of two minutes for both males and females, and going from middle age to old leads to an increase of five minutes for both males and females.) Note that it is *not* required that the changes, say, between levels 1 and 2 and between levels 2 and 3 of factor B are the same. These, of course, may differ depending upon the nature of the factor B effect.
2. The difference between the mean responses for any two levels of factor A is the same for all levels of factor B. (For instance, in Table 19.2a, going from male to female leads to a decrease of four minutes for all three age groups.)
3. The curves of the mean responses for the different levels of a factor are all parallel (such as the two gender curves in Figure 19.4).

All of these conditions are equivalent, implying that the two factors do not interact.

Interacting Factor Effects

Table 19.3a contains an illustration for the learning example where the factor effects do interact. The mean learning times for the different gender-age combinations in Table 19.3a indicate that gender has no effect on learning time for young persons but has a substantial effect for old persons. This differential influence of gender, which depends on the age of the person, implies that the age and gender factors interact in their effect on learning time.

TABLE 19.3 Age and Gender Effects, with Interactions—Learning Example.

(a) Mean Learning Times (in minutes)

	Factor B—Age				
Factor A—Gender	$j=1$ Young	$j=2$ Middle	$j=3$ Old	Row Average	Main Gender Effect
$i=1$ Male	9 (μ_{11})	12 (μ_{12})	18 (μ_{13})	13 ($\mu_{1\cdot}$)	1 (α_1)
$i=2$ Female	9 (μ_{21})	10 (μ_{22})	14 (μ_{23})	11 ($\mu_{2\cdot}$)	−1 (α_2)
Column average	9 ($\mu_{\cdot 1}$)	11 ($\mu_{\cdot 2}$)	16 ($\mu_{\cdot 3}$)	12 ($\mu_{\cdot\cdot}$)	
Main age effect	−3 (β_1)	−1 (β_2)	4 (β_3)		

(b) Interactions (in minutes)

	$j=1$	$j=2$	$j=3$	Row Average
$i=1$	−1	0	1	0
$i=2$	1	0	−1	0
Column average	0	0	0	0

Definition of Interaction. We can study the existence of interacting factor effects formally by examining whether or not all treatment means μ_{ij} can be expressed according to (19.7):

$$\mu_{ij} = \mu_{..} + \alpha_i + \beta_j$$

If they can, the factor effects are additive; otherwise, the factor effects are interacting.

For the learning example in Table 19.3a, the main factor effects α_i and β_j are shown in the margins of the table. It is clear that the factors interact. For instance, $\mu_{11} = 9$ while:

$$\mu_{..} + \alpha_1 + \beta_1 = 12 + 1 + (-3) = 10$$

If the two factors were additive, these would be the same.

The difference between the treatment mean μ_{ij} and the value $\mu_{..} + \alpha_i + \beta_j$ that would be expected if the two factors were additive is called the *interaction effect*, or more simply the *interaction*, of the ith level of factor A with the jth level of factor B, and is denoted by $(\alpha\beta)_{ij}$. Thus, we define $(\alpha\beta)_{ij}$ as follows:

$$(\alpha\beta)_{ij} = \mu_{ij} - (\mu_{..} + \alpha_i + \beta_j) \tag{19.8}$$

Replacing α_i and β_j by their definitions in (19.4) and (19.5), respectively, we obtain an alternative definition:

$$(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} \tag{19.8a}$$

To repeat, the interaction of the ith level of A with the jth level of B, denoted by $(\alpha\beta)_{ij}$, is simply the difference between the treatment mean μ_{ij} and the value that would be expected if the factors were additive. If in fact the two factors are additive, all interactions equal zero; i.e., $(\alpha\beta)_{ij} \equiv 0$.

The interactions for the learning example in Table 19.3a are shown in Table 19.3b. We have, for instance:

$$\begin{aligned}(\alpha\beta)_{13} &= \mu_{13} - (\mu_{..} + \alpha_1 + \beta_3)\\ &= 18 - (12 + 1 + 4)\\ &= 1\end{aligned}$$

Recognition of Interactions. We may recognize whether or not interactions are present in one of the following equivalent fashions:

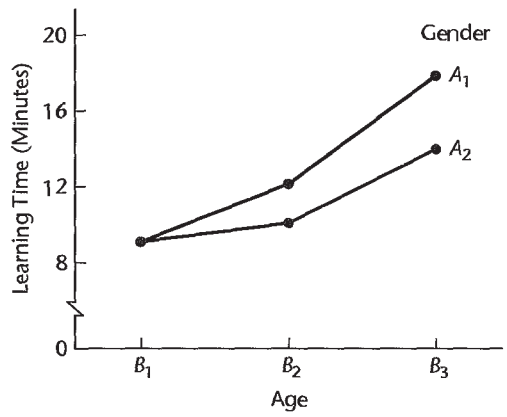
1. By examining whether all μ_{ij} can be expressed as the sums $\mu_{..} + \alpha_i + \beta_j$.

2. By examining whether the difference between the mean responses for any two levels of factor B is the same for all levels of factor A. (For instance, note in Table 19.3a that the mean learning time increases when going from young to middle-aged persons by three minutes for men but only by one minute for women.)

3. By examining whether the difference between the mean responses for any two levels of factor A is the same for all levels of factor B. (For instance, note in Table 19.3a that there is no difference between genders for young persons, but there is a difference of four minutes for old persons.)

4. By examining whether the treatment means curves for the different factor levels in a treatment means plot are parallel. (Figure 19.5 presents a plot of the treatment means in Table 19.3a, with age on the X axis. Note that the treatment means curves for the two genders are not parallel.)

FIGURE 19.5
Age and Gender Effects, with Important Interactions—Learning Example.



Comments

1. Note from Table 19.3b that some interactions are zero even though the two factors are interacting. *All* interactions must equal zero in order for the two factors to be additive.

2. Table 19.3b illustrates that interactions sum to zero when added over either rows or columns:

$$\sum_i (\alpha\beta)_{ij} = 0 \qquad j = 1, \ldots, b \tag{19.9a}$$

$$\sum_j (\alpha\beta)_{ij} = 0 \qquad i = 1, \ldots, a \tag{19.9b}$$

Consequently, the sum of all interactions is also zero:

$$\sum_i \sum_j (\alpha\beta)_{ij} = 0 \tag{19.9c}$$

We show this for (19.9a):

$$\begin{aligned}\sum_i (\alpha\beta)_{ij} &= \sum_{i=1}^{a} (\mu_{ij} - \mu_{\cdot\cdot} - \alpha_i - \beta_j) \\ &= \sum_i \mu_{ij} - a\mu_{\cdot\cdot} - \sum_i \alpha_i - a\beta_j\end{aligned}$$

Now $\sum_i \mu_{ij} = a\mu_{\cdot j}$ by (19.1) and $\sum \alpha_i = 0$ by (19.6). Finally, $\beta_j = \mu_{\cdot j} - \mu_{\cdot\cdot}$ by (19.5). Hence, we obtain:

$$\sum_i (\alpha\beta)_{ij} = a\mu_{\cdot j} - a\mu_{\cdot\cdot} - a(\mu_{\cdot j} - \mu_{\cdot\cdot}) = 0 \qquad \blacksquare$$

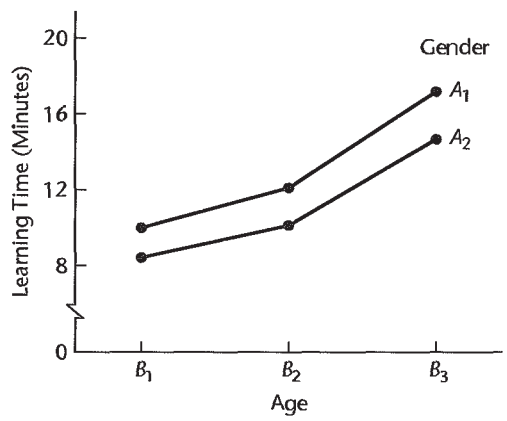
Important and Unimportant Interactions

When two factors interact, the question arises whether the factor level means are still meaningful measures. In Table 19.3a, for instance, it may well be argued that the gender factor level means 13 and 11 are misleading measures. They indicate that some difference exists in learning time for men and women, but that this difference is not too great. These factor level means hide the fact that there is no difference in mean learning time between

E 19.4
d
er Effects,
portant
actions—
ng
ple.

Factor *A*—Gender	Factor *B*—Age $j = 1$ Young	$j = 2$ Middle	$j = 3$ Old	Row Average
$i = 1$ Male	9.75	12.00	17.25	13.00
$i = 2$ Female	8.25	10.00	14.75	11.00
Column average	9.00	11.00	16.00	12.00

IGURE 19.6
ge and
Gender Effects,
th
mportant
teractions
curves almost
arallel)—
arning
xample.



genders for young persons, but there is a relatively large difference for old persons. The interactions in Table 19.3a would therefore be considered *important interactions,* implying that one should not ordinarily examine the effects of each factor separately in terms of the factor level means. A treatment means plot, such as in Figure 19.5, presents effectively a description of the nature of the interacting effects of the two factors.

Sometimes when two factors interact, the interaction effects are so small that they are considered to be *unimportant interactions.* Table 19.4 and Figure 19.6 present such a case. Note from Figure 19.6 that the curves are *almost* parallel. For practical purposes, one may say that the mean learning time for women is two minutes less than that for men, and this statement is approximately true for all age groups. Similarly, statements based on average learning time for different age groups will hold approximately for both genders.

Thus, in the case of unimportant interactions, the analysis of factor effects can proceed as for the case of no interactions. Each factor can be studied separately, based on the factor level means $\mu_{i\cdot}$ and $\mu_{\cdot j}$, respectively. This separate analysis of factor effects is, of course, much simpler than a joint analysis for the two factors based on the treatment means μ_{ij}, which is required when the interactions are important.

Comments

1. The determination of whether interactions are important or unimportant is admittedly sometimes difficult because it depends on the context of the application, just as the determination of whether an effect in a single-factor study is important. The subject area specialist (researcher) needs to play a prominent role in deciding whether an interaction is important or unimportant. The advantage of

unimportant (or no) interactions, namely, that one is then able to analyze the factor effects separatel is especially great when the study contains more than two factors.

2. Occasionally, it is meaningful to consider the effects of each factor in terms of the factor leve means even when important interactions are present. For example, two methods of teaching college mathematics (abstract and standard) were used in teaching students of excellent, good, and moderate quantitative ability. Important interactions between teaching method and student's quantitative abili were found to be present. Students with excellent quantitative ability tended to perform equally well with the two teaching methods, whereas students of moderate or good quantitative ability tended to perform better when taught by the standard method. If equal numbers of students with moderate, good, and excellent quantitative ability are to be taught by one of the two teaching methods, then the method that produces the best average result for all students might be of interest even in the presence of important interactions. A comparison of the teaching method factor level means would then be relevant, even though important interactions are present. ■

Transformable and Nontransformable Interactions

When important interactions exist, they are sometimes the result of the scale on which the response variable is measured. Consider, for instance, factor effects that act multiplicatively, rather than additively as in (19.7):

$$\mu_{ij} = \mu_{..}\alpha_i\beta_j \qquad \text{Multiplicative factor effects} \tag{19.10}$$

If we were to assume here that the factor effects are additive, we would find that condition (19.7) does not hold and therefore that interactions are present. These interactions can be removed, however, by applying a logarithmic transformation to (19.10):

$$\log \mu_{ij} = \log \mu_{..} + \log \alpha_i + \log \beta_j \tag{19.11}$$

This result can be restated equivalently as follows:

$$\mu'_{ij} = \mu'_{..} + \alpha'_i + \beta'_j \tag{19.11a}$$

where:

$$\begin{aligned} \mu'_{ij} &= \log \mu_{ij} \\ \mu'_{..} &= \log \mu_{..} \\ \alpha'_i &= \log \alpha_i \\ \beta'_j &= \log \beta_j \end{aligned}$$

The result in (19.11a) suggests that the original measurement scale for the response variable Y may not be the most appropriate one in the sense of leading to easily understood results. Rather, use of $Y' = \log Y$ for the response variable may be better, making the additive model (19.7) then more appropriate.

We say that the interactions present when the factor effects are actually multiplicative are *transformable interactions* because a simple transformation of Y will remove most of these interaction effects and thus make them unimportant.

Another instance of transformable interactions occurs when each interaction effect equals the product of functions of the main effects, for example:

$$\mu_{ij} = \alpha_i + \beta_j + 2\sqrt{\alpha_i}\sqrt{\beta_j} \qquad \text{Multiplicative interactions} \tag{19.12}$$

9.5
on of a
able
ction.

(a) Treatment Means—Original Scale			(b) Treatment Means after Square Root Transformation		
	Factor B			Factor B	
Factor A	$j=1$	$j=2$	Factor A	$j=1$	$j=2$
$i=1$	16	64	$i=1$	4	8
$i=2$	49	121	$i=2$	7	11
$i=3$	64	144	$i=3$	8	12

An equivalent form of (19.12) is:

$$\mu_{ij} = \left(\sqrt{\alpha_i} + \sqrt{\beta_j}\right)^2 \tag{19.12a}$$

If we now apply the square root transformation, we obtain an additive effects model:

$$\mu'_{ij} = \alpha'_i + \beta'_j \tag{19.13}$$

where:

$$\mu'_{ij} = \sqrt{\mu_{ij}}$$
$$\alpha'_i = \sqrt{\alpha_i}$$
$$\beta'_j = \sqrt{\beta_j}$$

Some simple transformations that may be helpful in making important interactions unimportant are the square, square root, logarithmic, and reciprocal transformations. When interactions cannot be largely removed by a transformation, they are called *nontransformable interactions*.

Table 19.5a contains an example of important interactions that are transformable. When a square root transformation is applied to these means, the resulting treatment means in Table 19.5b show no interacting effects. Ordinarily, of course, one cannot hope that a simple transformation of scale removes all interactions as in Table 19.5, but only that interactions become unimportant after the transformation.

Interpretation of Interactions

The interpretation of interactions can be quite difficult when the interacting effects are complex. There are many occasions, however, when the interactions have a simple structure, such as in Table 19.3a, so that the joint factor effects can be described in a straightforward manner. Table 19.6 provides several additional illustrations. The corresponding treatment means plots are shown in Figure 19.7.

In Table 19.6a and Figure 19.7a, we have a situation where either raising the pay or increasing the authority of low-paid executives with small authority leads to increased productivity. However, combining both higher pay and greater authority does not lead to any substantial further improvement in productivity than increasing either one alone. Table 19.6b and Figure 19.7b represent a case where both higher pay and greater authority are required before any substantial increase in productivity takes place.

TABLE 19.6
Examples of Different Types of Interactions.

(a) Productivity of Executives

	Factor *B*—Authority	
Factor *A*—Pay	Small	Great
Low	50	72
High	74	75

(b) Productivity of Executives

	Factor *B*—Authority	
Factor *A*—Pay	Small	Great
Low	50	52
High	53	75

(c) Productivity of Executives

	Factor *B*—Authority	
Factor *A*—Pay	Small	Great
Low	50	72
High	72	50

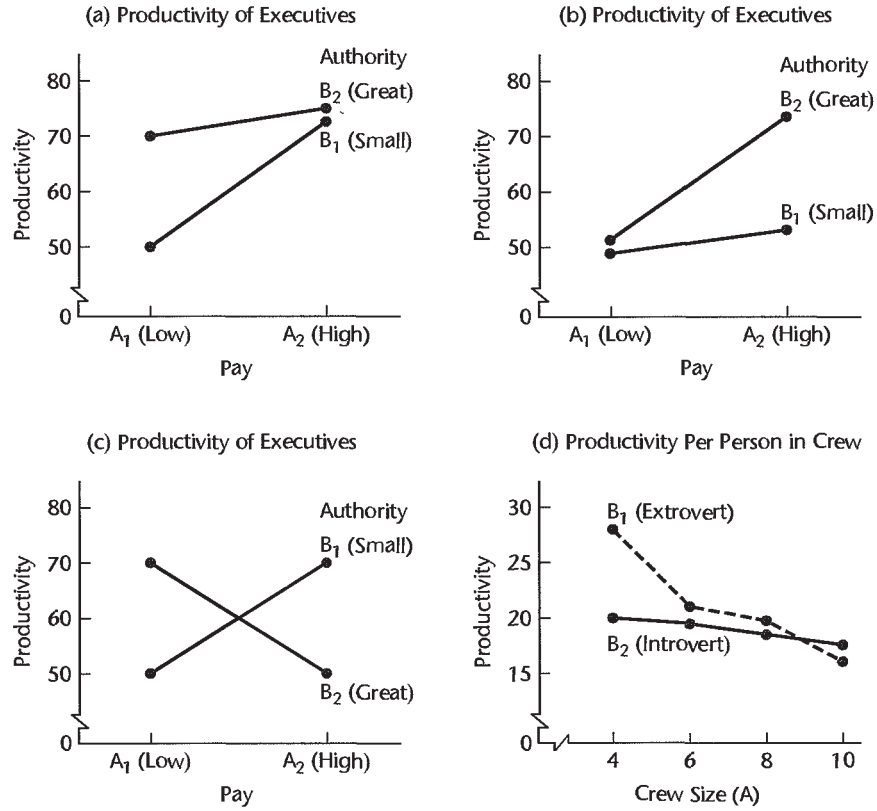
(d) Productivity per Person in Crew

	Factor *B*—Personality of Crew Chief	
Factor *A*—Crew Size	Extrovert	Introvert
4 persons	28	20
6 persons	22	20
8 persons	20	19
10 persons	17	18

It is possible that two factors interact, yet the main effects for one (or both) factors are zero. This would be the result of interactions in opposite directions that balance out over one (or both) factors. Thus, there would be definite factor effects, but these would not be disclosed by the factor level means. Table 19.6c and Figure 19.7c represent this situation where neither factor effect is present and the two factors interact. The case of interacting factors with no main effects for one (or both) factors fortunately is unusual. Typically, interaction effects are smaller than main effects.

Table 19.6d and Figure 19.7d portray a situation where size of crew and personality of crew chief interact in a complex fashion. Productivity with an extrovert crew chief and a crew of four is substantially larger than with an introvert crew chief. The advantage becomes small with crews of six and eight, and with a crew of 10 an introvert crew chief leads to a slightly larger productivity.

IGURE 19.7 Treatment eans Plots—xamples of teractions from able 19.6.



Comment

The terminology of reinforcement and interference interactions described in Chapter 8 for regression models where both predictor variables are quantitative is applicable to analysis of variance models if the two factors are quantitative or can be ordered on a measurement scale. In Figures 19.7a and 19.7b, pay level and authority both can be ordered on a scale. Hence, the interaction in Figure 19.7a can be described as an *interference* or *antagonistic* interaction (the slope decreases for higher levels of factor B), while that in Figure 19.7b can be described as a *reinforcement* or *synergistic* interaction (the slope increases for higher levels of factor B).

Similarly, the terminology of ordinal and disordinal interactions described in Chapter 8 for regression models where one predictor variable is quantitative and the other qualitative is applicable to analysis of variance models if one factor is quantitative or can be ordered on a measurement scale and the other factor is qualitative. In Figure 19.7d, crew size is a quantitative factor and personality is a qualitative factor. Therefore, the interaction in Figure 19.7d can be described as disordinal because the treatment means curves intersect. ■

19.3 Model I (Fixed Factor Levels) for Two-Factor Studies

Having explained the model elements, we are now ready to develop ANOVA model I with fixed factor levels for two-factor studies *when all treatment sample sizes are equal and all treatment means are of equal importance*. This ANOVA model is applicable to observational

studies and to experimental studies based on a completely randomized design. In Part VI we shall consider ANOVA models for some other experimental designs.

The basic situation is as follows: Factor A is studied at a levels, and these are of intrinsic interest in themselves; in other words, the a levels are not considered to be a sample from a larger population of factor A levels. Similarly, factor B is studied at b levels that are of intrinsic interest in themselves. All ab factor level combinations are included in the study. The number of cases for each of the ab treatments is the same, denoted by n, and it is required that $n > 1$. Thus, the total number of cases for the study is:

$$n_T = abn \tag{19.14}$$

The kth observation ($k = 1, \ldots, n$) for the treatment, where A is at the ith level, and B is at the jth level, is denoted by Y_{ijk} ($i = 1, \ldots, a$; $j = 1, \ldots, b$). Table 19.7 on page 833 illustrates this notation for an example where A is at three levels, B is at two levels, and two replications have been made for each treatment.

We shall state the fixed ANOVA model for two-factor studies in two equivalent versions—the cell means version and the factor effects version—and later will use one or the other as convenience dictates.

Cell Means Model

Model Formulation. When we regard the ab treatments without explicitly considering the factorial structure of the study, we express the analysis of variance model in terms of the cell (treatment) means μ_{ij}:

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk} \tag{19.15}$$

where:

μ_{ij} are parameters
ε_{ijk} are independent $N(0, \sigma^2)$
$i = 1, \ldots, a$; $j = 1, \ldots, b$; $k = 1, \ldots, n$

Important Features of Model. Some important features of the cell means model are:

1. The parameter μ_{ij} is the mean response for the treatment in which factor A is at the ith level and factor B is at the jth level. This follows because $E\{\varepsilon_{ijk}\} = 0$:

$$E\{Y_{ijk}\} = \mu_{ij} \tag{19.16}$$

2. Since μ_{ij} is a constant, the variance of Y_{ijk} is:

$$\sigma^2\{Y_{ijk}\} = \sigma^2\{\varepsilon_{ijk}\} = \sigma^2 \tag{19.17}$$

3. Since the error terms ε_{ijk} are independent and normally distributed, so are the observations Y_{ijk}. Hence, we can state ANOVA model (19.15) also as follows:

$$Y_{ijk} \text{ are independent } N(\mu_{ij}, \sigma^2) \tag{19.18}$$

4. ANOVA model (19.15) is a linear model because it can be expressed in the form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Consider a two-factor study with each factor having two levels (i.e., $a = b = 2$)

and two trials for each treatment (i.e., $n = 2$). Then $\mathbf{Y}$, $\mathbf{X}$, $\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}$ are defined as follows:

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{221} \\ \varepsilon_{222} \end{bmatrix} \tag{19.19}$$

Recall that the $\mathbf{E}\{\mathbf{Y}\}$ vector, which consists of the elements $E\{Y_{ijk}\}$, equals $\mathbf{X}\boldsymbol{\beta}$ according to (6.20). This vector here is:

$$\mathbf{E}\{\mathbf{Y}\} = \mathbf{X}\boldsymbol{\beta} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} = \begin{bmatrix} \mu_{11} \\ \mu_{11} \\ \mu_{12} \\ \mu_{12} \\ \mu_{21} \\ \mu_{21} \\ \mu_{22} \\ \mu_{22} \end{bmatrix} \tag{19.20}$$

Thus, $E\{Y_{ijk}\} = \mu_{ij}$, as it must according to (19.16), and we have the proper matrix representation for the two-factor ANOVA model (19.15):

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \begin{bmatrix} \mu_{11} \\ \mu_{11} \\ \mu_{12} \\ \mu_{12} \\ \mu_{21} \\ \mu_{21} \\ \mu_{22} \\ \mu_{22} \end{bmatrix} + \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{221} \\ \varepsilon_{222} \end{bmatrix} \tag{19.21}$$

In view of the error terms being independent with constant variance σ^2, the variance-covariance matrix of the error terms is $\boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\} = \sigma^2\mathbf{I}$, as in (16.9) for the single-factor ANOVA model. Also as before, we have $\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \boldsymbol{\sigma}^2\{\boldsymbol{\varepsilon}\}$ for two-factor ANOVA model (19.15).

5. ANOVA model (19.15) is therefore similar to the single-factor ANOVA model (16.2), except for the two subscripts now needed to identify the treatment. Normality, independent error terms, and constant variances for the error terms are properties of the ANOVA models for both single-factor and two-factor studies.

Factor Effects Model

Model Formulation. An equivalent version of cell means model (19.15) can be obtained by replacing each treatment mean μ_{ij} with an identical expression in terms of factor effects based on the definition of an interaction in (19.8):

$$(\alpha\beta)_{ij} = \mu_{ij} - (\mu_{..} + \alpha_i + \beta_j)$$

Rearranging terms, we obtain the identity:

$$\mu_{ij} \equiv \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} \tag{19.22}$$

where:

$$\begin{aligned} \mu_{..} &= \frac{\sum_i \sum_j \mu_{ij}}{ab} \\ \alpha_i &= \mu_{i.} - \mu_{..} \\ \beta_j &= \mu_{.j} - \mu_{..} \\ (\alpha\beta)_{ij} &= \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu_{..} \end{aligned}$$

This formulation indicates that each cell mean μ_{ij} can be viewed as the sum of four component factor effects. Specifically, (19.22) states that the mean response for the treatment where factor A is at the ith level and factor B is at the jth level is the sum of:

1. An overall mean $\mu_{...}$
2. The main effect α_i for factor A at the ith level.
3. The main effect β_j for factor B at the jth level.
4. The interaction effect $(\alpha\beta)_{ij}$ when factor A is at the ith level and factor B is at the jth level.

Replacing μ_{ij} in ANOVA model (19.15) by the expression in (19.22), we obtain an equivalent factor effects ANOVA model for two-factor studies:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \tag{19.23}$$

where:

$\mu_{..}$ is a constant
α_i are constants subject to the restriction $\sum \alpha_i = 0$
β_j are constants subject to the restriction $\sum \beta_j = 0$
$(\alpha\beta)_{ij}$ are constants subject to the restrictions:

$$\begin{aligned} \textstyle\sum_i (\alpha\beta)_{ij} &= 0 \qquad j = 1, \ldots, b \\ \textstyle\sum_j (\alpha\beta)_{ij} &= 0 \qquad i = 1, \ldots, a \end{aligned}$$

ε_{ijk} are independent $N(0, \sigma^2)$
$i = 1, \ldots, a;\ j = 1, \ldots, b;\ k = 1, \ldots, n$

Important Features of Model. Some important features of the factor effects model are:

1. ANOVA model (19.23) corresponds to the fixed factor effects ANOVA model (16.62) for a single-factor study except that the single-factor treatment effect is here replaced by the sum of a factor A effect, a factor B effect, and an interaction effect.

2. The properties of the observations Y_{ijk} for factor effects model (19.23) are the same as those for the equivalent cell means model (19.15). Since $E\{\varepsilon_{ijk}\} = 0$, we have:

$$E\{Y_{ijk}\} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} = \mu_{ij} \tag{19.24}$$

The second equality follows from identity (19.22). Further, we have:

$$\sigma^2\{Y_{ijk}\} = \sigma^2 \tag{19.25}$$

because the error term ε_{ijk} is the only random term on the right-hand side in (19.23) and $\sigma^2\{\varepsilon_{ijk}\} = \sigma^2$. Finally, the Y_{ijk} are independent normal random variables because the error terms are independent normal random variables. Hence, we can also state ANOVA model (19.23) as follows:

$$Y_{ijk} \text{ are independent } N[\mu.. + \alpha_i + \beta_j + (\alpha\beta)_{ij},\ \sigma^2] \qquad \textbf{(19.26)}$$

3. ANOVA model (19.23) is a linear model because it can be stated in the form $\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon}$. We shall show this explicitly in Section 23.2.

19.4 Analysis of Variance

Illustration

Table 19.7 contains an illustration that we shall employ in this chapter and the next. The Castle Bakery Company supplies wrapped Italian bread to a large number of supermarkets in a metropolitan area. An experimental study was made of the effects of height of the shelf display (factor A: bottom, middle, top) and the width of the shelf display (factor B: regular, wide) on sales of this bakery's bread during the experimental period (Y, measured in cases). Twelve supermarkets, similar in terms of sales volume and clientele, were utilized in the study. The six treatments were assigned at random to two stores each according to a completely randomized design, and the display of the bread in each store followed the treatment specifications for that store. Sales of the bread were recorded, and these results are presented in Table 19.7.

TABLE 19.7 Sample Data and Notation for Two-Factor Study—Castle Bakery Example (sales in cases).

Factor A (display height) i	Factor B (display width) j: B_1 (regular)	B_2 (wide)	Row Total	Display Height Average
A_1 (bottom)	47 (Y_{111}) 43 (Y_{112})	46 (Y_{121}) 40 (Y_{122})		
Total	90 ($Y_{11.}$)	86 ($Y_{12.}$)	176 ($Y_{1..}$)	
Average	45 ($\bar{Y}_{11.}$)	43 ($\bar{Y}_{12.}$)		44 ($\bar{Y}_{1..}$)
A_2 (middle)	62 (Y_{211}) 68 (Y_{212})	67 (Y_{221}) 71 (Y_{222})		
Total	130 ($Y_{21.}$)	138 ($Y_{22.}$)	268 ($Y_{2..}$)	
Average	65 ($\bar{Y}_{21.}$)	69 ($\bar{Y}_{22.}$)		67 ($\bar{Y}_{2..}$)
A_3 (top)	41 (Y_{311}) 39 (Y_{312})	42 (Y_{321}) 46 (Y_{322})		
Total	80 ($Y_{31.}$)	88 ($Y_{32.}$)	168 ($Y_{3..}$)	
Average	40 ($\bar{Y}_{31.}$)	44 ($\bar{Y}_{32.}$)		42 ($\bar{Y}_{3..}$)
Column total	300 ($Y_{.1.}$)	312 ($Y_{.2.}$)	612 ($Y_{...}$)	
Display width average	50 ($\bar{Y}_{.1.}$)	52 ($\bar{Y}_{.2.}$)		51 ($\bar{Y}_{...}$)

Notation

Table 19.7 illustrates the notation we shall use for two-factor studies. It is a straightforward extension of the notation for single-factor studies. An observation is denoted by Y_{ijk}. The subscripts i and j specify the levels of factors A and B, respectively, and the subscript k refers to the given case or trial for a particular treatment (i.e., factor level combination).

A dot in the subscript indicates aggregation or averaging over the variable represented by the index. For instance, the sum of the observations for the treatment corresponding to the ith level of factor A and the jth level of factor B is:

$$Y_{ij\cdot} = \sum_{k=1}^{n} Y_{ijk} \tag{19.27a}$$

The corresponding mean is:

$$\bar{Y}_{ij\cdot} = \frac{Y_{ij\cdot}}{n} \tag{19.27b}$$

The total of all observations for the ith factor level of A is:

$$Y_{i\cdot\cdot} = \sum_{j}^{b} \sum_{k}^{n} Y_{ijk} \tag{19.27c}$$

and the corresponding mean is:

$$\bar{Y}_{i\cdot\cdot} = \frac{Y_{i\cdot\cdot}}{bn} \tag{19.27d}$$

Similarly, for the jth factor level of B the sum of all observations and their mean are denoted by:

$$Y_{\cdot j\cdot} = \sum_{i}^{a} \sum_{k}^{n} Y_{ijk} \tag{19.27e}$$

$$\bar{Y}_{\cdot j\cdot} = \frac{Y_{\cdot j\cdot}}{an} \tag{19.27f}$$

Finally, the sum of all observations in the study is:

$$Y_{\cdot\cdot\cdot} = \sum_{i}^{a} \sum_{j}^{b} \sum_{k}^{n} Y_{ijk} \tag{19.27g}$$

and the overall mean is:

$$\bar{Y}_{\cdot\cdot\cdot} = \frac{Y_{\cdot\cdot\cdot}}{nab} \tag{19.27h}$$

Fitting of ANOVA Model

Cell Means Model (19.15). Fitting the two-factor cell means model (19.15) to the sample data by either the method of least squares or the method of maximum likelihood leads to minimizing the criterion:

$$Q = \sum_{i} \sum_{j} \sum_{k} (Y_{ijk} - \mu_{ij})^2 \tag{19.28}$$

When we perform the minimization of Q, we obtain the least squares and maximum likelihood estimators:

$$\hat{\mu}_{ij} = \bar{Y}_{ij\cdot} \tag{19.29}$$

Thus, the *fitted values* are the estimated treatment means:

$$\hat{Y}_{ijk} = \bar{Y}_{ij\cdot} \tag{19.30}$$

The *residuals,* as usual, are defined as the difference between the observed and fitted values:

$$e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - \bar{Y}_{ij\cdot} \tag{19.31}$$

Residuals are highly useful for assessing the appropriateness of two-factor ANOVA model (19.15), as they also are for the statistical models considered earlier.

Factor Effects Model (19.23). For the equivalent factor effects model (19.23), the least squares and maximum likelihood methods both lead to minimizing the criterion:

$$Q = \sum_i \sum_j \sum_k [Y_{ijk} - \mu_{\cdot\cdot} - \alpha_i - \beta_j - (\alpha\beta)_{ij}]^2 \tag{19.32}$$

subject to the restrictions:

$$\sum_i \alpha_i = 0 \qquad \sum_j \beta_j = 0 \qquad \sum_i (\alpha\beta)_{ij} = 0 \qquad \sum_j (\alpha\beta)_{ij} = 0$$

When we perform this minimization, we obtain the following least squares and maximum likelihood estimators of the parameters:

Parameter	Estimator	
$\mu_{\cdot\cdot}$	$\hat{\mu}_{\cdot\cdot} = \bar{Y}_{\cdot\cdot\cdot}$	(19.33a)
$\alpha_i = \mu_{i\cdot} - \mu_{\cdot\cdot}$	$\hat{\alpha}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}$	(19.33b)
$\beta_j = \mu_{\cdot j} - \mu_{\cdot\cdot}$	$\hat{\beta}_j = \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot}$	(19.33c)
$(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot}$	$\widehat{(\alpha\beta)}_{ij} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot}$	(19.33d)

The correspondences of these estimators to the definitions of the parameters are readily apparent.

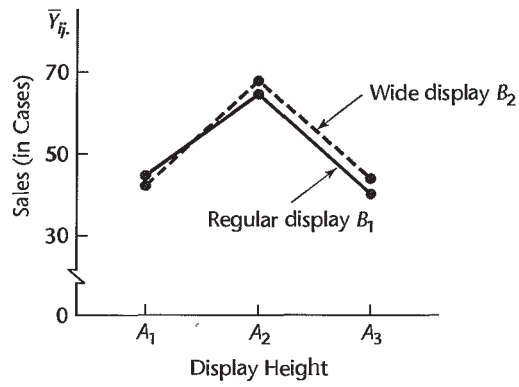
The fitted values and residuals for factor effects model (19.23) are exactly the same as those for cell means model (19.15). Specifically, the fitted values for ANOVA model (19.23) are:

$$\hat{Y}_{ijk} = \bar{Y}_{\cdot\cdot\cdot} + (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}) + (\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot}) + (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot}) = \bar{Y}_{ij\cdot} \tag{19.34}$$

so that the residuals are again:

$$e_{ijk} = Y_{ijk} - \bar{Y}_{ij\cdot} \tag{19.35}$$

FIGURE 19.8 Estimated Treatment Means Plot—Castle Bakery Example.



Example

For the Castle Bakery example, the fitted values, i.e., the estimated treatment means $\bar{Y}_{ij.}$, are shown in Table 19.7. A plot of these estimated treatment means is presented in Figure 19.8. We see from this estimated treatment means plot that, for both display widths, mean sales for the middle display height are substantially larger than those for the other two display heights. The effect of display width does not appear to be large. Indeed, there may be no effect of display width; the variations between the estimated treatment means for any given display height may be solely of a random nature. In that event, there would be no interactions between display height and display width in their effects on sales.

Figure 19.8 differs from the earlier treatment means plots because the earlier figures presented the true treatment means μ_{ij}, while Figure 19.8 presents sample estimates. We therefore need to test whether or not the effects shown in Figure 19.8 are real effects or represent only random variations. To conduct these tests, we require a partitioning of the total sum of squares, to be discussed next.

Partitioning of Total Sum of Squares

Partitioning of Total Deviation. We shall partition the total deviation of an observation Y_{ijk} from the overall mean $\bar{Y}_{...}$ in two stages. First, we shall obtain a decomposition of the total deviation $Y_{ijk} - \bar{Y}_{...}$ by viewing the study as consisting of ab treatments:

$$\underbrace{Y_{ijk} - \bar{Y}_{...}}_{\text{Total deviation}} = \underbrace{\bar{Y}_{ij.} - \bar{Y}_{...}}_{\text{Deviation of estimated treatment mean around overall mean}} + \underbrace{Y_{ijk} - \bar{Y}_{ij.}}_{\text{Deviation around estimated treatment mean}} \tag{19.36}$$

Note that the deviation around the estimated treatment mean is simply the residual e_{ijk} in (19.35):

$$e_{ijk} = Y_{ijk} - \bar{Y}_{ij.}$$

Treatment and Error Sums of Squares. When we square (19.36) and sum over all cases, the cross-product term drops out and we obtain:

$$SSTO = SSTR + SSE \tag{19.37}$$

where:

$$SSTO = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2 \tag{19.37a}$$

$$SSTR = n \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{...})^2 \tag{19.37b}$$

$$SSE = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2 = \sum_i \sum_j \sum_k e_{ijk}^2 \tag{19.37c}$$

SSTR reflects the variability between the *ab* estimated treatment means and is the ordinary *treatment sum of squares,* and *SSE* reflects the variability within treatments and is the usual *error sum of squares.* The only difference between these formulas and those for the single-factor case is the use of the two subscripts i and j to designate a treatment.

Example

For the Castle Bakery example, the decomposition of the total sum of squares in (19.37) is obtained as follows, using the data in Table 19.7:

$$SSTO = (47 - 51)^2 + (43 - 51)^2 + (46 - 51)^2 + \cdots + (46 - 51)^2 = 1{,}642$$
$$SSTR = 2[(45 - 51)^2 + (43 - 51)^2 + (65 - 51)^2 + \cdots + (44 - 51)^2] = 1{,}580$$
$$SSE = (47 - 45)^2 + (43 - 45)^2 + (46 - 43)^2 + \cdots + (46 - 44)^2 = 62$$

Partitioning of Treatment Sum of Squares. Next, we shall decompose the estimated treatment mean deviation $\bar{Y}_{ij.} - \bar{Y}_{...}$ in terms of components reflecting the factor A main effect, the factor B main effect, and the AB interaction effect:

$$\underbrace{\bar{Y}_{ij.} - \bar{Y}_{...}}_{\substack{\text{Deviation of}\\ \text{estimated treatment}\\ \text{mean around}\\ \text{overall mean}}} = \underbrace{\bar{Y}_{i..} - \bar{Y}_{...}}_{\substack{A \text{ main}\\ \text{effect}}} + \underbrace{\bar{Y}_{.j.} - \bar{Y}_{...}}_{\substack{B \text{ main}\\ \text{effect}}} + \underbrace{\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}}_{\substack{AB \text{ interaction}\\ \text{effect}}} \tag{19.38}$$

When we square (19.38) and sum over all treatments and over the n cases associated with each estimated treatment mean $\bar{Y}_{ij.}$, all cross-product terms drop out and we obtain:

$$SSTR = SSA + SSB + SSAB \tag{19.39}$$

where:

$$SSA = nb \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2 \tag{19.39a}$$

$$SSB = na \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2 \tag{19.39b}$$

$$SSAB = n \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 \tag{19.39c}$$

The interaction sum of squares can also be obtained as a remainder:

$$SSAB = SSTO - SSE - SSA - SSB \tag{19.39d}$$

or from:

$$SSAB = SSTR - SSA - SSB \tag{19.39e}$$

where *SSTO* and *SSTR* are given in (19.37a) and (19.37b), respectively.

SSA, called the *factor A sum of squares,* measures the variability of the estimated factor *A* level means $\bar{Y}_{i..}$. The more variable they are, the bigger will be *SSA*. Similarly, *SSB*, called the *factor B sum of squares,* measures the variability of the estimated factor *B* level means $\bar{Y}_{.j.}$. Finally, *SSAB*, called the *AB interaction sum of squares,* measures the variability of the estimated interactions $\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}$ for the *ab* treatments. Since the mean of all estimated interactions is zero, the deviations of the estimated interactions around their mean is not explicitly shown, as it was in *SSA* and *SSB*. The larger absolutely are the estimated interactions, the larger will be *SSAB*.

The partitioning of *SSTR* into the components *SSA*, *SSB*, and *SSAB* is called an *orthogonal decomposition*. An orthogonal decomposition is one where the component sums of squares add to the total sum of squares (*SSTR* here), and likewise for the degrees of freedom. Thus, the decompositions of *SSTO* into *SSTR* and *SSE* for single-factor and two-factor studies are also orthogonal decompositions. While many different orthogonal decompositions of *SSTR* are possible here, the one into the *SSA*, *SSB*, and *SSAB* components is of interest because these three components provide information about the factor *A* main effects, the factor *B* main effects, and the *AB* interactions, respectively, as will be seen shortly.

Example

For the Castle Bakery example, we obtain the following decomposition of *SSTR*, using the data in Table 19.7 and the formulas in (19.39):

$$\begin{aligned} SSA &= 2(2)[(44-51)^2 + (67-51)^2 + (42-51)^2] = 1{,}544 \\ SSB &= 2(3)[(50-51)^2 + (52-51)^2] = 12 \\ SSAB &= 1{,}580 - 1{,}544 - 12 = 24 \end{aligned}$$

Hence, we have:

$$\begin{aligned} 1{,}580 &= 1{,}544 + 12 + 24 \\ SSTR &= SSA + SSB + SSAB \end{aligned}$$

Combined Partitioning. Combining the decompositions in (19.37) and (19.39), we have established that:

$$SSTO = SSA + SSB + SSAB + SSE \tag{19.40}$$

where the component sums of squares are defined in (19.37) and (19.39).

Example

For the Castle Bakery example, we have found:

$$\begin{aligned} 1{,}642 &= 1{,}544 + 12 + 24 + 62 \\ SSTO &= SSA + SSB + SSAB + SSE \end{aligned}$$

Thus, much of the total variability in this instance is associated with the factor *A* (display height) effects.

Partitioning of Degrees of Freedom

We are familiar from single-factor analysis of variance with how the degrees of freedom are divided between the treatment and error components. For two-factor studies with n cases for each treatment, there are a total of $n_T = nab$ cases and $r = ab$ treatments; hence, the degrees of freedom associated with *SSTO*, *SSTR*, and *SSE* are $nab - 1$, $ab - 1$, and $nab - ab = (n - 1)ab$, respectively. These degrees of freedom for the Castle Bakery example are $2(3)(2) - 1 = 11$, $3(2) - 1 = 5$, and $(2 - 1)(3)(2) = 6$, respectively.

Corresponding to the further partitioning of the treatment sum of squares in (19.39), we can also obtain a breakdown of the associated $ab - 1$ degrees of freedom. *SSA* has $a - 1$ degrees of freedom associated with it. There are a factor level deviations $\bar{Y}_{i..} - \bar{Y}_{...}$, but one degree of freedom is lost because the deviations are subject to one restriction, i.e., $\sum(\bar{Y}_{i..} - \bar{Y}_{...}) = 0$. Similarly, *SSB* has $b - 1$ degrees of freedom associated with it. The degrees of freedom associated with *SSAB*, the interaction sum of squares, is the remainder:

$$(ab - 1) - (a - 1) - (b - 1) = (a - 1)(b - 1)$$

The degrees of freedom associated with *SSAB* may be understood as follows: There are ab interaction terms. These are subject to b restrictions since:

$$\sum_i (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) = 0 \qquad j = 1, \ldots, b$$

There are a additional restrictions since:

$$\sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...}) = 0 \qquad i = 1, \ldots, a$$

However, only $a - 1$ of these latter restrictions are independent since the last one is implied by the previous b restrictions. Altogether, therefore, there are $b + (a - 1)$ independent restrictions. Hence, the degrees of freedom are:

$$ab - (b + a - 1) = (a - 1)(b - 1)$$

Example

For the Castle Bakery example, *SSA* has $3 - 1 = 2$ degrees of freedom associated with it, *SSB* has $2 - 1 = 1$ degree of freedom, and *SSAB* has $(3 - 1)(2 - 1) = 2$ degrees of freedom.

Mean Squares

Mean squares are obtained in the usual way by dividing the sums of squares by their associated degrees of freedom. We thus obtain:

$$MSA = \frac{SSA}{a - 1} \tag{19.41a}$$

$$MSB = \frac{SSB}{b - 1} \tag{19.41b}$$

$$MSAB = \frac{SSAB}{(a - 1)(b - 1)} \tag{19.41c}$$

Example For the Castle Bakery example, these mean squares are:

$$MSA = \frac{1{,}544}{2} = 772$$

$$MSB = \frac{12}{1} = 12$$

$$MSAB = \frac{24}{2} = 12$$

Expected Mean Squares

It can be shown, along the same lines used for single-factor ANOVA, that the mean squares for two-factor ANOVA model (19.23) have the following expectations:

$$E\{MSE\} = \sigma^2 \tag{19.42a}$$

$$E\{MSA\} = \sigma^2 + nb\frac{\sum \alpha_i^2}{a-1} = \sigma^2 + nb\frac{\sum(\mu_{i\cdot} - \mu_{\cdot\cdot})^2}{a-1} \tag{19.42b}$$

$$E\{MSB\} = \sigma^2 + na\frac{\sum \beta_j^2}{b-1} = \sigma^2 + na\frac{\sum(\mu_{\cdot j} - \mu_{\cdot\cdot})^2}{b-1} \tag{19.42c}$$

$$E\{MSAB\} = \sigma^2 + n\frac{\sum\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)} \tag{19.42d}$$

$$= \sigma^2 + n\frac{\sum\sum(\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot})^2}{(a-1)(b-1)}$$

These expectations show that if there are no factor A main effects (i.e., if all $\mu_{i\cdot}$ are equal, or all $\alpha_i = 0$), *MSA* and *MSE* have the same expectation; otherwise *MSA* tends to be larger than *MSE*. Similarly, if there are no factor B main effects, *MSB* and *MSE* have the same expectation; otherwise *MSB* tends to be larger than *MSE*. Finally, if there are no interactions [i.e., if all $(\alpha\beta)_{ij} = 0$] so that the factor effects are additive, *MSAB* has the same expectation as *MSE*; otherwise, *MSAB* tends to be larger than *MSE*. This suggests that F^* test statistics based on the ratios *MSA/MSE*, *MSB/MSE*, and *MSAB/MSE* will provide information about the main effects and interactions of the two factors, with large values of the test statistics indicating the presence of factor effects. We shall see shortly that tests based on these statistics are regular F tests.

Analysis of Variance Table

The decomposition of the total sum of squares in (19.40) into the several factor and error components is shown in Table 19.8. Also shown there are the associated degrees of freedom, the mean squares, and the expected mean squares. Table 19.9 contains the two-factor analysis of variance for the Castle Bakery example.

Figure 19.9 presents MINITAB output for the Castle Bakery example. The first output block shows ANOVA results similar to those presented in Table 19.9. The second block presents various estimated means.

TABLE 19.8 ANOVA Table for Two-Factor Study with Fixed Factor Levels.

Source of Variation	SS	df	MS	E{MS}
Factor A	$SSA = nb\sum(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$a-1$	$MSA = \frac{SSA}{a-1}$	$\sigma^2 + bn\frac{\sum(\mu_{i\cdot} - \mu_{\cdot\cdot})^2}{a-1}$
Factor B	$SSB = na\sum(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$b-1$	$MSB = \frac{SSB}{b-1}$	$\sigma^2 + an\frac{\sum(\mu_{\cdot j} - \mu_{\cdot\cdot})^2}{b-1}$
AB interactions	$SSAB = n\sum\sum(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$	$(a-1)(b-1)$	$MSAB = \frac{SSAB}{(a-1)(b-1)}$	$\sigma^2 + n\frac{\sum\sum(\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot})^2}{(a-1)(b-1)}$
Error	$SSE = \sum\sum\sum(Y_{ijk} - \bar{Y}_{ij\cdot})^2$	$ab(n-1)$	$MSE = \frac{SSE}{ab(n-1)}$	σ^2
Total	$SSTO = \sum\sum\sum(Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$	$nab-1$		

TABLE 19.9 ANOVA Table for Two-Factor Study—Castle Bakery Example.

Source of Variation	*SS*	*df*	*MS*
Factor *A* (display height)	1,544	2	772
Factor *B* (display width)	12	1	12
AB interactions	24	2	12
Error	62	6	10.3
Total	1,642	11	

FIGURE 19.9 MINITAB Computer Output for Two-Factor Analysis of Variance—Castle Bakery Example.

Analysis of Variance for Cases Sold

Source	DF	SS	MS	F	P
Height	2	1544.00	772.00	74.71	0.000
Width	1	12.00	12.00	1.16	0.323
Height*Width	2	24.00	12.00	1.16	0.375
Error	6	62.00	10.33		
Total	11	1642.00			

Means

Height	N	Cases So
1	4	44.000
2	4	67.000
3	4	42.000

Width	N	Cases So
1	6	50.000
2	6	52.000

Height	Width	N	Cases So
1	1	2	45.000
1	2	2	43.000
2	1	2	65.000
2	2	2	69.000
3	1	2	40.000
3	2	2	44.000

19.5 Evaluation of Appropriateness of ANOVA Model

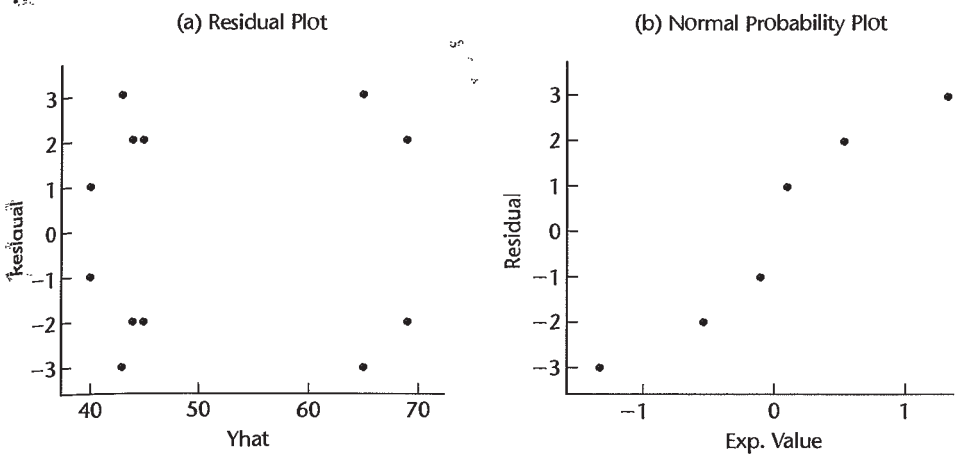
Before undertaking formal inference procedures, we need to evaluate the appropriateness of two-factor ANOVA model (19.23). No new problems arise here. The residuals in (19.35):

$$e_{ijk} = Y_{ijk} - \bar{Y}_{ij.}$$

are examined for normality, constancy of error variance, and independence of error terms in the same fashion as for a single-factor study.

Weighted least squares is a standard remedial measure when the error terms are normally distributed but do not have constant variance. When both the assumptions of normality and constancy of the error variance are violated, a transformation of the response variable may be sought to stabilize the error variance and to bring the distribution of the error terms closer to a normal distribution. Our discussion of these topics in Chapter 18 for single-factor ANOVA applies completely to two-factor ANOVA.

FIGURE 19.10 MINITAB Diagnostic Residual Plots—Castle Bakery Example.



Our earlier discussion on the effects of departures from the single-factor ANOVA model applies fully to two-factor ANOVA. In particular, the employment of equal sample sizes for each treatment minimizes the effect of unequal error variances.

Example

In the Castle Bakery example, there are only two replications for each treatment. Also, the data are rounded to keep the illustrative computations simple. As a result, the analysis of residuals will only be of limited value here. The residuals are obtained according to (19.35). Using the data in Table 19.7, we have, for instance:

$$e_{111} = 47 - 45 = 2$$
$$e_{121} = 46 - 43 = 3$$

A plot of the residuals against the fitted values $\hat{Y}_{ijk} = \bar{Y}_{ij\cdot}$ is presented in Figure 19.10a. There is no strong evidence of unequal error variances for the different treatments here. A normal probability plot of the residuals is presented in Figure 19.10b. The plot is moderately linear; the fact that only six plot points are visible is due to the rounded nature of the data. The coefficient of correlation between the ordered residuals and their expected values under normality is .966, which tends to support the reasonableness of approximate normality.

On the basis of these diagnostics and since the inference procedures for ANOVA model (19.23) are robust, it appears to be reasonable to proceed with tests for factor effects and other inference procedures.

19.6 *F* Tests

In view of the additivity of sums of squares and degrees of freedom, Cochran's theorem (2.61) applies when no factor effects are present. Hence, the F^* test statistics based on the appropriate mean squares then follow the F distribution, leading to the usual type of F tests for factor effects.

Test for Interactions

Ordinarily, the analysis of a two-factor study begins with a test to determine whether or not the two factors interact:

$$\begin{aligned} H_0&: \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot} = 0 \qquad \text{for all } i, j \\ H_a&: \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot} \neq 0 \qquad \text{for some } i, j \end{aligned} \tag{19.43}$$

or equivalently:

$$\begin{aligned} H_0&: \text{all } (\alpha\beta)_{ij} = 0 \\ H_a&: \text{not all } (\alpha\beta)_{ij} \text{ equal zero} \end{aligned} \tag{19.43a}$$

As we noted from an examination of the expected mean squares in Table 19.8, the appropriate test statistic is:

$$F^* = \frac{MSAB}{MSE} \tag{19.44}$$

Large values of F^* indicate the existence of interactions. When H_0 holds, F^* is distributed as $F[(a-1)(b-1), (n-1)ab]$. Hence, the appropriate decision rule to control the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; (a-1)(b-1), (n-1)ab], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; (a-1)(b-1), (n-1)ab], \text{ conclude } H_a \end{aligned} \tag{19.45}$$

where $F[1-\alpha; (a-1)(b-1), (n-1)ab]$ is the $(1-\alpha)100$ percentile of the appropriate F distribution.

Test for Factor *A* Main Effects

Tests for factor A main effects and for factor B main effects ordinarily follow the test for interactions when no important interactions exist. To test whether or not A main effects are present:

$$\begin{aligned} H_0&: \mu_{1\cdot} = \mu_{2\cdot} = \cdots = \mu_{a\cdot} \\ H_a&: \text{not all } \mu_{i\cdot} \text{ are equal} \end{aligned} \tag{19.46}$$

or equivalently:

$$\begin{aligned} H_0&: \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0 \\ H_a&: \text{not all } \alpha_i \text{ equal zero} \end{aligned} \tag{19.46a}$$

we use the test statistic:

$$F^* = \frac{MSA}{MSE} \tag{19.47}$$

Again, large values of F^* indicate the existence of factor A main effects. Since F^* is distributed as $F[a-1, (n-1)ab]$ when H_0 holds, the appropriate decision rule for controlling the risk of making a Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; a-1, (n-1)ab], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; a-1, (n-1)ab], \text{ conclude } H_a \end{aligned} \tag{19.48}$$

Test for Factor *B* Main Effects

This test is similar to the one for factor A main effects. The alternatives are:

$$H_0: \mu_{\cdot 1} = \mu_{\cdot 2} = \cdots = \mu_{\cdot b}$$
$$H_a: \text{not all } \mu_{\cdot j} \text{ are equal} \tag{19.49}$$

or equivalently:

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_b = 0$$
$$H_a: \text{not all } \beta_j \text{ equal zero} \tag{19.49a}$$

The test statistic is:

$$F^* = \frac{MSB}{MSE} \tag{19.50}$$

and the appropriate decision rule for controlling the risk of a Type I error at α is:

$$\text{If } F^* \leq F[1-\alpha; b-1, (n-1)ab], \text{ conclude } H_0$$
$$\text{If } F^* > F[1-\alpha; b-1, (n-1)ab], \text{ conclude } H_a \tag{19.51}$$

Example

We shall investigate in the Castle Bakery example the presence of display height and display width effects, using a level of significance of $\alpha = .05$ for each test. First, we begin by testing whether or not interaction effects are present:

$$H_0: \text{all } (\alpha\beta)_{ij} = 0$$
$$H_a: \text{not all } (\alpha\beta)_{ij} \text{ equal zero}$$

Using the ANOVA results from Table 19.9 in test statistic (19.44), we obtain:

$$F^* = \frac{12}{10.3} = 1.17$$

For $\alpha = .05$, we require $F(.95; 2, 6) = 5.14$, so that the decision rule is:

$$\text{If } F^* \leq 5.14, \text{ conclude } H_0$$
$$\text{If } F^* > 5.14, \text{ conclude } H_a$$

Since $F^* = 1.17 \leq 5.14$, we conclude H_0, that display height and display width do not interact in their effects on sales. The P-value of this test is $P\{F(2, 6) > 1.17\} = .37$.

Since the two factors do not interact, we turn to test for display height (factor A) main effects; the alternative conclusions are given in (19.46). Test statistic (19.47) for our example becomes:

$$F^* = \frac{772}{10.3} = 75.0$$

For $\alpha = .05$, we require $F(.95; 2, 6) = 5.14$. Since $F^* = 75.0 > 5.14$, we conclude H_a, that the factor A level means $\mu_{i\cdot}$ are not equal, or that some definite effects associated with height of display level exist. The P-value of this test is $P\{F(2, 6) > 75.0\} = .0001$.

Next, we test for display width (factor B) main effects; the alternative conclusions are given in (19.49). Test statistic (19.50) becomes for our example:

$$F^* = \frac{12}{10.3} = 1.17$$

For $\alpha = .05$, we require $F(.95; 1, 6) = 5.99$. Since $F^* = 1.17 \leq 5.99$, we conclude H_0, that all $\mu_{.j}$ are equal, or that display width has no effect on sales. The P-value of this test is $P\{F(1, 6) > 1.17\} = .32$.

Thus, the analysis of variance tests confirm the impressions from the estimated treatment means plot in Figure 19.8 that only display height has an effect on sales for the treatments studied. At this point, it is clearly desirable to conduct further analyses of the nature of the display height effects. We shall discuss analyses of the nature of the factor effects in Sections 19.8 and 19.9.

Kimball Inequality

If the test for interactions is conducted with level of significance α_1, that for factor A main effects with level of significance α_2, and that for factor B main effects with level of significance α_3, the level of significance α for the *family* of three tests is greater than the individual levels of significance. From the Bonferroni inequality in (4.4), we can derive the inequality:

$$\alpha \leq \alpha_1 + \alpha_2 + \alpha_3 \tag{19.52}$$

For the case considered here, a somewhat tighter inequality can be used, the *Kimball inequality*, which utilizes the fact that the numerators of the three test statistics are independent and the denominator is the same in each case. This inequality states:

$$\alpha \leq 1 - (1 - \alpha_1)(1 - \alpha_2)(1 - \alpha_3) \tag{19.53}$$

For the Castle Bakery example, where $\alpha_1 = \alpha_2 = \alpha_3 = .05$, the Bonferroni inequality yields as the bound for the family level of significance:

$$\alpha \leq .05 + .05 + .05 = .15$$

while the Kimball inequality yields the bound:

$$\alpha \leq 1 - (.95)(.95)(.95) = .143$$

This illustration makes it clear that the level of significance for the family of three tests may be substantially higher than the levels of significance for the individual tests.

Comment

The F^* test statistics in (19.44), (19.47), and (19.50) can be obtained by the general linear test approach explained in Chapter 2. For example, in testing for the presence of interaction effects, the alternatives are those given in (19.43) and the full model is ANOVA model (19.23):

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad \text{Full model} \tag{19.54}$$

Fitting this full model leads to the fitted values $\hat{Y}_{ijk} = \bar{Y}_{ij.}$ and the error sum of squares:

$$SSE(F) = \sum\sum\sum(Y_{ijk} - \hat{Y}_{ijk})^2 = \sum\sum\sum(Y_{ijk} - \bar{Y}_{ij.})^2 = SSE \tag{19.55}$$

which is the usual ANOVA error sum of squares in (19.37c). This error sum of squares has $ab(n-1)$ degrees of freedom associated with it.

The reduced model under H_0: $(\alpha\beta)_{ij} \equiv 0$ is:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \varepsilon_{ijk} \qquad \text{Reduced model} \tag{19.56}$$

It can be shown that the fitted values for the reduced model are $\hat{Y}_{ijk} = \bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{...}$, so that the error sum of squares for the reduced model is:

$$SSE(R) = \sum\sum\sum(Y_{ijk} - \hat{Y}_{ijk})^2 = \sum\sum\sum(Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 \tag{19.57}$$

This error sum of squares can be shown to have $nab - a - b + 1$ degrees of freedom associated with it. Test statistic (2.70) then simplifies to $F^* = MSAB/MSE$ in (19.44). ■

19.7 Strategy for Analysis

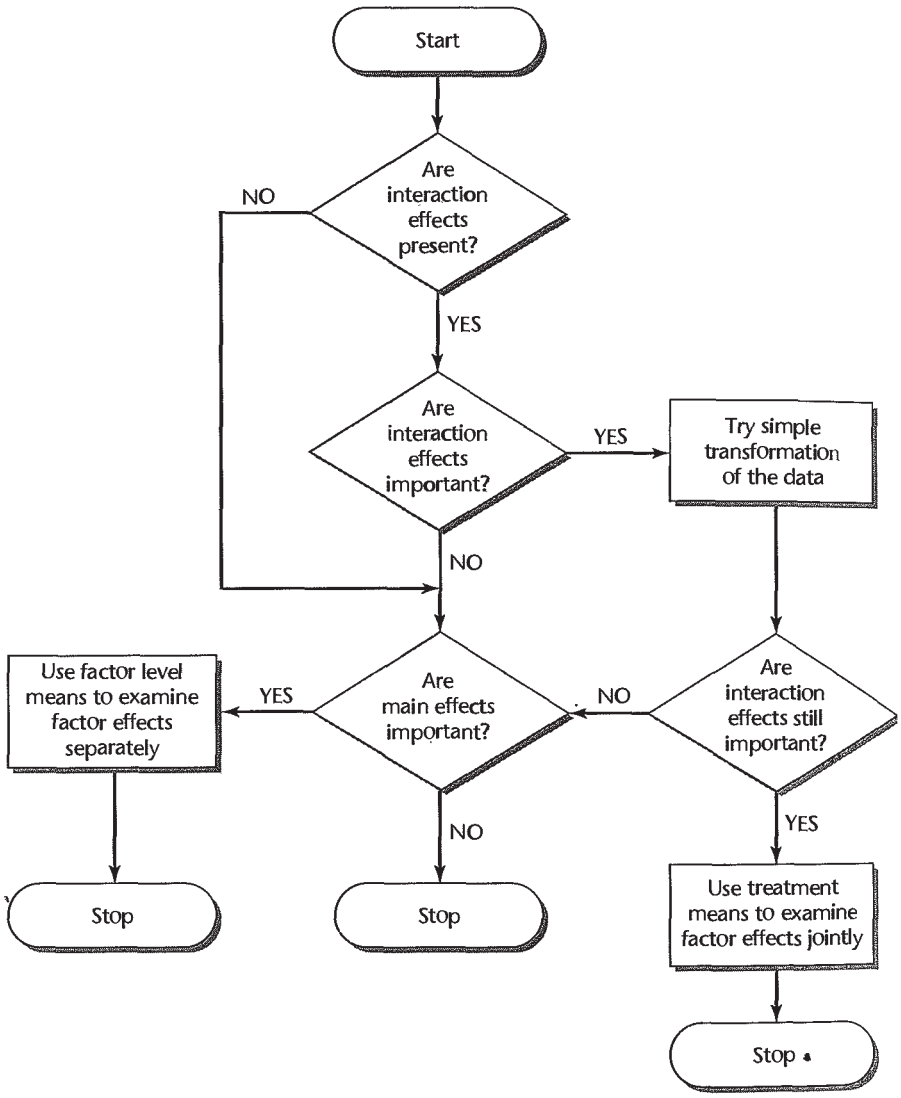
Scientific inquiry is often guided by the principle that the simplest explanations of observed phenomena tend to be the most effective. Data analysis is guided by this principle, seeking to obtain a simple, clear explanation of the data. In the context of ANOVA studies, additive effects provide a much simpler explanation of factor effects than do interacting effects. The presence of interacting effects complicates the explanation of the factor effects because they must then be described in terms of the *combined* effects of the two factors. Of course, some phenomena are complex so that the factor effects cannot be described simply by additive effects. The desire for a simple, parsimonious explanation, when possible, suggests the following basic strategy for analyzing factor effects in two-factor studies:

1. Examine whether the two factors interact.
2. If they do not interact, examine whether the main effects for factors A and B are important. For important A or B main effects, describe the nature of these effects in terms of the factor level means $\mu_{i.}$ or $\mu_{.j}$, respectively. In some special cases, there may also be interest in the treatment means μ_{ij}.
3. If the factors do interact, examine if the interactions are important or unimportant.
4. If the interactions are unimportant, proceed as in step 2.
5. If the interactions are important, consider whether they can be made unimportant by a meaningful simple transformation of scale. If so, make the transformation and proceed as in step 2.
6. For important interactions that cannot be made unimportant by a simple transformation, analyze the two factor effects jointly in terms of the treatment means μ_{ij}. In some special cases, there may also be interest in the factor level means $\mu_{i.}$ and $\mu_{.j}$.

A flowchart of this strategy is presented in Figure 19.11.

We have already discussed the testing for interaction effects, the possible diminution of important interactions by a meaningful simple transformation, as well as how to test for the presence of factor main effects. Now we turn to steps 2 and 6 of the strategy for analysis, namely, how to compare factor level means $\mu_{i.}$ or $\mu_{.j}$ when there are no interactions or only unimportant ones, and how to compare treatment means μ_{ij} when there are important interactions. We begin with a discussion of the analysis of factor effects when the factors do not interact or interact only in an unimportant fashion.

FIGURE 19.11
Strategy for Analysis of Two-Factor Studies.



19.8 Analysis of Factor Effects when Factors Do Not Interact

As just noted, the analysis of factor effects usually only involves the factor level mea and $\mu_{\cdot j}$ when the two factors do not interact, or when they interact only in an unimp fashion.

Estimation of Factor Level Mean

Unbiased point estimators of $\mu_{i\cdot}$ and $\mu_{\cdot j}$ are:

$$\hat{\mu}_{i\cdot} = \bar{Y}_{i\cdot\cdot}$$

$$\hat{\mu}_{\cdot j} = \bar{Y}_{\cdot j\cdot}$$

where $\bar{Y}_{i..}$ and $\bar{Y}_{.j.}$ are defined in (19.27d) and (19.27f), respectively. The variance of $\bar{Y}_{i..}$ is:

$$\sigma^2\{\bar{Y}_{i..}\} = \frac{\sigma^2}{bn} \tag{19.58a}$$

since $\bar{Y}_{i..}$ contains bn independent observations, each with variance σ^2. Similarly, we have:

$$\sigma^2\{\bar{Y}_{.j.}\} = \frac{\sigma^2}{an} \tag{19.58b}$$

Unbiased estimators of these variances are obtained by replacing σ^2 with MSE:

$$s^2\{\bar{Y}_{i..}\} = \frac{MSE}{bn} \tag{19.59a}$$

$$s^2\{\bar{Y}_{.j.}\} = \frac{MSE}{an} \tag{19.59b}$$

Confidence limits for $\mu_{i.}$ and $\mu_{.j}$ utilize, as usual, the t distribution:

$$\bar{Y}_{i..} \pm t[1-\alpha/2; (n-1)ab]s\{\bar{Y}_{i..}\} \tag{19.60a}$$

$$\bar{Y}_{.j.} \pm t[1-\alpha/2; (n-1)ab]s\{\bar{Y}_{.j.}\} \tag{19.60b}$$

The degrees of freedom $(n-1)ab$ are those associated with MSE.

Estimation of Contrast of Factor Level Means

A contrast among the factor level means $\mu_{i.}$:

$$L = \sum c_i \mu_{i.} \qquad \text{where } \sum c_i = 0 \tag{19.61}$$

is estimated unbiasedly by:

$$\hat{L} = \sum c_i \bar{Y}_{i..} \tag{19.62}$$

Because of the independence of the $\bar{Y}_{i..}$, the variance of this estimator is:

$$\sigma^2\{\hat{L}\} = \sum c_i^2 \sigma^2\{\bar{Y}_{i..}\} = \frac{\sigma^2}{bn} \sum c_i^2 \tag{19.63}$$

An unbiased estimator of this variance is:

$$s^2\{\hat{L}\} = \frac{MSE}{bn} \sum c_i^2 \tag{19.64}$$

Finally, the appropriate $1-\alpha$ confidence limits for L are:

$$\hat{L} \pm t[1-\alpha/2; (n-1)ab]s\{\hat{L}\} \tag{19.65}$$

To estimate a contrast among the factor level means $\mu_{.j}$:

$$L = \sum c_j \mu_{.j} \qquad \text{where } \sum c_j = 0 \tag{19.66}$$

we use the estimator:

$$\hat{L} = \sum c_j \bar{Y}_{.j.} \tag{19.67}$$

whose estimated variance is:

$$s^2\{\hat{L}\} = \frac{MSE}{an}\sum c_j^2 \tag{19.68}$$

The $1-\alpha$ confidence limits for L in (19.65) are still appropriate, with $\hat{L}$ and $s\{\hat{L}\}$ now defined in (19.67) and (19.68), respectively.

Estimation of Linear Combination of Factor Level Means

A linear combination of the factor level means $\mu_{i\cdot}$:

$$L = \sum c_i \mu_{i\cdot} \tag{19.69}$$

is estimated unbiasedly by $\hat{L}$ in (19.62). The variance of this estimator is given in (19.63), and an unbiased estimator of this variance is given in (19.64). The appropriate $1-\alpha$ confidence limits for L are given in (19.65).

Analogous results follow for a linear combination of the factor level means $\mu_{\cdot j}$:

$$L = \sum c_j \mu_{\cdot j} \tag{19.70}$$

Multiple Pairwise Comparisons of Factor Level Means

Usually, more than one pairwise comparison is of interest, and the multiple comparison procedures discussed in Chapter 17 for single-factor ANOVA studies can be employed with only minor modifications for two-factor studies. If all or a large number of pairwise comparisons among the factor level means $\mu_{i\cdot}$ or $\mu_{\cdot j}$ are to be made, the Tukey procedure of Section 17.5 is appropriate. When only a few pairwise comparisons are to be made that are specified in advance of the analysis, the Bonferroni procedure of Section 17.7 may be best. Often, tests for differences between pairs of factor level means precede the construction of interval estimates so that the analysis of the interval estimates can be confined to active comparisons. Finally, when a large number of comparisons among the factor-level means is of interest, the Scheffé method is usually preferred.

Tukey Procedure. The Tukey multiple comparison confidence limits for all pairwise comparisons:

$$D = \mu_{i\cdot} - \mu_{i'\cdot} \tag{19.71}$$

with family confidence coefficient of at least $1-\alpha$ are:

$$\hat{D} \pm Ts\{\hat{D}\} \tag{19.72}$$

where:

$$\hat{D} = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot} \tag{19.72a}$$

$$s^2\{\hat{D}\} = \frac{2MSE}{bn} \tag{19.72b}$$

$$T = \frac{1}{\sqrt{2}} q[1-\alpha; a, (n-1)ab] \tag{19.72c}$$

To use the Tukey procedure to conduct all simultaneous tests of the form:

$$\begin{aligned} H_0&: D = \mu_{i\cdot} - \mu_{i'\cdot} = 0 \\ H_a&: D = \mu_{i\cdot} - \mu_{i'\cdot} \neq 0 \end{aligned} \tag{19.73}$$

the test statistic and decision rule are:

$$q^* = \frac{\sqrt{2}\hat{D}}{s\{\hat{D}\}}; \qquad \text{If } |q^*| > q[1-\alpha; a, (n-1)ab], \text{ conclude } H_a \tag{19.73a}$$

For conciseness in this chapter, we state only the portion of the decision rule leading to conclusion H_a. As for single-factor ANOVA, the family level of significance for all pairwise tests here is $1-\alpha$; in other words, the probability of concluding that there exist any pairwise differences when there are none is α.

For pairwise comparisons of the factor level means $\mu_{\cdot j}$, the only changes are:

$$D = \mu_{\cdot j} - \mu_{\cdot j'} \tag{19.74}$$

$$\hat{D} = \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot j'\cdot} \tag{19.75}$$

$$s^2\{\hat{D}\} = \frac{2MSE}{an} \tag{19.76}$$

$$T = \frac{1}{\sqrt{2}} q[1-\alpha; b, (n-1)ab] \tag{19.77}$$

$$q^* = \frac{\sqrt{2}\hat{D}}{s\{\hat{D}\}}; \qquad \text{If } |q^*| > q[1-\alpha; b, (n-1)ab], \text{ conclude } H_a \tag{19.78}$$

Bonferroni Procedure. When only a few pairwise comparisons specified in advance are to be made, the Bonferroni method may be best. The simultaneous estimation formulas above still apply, with the Tukey multiple T replaced by the Bonferroni multiple B:

$$B = t[1-\alpha/2g; (n-1)ab] \tag{19.79}$$

where g is the number of statements in the family.

To test simultaneously each of g pairwise differences with the Bonferroni procedure, the test statistic and decision rule are:

$$t^* = \frac{\hat{D}}{s\{\hat{D}\}}; \qquad \text{If } |t^*| > t[1-\alpha/2g; (n-1)ab], \text{ conclude } H_a \tag{19.80}$$

Combined Factor *A* and Factor *B* Family. When important factor A and factor B effects both are present, it is often desired to have a family confidence coefficient $1-\alpha$, or family significance level α, for the joint set of pairwise comparisons involving *both* factor A and factor B means. The Bonferroni method can be used directly for this purpose, with g representing the total number of statements in the joint set.

Alternatively, the Bonferroni method can be used in conjunction with the Tukey method. To illustrate this use, if the pairwise comparisons for factor A are made with the Tukey procedure with a family confidence coefficient of .95, and likewise for the pairwise comparisons for factor B, the Bonferroni inequality then assures us that the family confidence coefficient for the joint set of comparisons for both factors is at least .90.

Multiple Contrasts of Factor Level Means

Scheffé Procedure. When a large number of contrasts among the factor level mean $\mu_{i\cdot}$ or $\mu_{\cdot j}$ are of interest, the Scheffé method should be used. If the contrasts involve the $\mu_{i\cdot}$ as in (19.61), the Scheffé confidence limits are:

$$\hat{L} \pm Ss\{\hat{L}\} \tag{19.81}$$

where:

$$S^2 = (a-1)F[1-\alpha; a-1, (n-1)ab] \tag{19.81a}$$

and $\hat{L}$ is given by (19.62) and $s^2\{\hat{L}\}$ is given by (19.64). The probability is then $1-\alpha$ that every confidence interval (19.81) in the family of all possible contrasts is correct. If the contrasts involve the $\mu_{\cdot j}$ as in (19.66), $\hat{L}$ is given by (19.67), $s^2\{\hat{L}\}$ is given by (19.68), and the Scheffé multiple in (19.81) is defined by:

$$S^2 = (b-1)F[1-\alpha; b-1, (n-1)ab] \tag{19.81b}$$

When the Scheffé procedure is employed to conduct simultaneous tests of the form:

$$\begin{aligned} H_0&: L = 0 \\ H_a&: L \neq 0 \end{aligned} \tag{19.82}$$

for contrasts involving the factor level means $\mu_{i\cdot}$, the test statistic and decision rule are:

$$F^* = \frac{\hat{L}^2}{(a-1)s^2\{\hat{L}\}}; \qquad \text{If } F^* > F[1-\alpha; a-1, (n-1)ab], \text{ conclude } H_a \tag{19.82a}$$

When the contrasts involve the factor level means $\mu_{\cdot j}$, the test statistic and decision rule are:

$$F^* = \frac{\hat{L}^2}{(b-1)s^2\{\hat{L}\}}; \qquad \text{If } F^* > F[1-\alpha; b-1, (n-1)ab], \text{ conclude } H_a \tag{19.82b}$$

Bonferroni Procedure. When the number of contrasts of interest is small and has been specified in advance, the Bonferroni procedure may be best. Confidence limits (19.81) are modified by replacing the Scheffé multiple S with the Bonferroni multiple B:

$$B = t[1-\alpha/2g; (n-1)ab] \tag{19.83}$$

where g is the number of statements in the family.

Simultaneous testing of g tests with the Bonferroni procedure is based on the following test statistic and decision rule:

$$t^* = \frac{\hat{L}}{s\{\hat{L}\}}; \qquad \text{If } |t^*| > t[1-\alpha/2g; (n-1)ab], \text{ conclude } H_a \tag{19.84}$$

Combined Factor *A* and Factor *B* Family. When important factor A and factor B effects are present and contrasts for each of the two factors are of interest, it is often desired that the inference procedure provide assurance for the combined family of factor A and factor B contrasts. Several possibilities exist to accomplish this:

1. The Bonferroni method may be used directly, with g representing the total number of statements in the joint set.

2. The Bonferroni method can be used to join the two sets of Scheffé multiple comparison families in the same way explained earlier for joining two Tukey sets.
3. The Scheffé confidence limits (19.81) can be modified to use the S multiple defined by:

$$S^2 = (a+b-2)F[1-\alpha; a+b-2, (n-1)ab] \qquad \textbf{(19.85)}$$

For simultaneous testing, the test statistics and decision rules in (19.82a) and (19.82b) can be replaced by:

$$F^* = \frac{\hat{L}^2}{(a+b-2)s^2\{\hat{L}\}}; \qquad \text{If } F^* > F[1-\alpha; a+b-2, (n-1)ab], \text{ conclude } H_a \qquad \textbf{(19.86)}$$

Estimates Based on Treatment Means

Occasionally in analyzing the factor effects in a two-factor study when no interactions are present, there is interest in particular treatment means μ_{ij}. For example, in a two-factor study of the effects of price and type of advertisement on sales, interest may exist in estimating the mean sales for two different price levels when a particular advertisement is used. In such cases, the methods of analysis for single-factor studies discussed in Chapter 17 are appropriate. The number of treatments now is simply $r = ab$, the degrees of freedom associated with MSE are $n_T - r = nab - ab = (n-1)ab$, and the estimated treatment means are $\bar{Y}_{ij\cdot}$, based on n observations each.

Example 1—Pairwise Comparisons of Factor Level Means

In the Castle Bakery, the estimated treatment means plot in Figure 19.8 suggested that no interaction effects are present and that display width may not have any effect. The formal analysis of variance based on Table 19.9 supported both of these conclusions. Our interest now is in examining the nature of the display height effects in more detail.

First, we shall obtain a preliminary view of the display height and width effects by plotting bar graphs of the estimated factor level means in Table 19.7. Figure 19.12a contains a bar graph of the estimated factor A level means $\bar{Y}_{i\cdot\cdot}$. For comparison, we show in Figure 19.12b a similar plot for the estimated factor B level means $\bar{Y}_{\cdot j\cdot}$. Figure 19.12a suggests that level 2 of factor A (middle shelf display height) leads to significantly larger sales than the other

FIGURE 19.12 **Bar Graphs of Estimated Factor Level Means—Castle Bakery Example.**

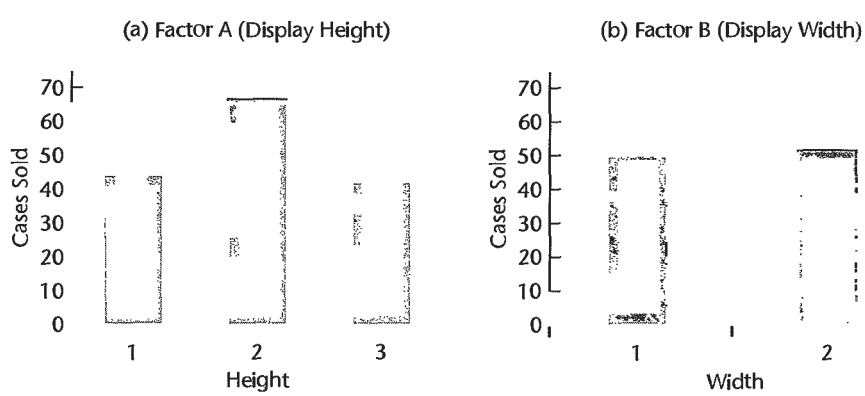


TABLE 19.10 Pairwise Testing of Factor *A* Level Means—Castle Bakery Example.

(1) Alternatives	(2) Test Statistic (19.73a)	(3) Decision Rule Conclude H_a if $\lvert q^* \rvert >$	(4) Conclusion
H_0: $D_1 = \mu_{2\cdot} - \mu_{1\cdot} = 0$ H_a: $D_1 = \mu_{2\cdot} - \mu_{1\cdot} \neq 0$	$q^* = \frac{\sqrt{2}(23)}{2.27} = 14.33$	$q(.95; 3, 6) = 4.34$	H_a
H_0: $D_2 = \mu_{1\cdot} - \mu_{3\cdot} = 0$ H_a: $D_2 = \mu_{1\cdot} - \mu_{3\cdot} \neq 0$	$q^* = \frac{\sqrt{2}(2)}{2.27} = 1.25$	$q(.95; 3, 6) = 4.34$	H_0
H_0: $D_3 = \mu_{2\cdot} - \mu_{3\cdot} = 0$ H_a: $D_3 = \mu_{2\cdot} - \mu_{3\cdot} \neq 0$	$q^* = \frac{\sqrt{2}(25)}{2.27} = 15.58$	$q(.95; 3, 6) = 4.34$	H_a

two factor levels. In addition, Figure 19.12a also suggests that the mean sales for display height levels 1 and 3 may not be different from each other.

Turning now to formal inference procedures, we shall first test simultaneously all pairwise differences among the shelf height means, using the Tukey multiple comparison procedure with family significance level $\alpha = .05$. The alternatives to be tested for the comparisons of display height means ($i = 1$—bottom, 2—middle, 3—top) are shown in Table 19.10, column 1. From Tables 19.7 and 19.9 we obtain the following information:

$$\hat{D}_1 = \bar{Y}_{2\cdot\cdot} - \bar{Y}_{1\cdot\cdot} = 67 - 44 = 23 \qquad MSE = 10.3$$
$$a = 3$$
$$\hat{D}_2 = \bar{Y}_{1\cdot\cdot} - \bar{Y}_{3\cdot\cdot} = 44 - 42 = 2 \qquad b = 2$$
$$n = 2$$
$$\hat{D}_3 = \bar{Y}_{2\cdot\cdot} - \bar{Y}_{3\cdot\cdot} = 67 - 42 = 25 \qquad (n-1)ab = 6$$

Hence, by (19.72b) we obtain:

$$s^2\{\hat{D}_1\} = s^2\{\hat{D}_2\} = s^2\{\hat{D}_3\} = \frac{2(10.3)}{2(2)} = 5.15$$

so that $s\{\hat{D}_1\} = s\{\hat{D}_2\} = s\{\hat{D}_3\} = 2.27$. The test statistics and decision rules based on (19.73a) are given in Table 19.10, columns 2 and 3, and the conclusions from the tests are shown in column 4.

It can be concluded from the tests in Table 19.10 with family significance level $\alpha = .05$ that for the product studied and the types of stores in the experiment, the middle shelf height is far better than either the bottom or the top heights, and that the latter two do not differ significantly in sales effectiveness. All of these conclusions are covered by the family significance level of .05.

Next, we wish to estimate how much greater are mean sales at the middle shelf height than at either of the other two shelf heights. We shall continue to use the Tukey multiple comparison procedure because the two pairwise comparisons now of interest are the result of the earlier testing of all pairwise comparisons. From our previous work, we have:

$$\hat{D}_1 = \bar{Y}_{2\cdot\cdot} - \bar{Y}_{1\cdot\cdot} = 23 \qquad \hat{D}_3 = \bar{Y}_{2\cdot\cdot} - \bar{Y}_{3\cdot\cdot} = 25 \qquad s\{\hat{D}_1\} = s\{\hat{D}_3\} = 2.27$$

We also require, from (19.72):

$$q(.95; 3, 6) = 4.34$$

$$T = \frac{4.34}{\sqrt{2}} = 3.07$$

$$Ts\{\hat{D}_1\} = Ts\{\hat{D}_3\} = 3.07(2.27) = 7.0$$

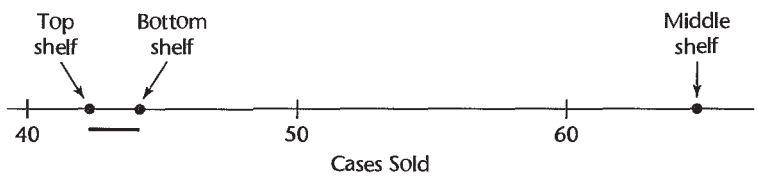
We therefore find the following confidence intervals for the two pairwise comparisons of the shelf height factor level means:

$$16 = 23 - 7.0 \leq \mu_{2\cdot} - \mu_{1\cdot} \leq 23 + 7.0 = 30$$

$$18 = 25 - 7.0 \leq \mu_{2\cdot} - \mu_{3\cdot} \leq 25 + 7.0 = 32$$

With family confidence coefficient of .95, we conclude that mean sales for the middle shelf height exceed those for the bottom shelf height by between 16 and 30 cases and those for the top shelf height by between 18 and 32 cases.

We can summarize the effects of shelf height on mean sales by the following line plot:



Example 2—Estimation of Treatment Means

The manager of a supermarket that has sales volume and clientele similar to the supermarkets included in the Castle Bakery study has room only for the regular shelf display width, and wishes to obtain estimates of mean sales for the middle and top shelf heights. We shall now obtain interval estimates with a 90 percent family confidence coefficient using the Bonferroni procedure.

From Tables 19.7 and 19.9, we have:

$$\bar{Y}_{21\cdot} = 65 \qquad \bar{Y}_{31\cdot} = 40 \qquad MSE = 10.3$$

Hence, we obtain:

$$s^2\{\bar{Y}_{21\cdot}\} = s^2\{\bar{Y}_{31\cdot}\} = \frac{MSE}{n} = \frac{10.3}{2} = 5.15$$

$$s\{\bar{Y}_{21\cdot}\} = s\{\bar{Y}_{31\cdot}\} = 2.27$$

For $g = 2$, we require $B = t[1 - \alpha/2g; (n-1)ab] = t(.975; 6) = 2.447$. Thus, we obtain the confidence limits:

$$65 \pm 2.447(2.27) \qquad 40 \pm 2.447(2.27)$$

and the desired confidence intervals are:

$$59.4 \leq \mu_{21} \leq 70.6 \qquad 34.4 \leq \mu_{31} \leq 45.6$$

19.9 Analysis of Factor Effects when Interactions Are Important

When important interactions exist that cannot be made unimportant by a simple transformation, the analysis of factor effects generally must be based on the treatment means μ_{ij}. Typically, this analysis will involve estimation of multiple comparisons of treatment means or single degree of freedom tests. Furthermore, one often compares the levels of one factor across levels of the other factor, referred to as the comparison of simple effects. For example, in a 2×3 factorial structure study, we compare individual cell means within levels of each factor, e.g., $\mu_{11} = \mu_{12} = \mu_{13}$ and $\mu_{21} = \mu_{22} = \mu_{23}$ and/or $\mu_{11} = \mu_{21}$, $\mu_{12} = \mu_{22}$, and $\mu_{13} = \mu_{23}$.

Multiple Pairwise Comparisons of Treatment Means

If pairs of treatment means μ_{ij} are to be compared, either the Tukey or the Bonferroni multiple comparison procedure may be used, depending on which is more advantageous. In effect, the analysis is equivalent to that for single-factor ANOVA, with the total number of treatments here equal to $r = ab$, the degrees of freedom associated with MSE here equal to $n_T - r = (n-1)ab$, and each estimated treatment mean, now denoted by $\bar{Y}_{ij\cdot}$, based on n cases.

Tukey Procedure. The Tukey $1-\alpha$ multiple comparison confidence limits for all pairwise comparisons:

$$D = \mu_{ij} - \mu_{i'j'} \qquad i, j \neq i', j' \tag{19.87}$$

are:

$$\hat{D} \pm Ts\{\hat{D}\} \tag{19.88}$$

where:

$$\hat{D} = \bar{Y}_{ij\cdot} - \bar{Y}_{i'j'\cdot} \tag{19.88a}$$

$$s^2\{\hat{D}\} = \frac{2MSE}{n} \tag{19.88b}$$

$$T = \frac{1}{\sqrt{2}} q[1-\alpha; ab, (n-1)ab] \tag{19.88c}$$

The test statistic and decision rule for all simultaneous Tukey tests of the form:

$$\begin{aligned} H_0&: D = 0 \\ H_a&: D \neq 0 \end{aligned} \tag{19.89}$$

are as follows when the family significance level is controlled at α:

$$q^* = \frac{\sqrt{2}\hat{D}}{s\{\hat{D}\}}; \qquad \text{If } |q^*| > q[1-\alpha; ab, (n-1)ab], \text{ conclude } H_a \tag{19.89a}$$

Bonferroni Procedure. If the Bonferroni method is employed for a family of g comparisons, the multiple T in confidence interval (19.88) is replaced by:

$$B = t[1-\alpha/2g; (n-1)ab] \tag{19.90}$$

and the test statistic and decision rule in (19.89a) become:

$$t^* = \frac{\hat{D}}{s\{\hat{D}\}}; \qquad \text{If } |t^*| > t[1-\alpha/2g; (n-1)ab], \text{ conclude } H_a \tag{19.91}$$

Multiple Contrasts of Treatment Means

Scheffé Procedure. The Scheffé multiple comparison procedure for single-factor studies is directly applicable to the estimation of contrasts involving the treatment means μ_{ij}. The joint confidence limits for contrasts of the form:

$$L = \sum\sum c_{ij}\mu_{ij} \qquad \text{where } \sum\sum c_{ij} = 0 \tag{19.92}$$

are:

$$\hat{L} \pm Ss\{\hat{L}\} \tag{19.93}$$

where:

$$\hat{L} = \sum\sum c_{ij}\bar{Y}_{ij}. \tag{19.93a}$$

$$s^2\{\hat{L}\} = \frac{MSE}{n}\sum\sum c_{ij}^2 \tag{19.93b}$$

$$S^2 = (ab-1)F[1-\alpha; ab-1, (n-1)ab] \tag{19.93c}$$

The test statistic and associated decision rule for all simultaneous Scheffé tests of the form:

$$\begin{aligned} H_0&: L = 0 \\ H_a&: L \neq 0 \end{aligned} \tag{19.94}$$

are as follows when the family significance level is controlled at α:

$$F^* = \frac{\hat{L}^2}{(ab-1)s^2\{\hat{L}\}}; \qquad \text{If } F^* > F[1-\alpha; ab-1, (n-1)ab], \text{ conclude } H_a \tag{19.94a}$$

Bonferroni Procedure. When the number of contrasts is small, the Bonferroni procedure may be preferable. The confidence intervals (19.93) are simply modified by replacing S with B as defined in (19.90). The test statistic and decision rule in (19.94a) are replaced by:

$$t^* = \frac{\hat{L}}{s\{\hat{L}\}}; \qquad \text{If } |t^*| > t[1-\alpha/2g; (n-1)ab], \text{ conclude } H_a \tag{19.95}$$

Example 1—Pairwise Comparisons of Treatment Means

A junior college system studied the effects of teaching method (factor A) and student's quantitative ability (factor B) on learning of college mathematics. Two teaching methods were studied—the standard method of teaching (to be called the standard method) and a method that emphasizes teaching of concepts in the abstract before going into drill routines

TABLE 19.11 Results—Mathematics Learning Example.

(a) Mean Learning Scores ($n = 21$)

Teaching Method i	Quantitative Ability (j)		
	Excellent	Good	Moderate
Abstract	92 ($\bar{Y}_{11\cdot}$)	81 ($\bar{Y}_{12\cdot}$)	73 ($\bar{Y}_{13\cdot}$)
Standard	90 ($\bar{Y}_{21\cdot}$)	86 ($\bar{Y}_{22\cdot}$)	82 ($\bar{Y}_{23\cdot}$)

(b) ANOVA Table

Source of Variation	*SS*	*df*	*MS*
Factor *A* (teaching methods)	504	1	504
Factor *B* (quantitative ability)	3,843	2	1,921.5
AB interactions	651	2	325.5
Error	3,360	120	28
Total	8,358	125	

(to be called the abstract method). The quantitative ability of a student was determined by a standard aptitude test, on the basis of which the student was classified as having excellent, good, or moderate quantitative ability. Thus, factor A (teaching method) has $a = 2$ levels, and factor B (student's quantitative ability) has $b = 3$ levels.

For each quantitative ability group, 42 students were selected and randomly placed into classes according to the designated teaching method, with each class containing equal numbers of students of each quantitative ability level. For simplicity, it is assumed that any effects associated with the classes are negligible.

This study has one experimental factor—teaching method—and one observational factor—quantitative ability. Equal numbers of students with excellent, good, and moderate quantitative ability are randomly selected and then within these categories, students are randomly assigned to a teaching method. Therefore, teaching ability is a blocking factor here with replication within blocks. This experimental study is called a generalized randomized block design and is discussed further in Section 21.6.

The response variable of interest is the amount of learning of college mathematics, as measured by a standard mathematics achievement test. The results of the study are summarized in Table 19.11 (the original data are not shown). The estimated treatment means are shown in Table 19.11a, and the analysis of variance table is presented in Table 19.11b.

Figure 19.13 contains two plots of the estimated treatment means $\bar{Y}_{ij\cdot}$. In Figure 19.13a, the two curves represent the different factor A levels, and in Figure 19.13b, the three curves represent the different factor B levels. The clear lack of parallelism of the curves suggests the presence of interaction effects between teaching method and student's quantitative ability on amount of mathematics learning. A formal test for interactions confirms this. From Table 19.11b, we have $F^* = MSAB/MSE = 325.5/28 = 11.625$. For $\alpha = .01$ we require $F(.99; 2,120) = 4.79$. Since $F^* = 11.625 > 4.79$, we conclude that interaction effects are present. The P-value of this test is 0+.

URE 19.13
ots of
timated
tment
eans—
athematics
arning
xample.

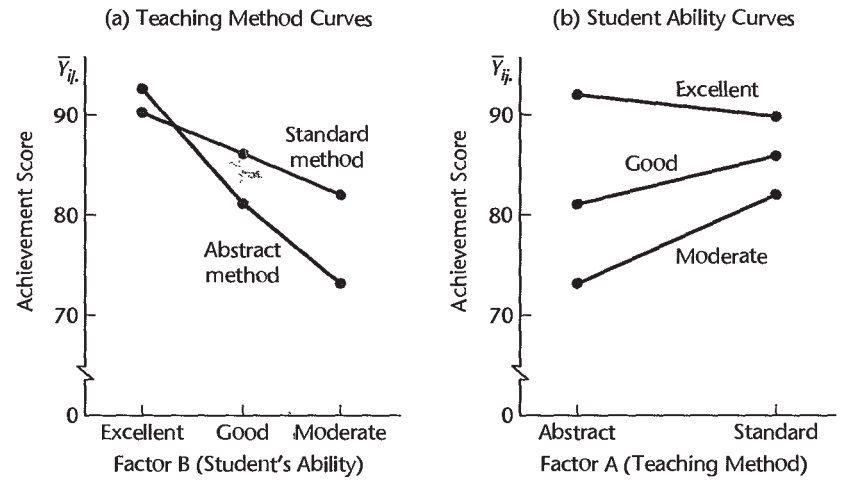


Figure 19.13 suggests that the interactions are important: students with excellent quantitative ability are but little affected by teaching method (perhaps doing slightly better with the abstract method); students with good or moderate abilities learn much better with the standard teaching method. Hence, we shall first investigate whether some simple transformation can make the interactions unimportant. We do this in an approximate fashion by considering the logarithmic and square root transformations of the response. In neither case did the interactions become unimportant, so it appears that the interactions here may be nontransformable.

We now wish to investigate the nature of the interaction effects in Figure 19.13. We shall do this by estimating separately for students with excellent, good, and moderate quantitative abilities how large is the difference in mean learning for the two teaching methods. Thus, we wish to estimate:

$$D_1 = \mu_{11} - \mu_{21}$$
$$D_2 = \mu_{12} - \mu_{22}$$
$$D_3 = \mu_{13} - \mu_{23}$$

We shall employ the Bonferroni multiple comparison procedure with family confidence coefficient .95. (Since only three pairwise comparisons are of interest, the Bonferroni method yields more precise estimates here than the Tukey method.)

For the data in Table 19.11a, the point estimates of the pairwise comparisons are:

$$\hat{D}_1 = 92 - 90 = 2$$
$$\hat{D}_2 = 81 - 86 = -5$$
$$\hat{D}_3 = 73 - 82 = -9$$

We find the estimated variances of these estimates by (19.88b), for $n = 21$:

$$s^2\{\hat{D}_1\} = s^2\{\hat{D}_2\} = s^2\{\hat{D}_3\} = \frac{2(28)}{21} = 2.667$$

so that:

$$s\{\hat{D}_1\} = s\{\hat{D}_2\} = s\{\hat{D}_3\} = 1.633$$

Finally, for family confidence coefficient $1-\alpha = .95$ and $g = 3$, we require $B = t[1-.05/2(3); 120] = t(.99167; 120) = 2.428$. Hence, the confidence limits are by (19.88) and (19.90):

$$2 \pm 2.428(1.633) \qquad -5 \pm 2.428(1.633) \qquad -9 \pm 2.428(1.633)$$

and the 95 percent confidence intervals for the family of comparisons are:

$$-1.96 \leq \mu_{11} - \mu_{21} \leq 5.96$$
$$-8.96 \leq \mu_{12} - \mu_{22} \leq -1.04$$
$$-12.96 \leq \mu_{13} - \mu_{23} \leq -5.04$$

For this family of confidence intervals, the following conclusions may be drawn with family confidence coefficient of 95 percent: (1) For students with excellent quantitative ability, the mean learning scores with the two teaching methods do not differ. (2) For students with either good or moderate quantitative abilities, the mean learning score with the abstract teaching method is lower than that with the standard method. The superiority of the standard teaching method may be particularly strong for students with moderate quantitative ability.

Example 2—Contrasts of Treatment Means

In the mathematics learning example, a school administrator also wished to know whether the amount of learning gain with the standard teaching method over the abstract method is greater for students with moderate quantitative ability than for students with good quantitative ability. This question had been raised before the study began. We shall estimate the single contrast:

$$L = (\mu_{23} - \mu_{13}) - (\mu_{22} - \mu_{12})$$

by means of a one-sided lower confidence interval. For the results in Table 19.11a, the point estimate of L is $\hat{L} = (82 - 73) - (86 - 81) = 4$. The estimated variance by (19.93b) is:

$$s^2\{\hat{L}\} = \frac{28}{21}[(1)^2 + (-1)^2 + (-1)^2 + (1)^2] = 5.333$$

so that the estimated standard deviation is $s\{\hat{L}\} = 2.309$. For a 95 percent confidence coefficient, we require $t(.05; 120) = -1.658$. Hence, the lower confidence limit is $4 - 1.658(2.309)$ and the desired confidence interval is:

$$L \geq .17$$

We conclude, therefore, with 95 percent confidence coefficient that the gain in learning with the standard teaching method over the abstract method is greater for students with moderate quantitative ability than for students with good quantitative ability, the difference in the mean gain being at least .17 point.

19.10 Pooling Sums of Squares in Two-Factor Analysis of Variance

The testing approach presented in this chapter assumes that ANOVA model (19.23) is the full model for all tests of factor effects, regardless of the conclusions reached in any of these tests. The rationale for this approach is that ANOVA model (19.23) is based on the identity (19.22) for the treatment means μ_{ij}. Once the analysis of residuals and other diagnostics demonstrate that this model is appropriate, it is used for all tests.

Some statisticians take the view that ANOVA model (19.23) should be revised when the test for interaction effects leads to the conclusion that no interactions are present. With this approach, the full model considered in testing for factor A and factor B main effects when the test for interaction effects leads to the conclusion that no interactions are present is the revised model:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \varepsilon_{ijk} \qquad \text{Revised full model} \tag{19.96}$$

As we just noted with the regression approach for the Castle Bakery example, the extra sums of squares for factor A and factor B main effects do not depend on the order of the extra sums of squares for factor effects when all treatment sample sizes are equal. Hence, the numerator sums of squares SSA and SSB of the test statistic F^* are not affected by this revision in the full model when the treatment sample sizes are equal. The denominator sum of squares of the F^* test statistic is affected, however, leading to the following error sum of squares for the full model:

$$SSE(F) = SSE + SSAB \tag{19.97}$$

Thus, the error sum of squares for the full model with this approach involves the *pooling* of the interaction and error sums of squares. Likewise, the degrees of freedom are pooled; the degrees of freedom associated with $SSE(F)$ are:

$$df_F = (a-1)(b-1) + (n-1)ab = nab - a - b + 1$$

For the Castle Bakery example, the pooled error sum of squares for testing factor A and factor B main effects would be (Table 19.9):

$$SSE(F) = 62 + 24 = 86$$

and the pooled degrees of freedom would be:

$$df_F = 6 + 2 = 8$$

Hence, the error mean square for testing factor A or factor B main effects with the model revision approach here would be $86/8 = 10.75$.

This pooling procedure affects both the level of significance and the power of the tests for factor A and factor B main effects, in ways not yet fully understood. It has been suggested

therefore by some statisticians that pooling should not be considered unless: (1) the degrees of freedom associated with MSE are small, perhaps 5 or less, and (2) the test statistic $MSAB/MSE$ falls substantially below the action limit of the decision rule, perhaps when $MSAB/MSE < 2$ for $\alpha = .05$. Part (1) of this rule is designed to limit pooling to cases where the gains may be substantial, while part (2) is designed to give reasonable assurance that there are indeed no interactions.

19.11 Planning of Sample Sizes for Two-Factor Studies

We introduced the power approach to sample size planning for single-factor studies in Section 16.10, and the estimation approach to sample size planning for single-factor studies was discussed in Section 17.8. We now consider these two approaches in the context of two-factor studies.

Power Approach

Power of F Test. Table B.11 can be used for determining the power of tests for multi-factor studies in the same fashion as for single-factor studies. The only differences arise in the definition of the noncentrality parameter and the degrees of freedom. For two-factor fixed effects ANOVA model (19.23) with equal treatment sample sizes, the noncentrality parameter ϕ and the degrees of freedom ν_1 and ν_2 for testing for interaction effects, factor A main effects, and factor B main effects are as follows:

Test for interactions:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{n\sum\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)+1}} = \frac{1}{\sigma}\sqrt{\frac{n\sum\sum(\mu_{ij}-\mu_{i\cdot}-\mu_{\cdot j}+\mu_{\cdot\cdot})^2}{(a-1)(b-1)+1}} \tag{19.98a}$$

$$\nu_1 = (a-1)(b-1) \qquad \nu_2 = ab(n-1)$$

Test for A main effects:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{nb\sum\alpha_i^2}{a}} = \frac{1}{\sigma}\sqrt{\frac{nb\sum(\mu_{i\cdot}-\mu_{\cdot\cdot})^2}{a}} \tag{19.98b}$$

$$\nu_1 = a-1 \qquad \nu_2 = ab(n-1)$$

Test for B main effects:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{na\sum\beta_j^2}{b}} = \frac{1}{\sigma}\sqrt{\frac{na\sum(\mu_{\cdot j}-\mu_{\cdot\cdot})^2}{b}} \tag{19.98c}$$

$$\nu_1 = b-1 \qquad \nu_2 = ab(n-1)$$

Use of Table B.12 for Two-factor Studies. When planning sample sizes for two-factor studies with the power approach, one is concerned typically with both the power of detecting factor A main effects and the power of detecting factor B main effects. One can first specify the minimum range of factor A level means for which it is important to detect factor A

main effects, and obtain the needed sample sizes from Table B.12, with $r = a$. The resulting sample size is bn, from which n can be obtained readily. The use of Table B.12 for this purpose is appropriate provided the resulting sample size is not small, specifically provided $a(bn - 1) \geq 20$. If this condition is not met, the ANOVA power tables in Table B.11 should be used. These tables, as noted earlier, require an iterative approach for determining needed sample sizes.

In the same way, the minimum range of factor B level means can then be specified for which it is important to detect factor B main effects, and the needed sample sizes found. If the sample sizes obtained from the factor A and factor B power specifications differ substantially, a judgment will need to be made as to the final sample sizes.

Estimation Approach

The estimation approach to planning sample sizes described in Section 17.8 for single-factor studies is readily adapted for use in two-factor studies. We specify the set of comparisons of interest and determine the expected widths of the confidence intervals for various advance planning values for the standard deviation, σ. Through an iterative, trial-and-error process, we determine a sample size plan that represents an acceptable compromise between the cost of running the study and the precision obtained for comparisons of interest. We illustrate this procedure with a two-factor study example.

Example

In a two-factor study, factor A has $a = 3$ levels and factor B has $b = 2$ levels. No interaction effects are anticipated, and all pairwise comparisons of factor level means are to be made for each of the two factors. A family confidence coefficient of .90 is specified for the $3 + 1 = 4$ pairwise comparisons. Equal treatment sample sizes of n experimental units are to be used. The width of each confidence interval is to be ± 30. A reasonable planning value for the standard deviation of the error terms is $\sigma = 50$.

We know from (19.63) that the variance of a comparison of factor A level means, $\hat{L} = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{i'\cdot\cdot}$, is:

$$\sigma^2\{\hat{L}\} = \frac{\sigma^2}{bn}\sum c_i^2 = \frac{2\sigma^2}{bn} \qquad \text{Factor } A \text{ comparisons}$$

Similarly, the variance of the comparison of the two factor B level means, $\hat{L} = \bar{Y}_{\cdot 1\cdot} - \bar{Y}_{\cdot 2\cdot}$, is:

$$\sigma^2\{\hat{L}\} = \frac{2\sigma^2}{an} \qquad \text{Factor } B \text{ comparison}$$

Since equal precision is specified for all pairwise comparisons and since $a = 3$ and $b = 2$, the variance for the factor A comparisons will be larger for any given treatment sample size n and hence will be the critical consideration.

Suppose that we begin the iterative process with $n = 30$. We then find for the factor A comparisons that $\sigma^2\{\hat{L}\} = 2(50)^2/2(30) = 83.33$ or $\sigma\{\hat{L}\} = 9.13$. For $n_T = 6(30) = 180$, $\alpha = .10$, and $g = 4$ comparisons, the Bonferroni multiple is $B = t(.9875; 174) = 2.26$. Hence, the anticipated width of the confidence intervals is $2.26(9.13) = \pm 20.6$. This

anticipated width is somewhat tighter than the specified width ±30, and a smaller treatment sample size should be tried in the next iteration.

Finding the "Best" Treatment

As we discussed earlier in Section 16.11 in the context of single-factor studies, there are occasions when the chief purpose of the study is to ascertain the treatment with the highest or lowest mean. This is also true for two-factor studies, where the objective is to identify the best of the $r = ab$ factor level combinations. We illustrate the use of this approach with an example.

Two-Factor Study Example. Suppose that in the Castle Bakery example, the chief objective is to identify the combination of shelf height and shelf width that maximizes sales (in cases). There are $3 \times 2 = 6$ treatment combinations. We anticipate that $\sigma = 10$. Further, we want to be able to detect an average difference of $\lambda = 8$ cases between the highest and second highest treatment means with probability $1 - \alpha = .90$ or greater.

The entry in Table B.13 is $\lambda\sqrt{n}/\sigma$. For $r = 6$ and probability $1 - \alpha = .90$, we find from Table B.13 that $\lambda\sqrt{n}/\sigma = 2.7100$. Hence, since $\lambda = 8$, we obtain:

$$\frac{(8)\sqrt{n}}{10} = 2.7100$$

$$\sqrt{n} = 3.3875 \qquad \text{or} \qquad n = 12$$

Thus, when the average number of cases for the best shelf height and shelf width treatment mean exceeds that of the second best by at least 8 cases and $\sigma = 10$, sample sizes of 12 supermarkets for each shelf height and shelf width combination are needed to provide an assurance of at least .90 that the highest estimated mean $\bar{Y}_{ij\cdot}$ corresponds to the highest population mean.

Problems

19.1. Refer to the **SENIC** data set in Appendix C.1. An analyst wishes to investigate the effects of medical school affiliation (factor A) and geographic region (factor B) on infection risk. All factor level combinations will be included in the study.

a. How many treatments are being studied?

b. What is the response variable here?

19.2. A student in a class discussion stated: "A treatment is a treatment, whether the study involves a single factor or multiple factors. The number of factors has little effect on the interpretation of the results." Discuss.

19.3. Verify the interactions in Table 19.3b.

*19.4. In a two-factor study, the treatment means μ_{ij} are as follows:

	Factor B		
Factor A	B_1	B_2	B_3
A_1	34	23	36
A_2	40	29	42

a. Obtain the factor A level means.

b. Obtain the main effects of factor A.

c. Does the fact that $\mu_{12} - \mu_{11} = -11$ while $\mu_{13} - \mu_{12} = 13$ imply that factors A and B interact? Explain.

d. Prepare a treatment means plot and determine whether the two factors interact. What do you find?

19.5. In a two-factor study, the treatment means μ_{ij} are as follows:

	Factor B			
Factor A	B_1	B_2	B_3	B_4
A_1	250	265	268	269
A_2	288	273	270	269

a. Obtain the factor B main effects. What do your results imply about factor B?

b. Prepare a treatment means plot and determine whether the two factors interact. How can you tell that interactions are present? Are the interactions important or unimportant?

c. Make a logarithmic transformation of the μ_{ij} and plot the transformed values to explore whether this transformation is helpful in reducing the interactions. What are your findings?

19.6. Three sets of treatment means μ_{ij} for students' grades in a course follow, where factor A is student's major (A_1: computer science; A_2: mathematics) and factor B is student's class affiliation (B_1: junior; B_2: senior; B_3: graduate).

	Set 1				Set 2				Set 3		
	B_1	B_2	B_3		B_1	B_2	B_3		B_1	B_2	B_3
A_1	80	80	80	A_1	75	80	90	A_1	75	80	85
A_2	90	90	90	A_2	80	86	97	A_2	75	85	100

Prepare a treatment means plot for each set of μ_{ij} to study interaction effects. Interpret each plot and state your findings. If interactions are present, describe their nature and indicate whether they are important or unimportant.

*19.7. Refer to Problem 19.4. Assume that $\sigma = 1.4$ and $n = 10$.

a. Obtain $E\{MSE\}$ and $E\{MSA\}$.

b. Is $E\{MSA\}$ substantially larger than $E\{MSE\}$? What is the implication of this?

19.8. Refer to Problem 19.5. Assume that $\sigma = 4$ and $n = 6$.

a. Obtain $E\{MSE\}$ and $E\{MSAB\}$.

b. Is $E\{MSAB\}$ substantially larger than $E\{MSE\}$? What is the implication of this?

19.9. A psychologist stated: "I feel uncomfortable about deciding in a research study whether the interactions are important or unimportant. I would rather have the statistician make that decision." Comment.

*19.10. Refer to **Cash offers** Problem 16.10. Six male and six female volunteers were used in each age group. The observations (in hundred dollars), classified by age (factor A) and gender of owner (factor B), follow.

Factor A (age)		Factor B (gender of owner) $j = 1$ Male	$j = 2$ Female
$i = 1$	Young	21	21
		23	22
		...	...
		23	25
$i = 2$	Middle	30	26
		29	29
		...	...
		27	29
$i = 3$	Elderly	25	23
		22	19
		...	...
		21	20

a. Obtain the fitted values for ANOVA model (19.23).

b. Obtain the residuals. Do they sum to zero for each treatment?

c. Prepare aligned residual dot plots for the treatments. What departures from ANOVA model (19.23) can be studied from these plots? What are your findings?

d. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

e. The observations for each treatment were obtained in the order shown. Prepare residual sequence plots and interpret them. What are your findings?

*19.11. Refer to **Cash offers** Problems 16.10 and 19.10. Assume that ANOVA model (19.23) is applicable.

a. Prepare an estimated treatment means plot. Does it appear that any factor effects are present? Explain.

b. Set up the analysis of variance table. Does any one source account for most of the total variability in cash offers in the study? Explain.

c. Test whether or not interaction effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test whether or not age and gender main effects are present. In each case, use $\alpha = .05$ and state the alternatives, decision rule, and conclusion. What is the P-value of the test? Is it meaningful here to test for main factor effects? Explain.

e. Obtain an upper bound on the family level of significance for the tests in parts (c) and (d); use the Kimball inequality (19.53).

f. Do the results in parts (c) and (d) confirm your graphic analysis in part (a)?

g. What are the relations between the sums of squares in the two-factor analysis of variance in part (b) and the sums of squares in the single-factor analysis of variance in Problem 16.10d? Do the same relations hold for the degrees of freedom?

19.12. **Eye contact effect.** In a study of the effect of applicant's eye contact (factor A) and personnel officer's gender (factor B) on the personnel officer's assessment of likely job success of applicant, 10 male and 10 female personnel officers were shown a front view photograph of an applicant's face and were asked to give the person in the photograph a success rating on a scale of 0 (total failure) to 20 (outstanding success). Half of the officers in each gender group were chosen at random to receive a version of the photograph in which the applicant made eye contact with the camera lens. The other half received a version in which there was no eye contact. The success ratings follow.

	Factor B (gender of officer)	
Factor A (eye contact)	$j = 1$ Male	$j = 2$ Female
$i = 1$ Present	11	15
	7	12
	...	...
	10	16
$i = 2$ Absent	12	14
	16	17
	...	...
	14	18

a. Obtain the fitted values for ANOVA model (19.23).

b. Obtain the residuals. Do they sum to zero for each treatment?

c. Prepare aligned residual dot plots for the treatments. What departures from ANOVA model (19.23) can be studied from these plots? What are your findings?

d. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

e. The observations for each treatment were obtained in the order shown. Prepare residual sequence plots and interpret them. What are your findings?

19.13. Refer to **Eye contact effect** Problem 19.12. Assume that ANOVA model (19.23) is applicable.

a. Prepare an estimated treatment means plot. Does it appear that any factor effects are present? Explain.

b. Set up the analysis of variance table. Does any one source account for most of the total variability in the success ratings in the study? Explain.

c. Test whether or not interaction effects are present; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test whether or not eye contact and gender main effects are present. In each case, use $\alpha = .01$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test? Is it meaningful here to test for main factor effects? Explain.

e. Obtain an upper bound on the family level of significance for the tests in parts (c) and (d); use the Kimball inequality (19.53).

f. Do the results in parts (c) and (d) confirm your graphic analysis in part (a)?

*19.14. **Hay fever relief.** A research laboratory was developing a new compound for the relief of severe cases of hay fever. In an experiment with 36 volunteers, the amounts of the two active ingredients (factors A and B) in the compound were varied at three levels each. Randomization was used in assigning four volunteers to each of the nine treatments. The data on hours of relief follow.

Factor A (ingredient 1)		Factor B (ingredient 2) $j = 1$ Low	$j = 2$ Medium	$j = 3$ High
$i = 1$	Low	2.4	4.6	4.8
		...	...	...
		2.5	4.7	4.6
$i = 2$	Medium	5.8	8.9	9.1
		...	...	.
		5.3	9.0	9.4
$i = 3$	High	6.1	9.9	13.5
		...	...	...
		6.2	10.1	13.2

a. Obtain the fitted values for ANOVA model (19.23).

b. Obtain the residuals.

c. Plot the residuals against the fitted values. What departures from ANOVA model (19.23) can be studied from this plot? What are your findings?

d. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

*19.15. Refer to **Hay fever relief** Problem 19.14. Assume that ANOVA model (19.23) is applicable.

a. Prepare an estimated treatment means plot. Does your graph suggest that any factor effects are present? Explain.

b. Obtain the analysis of variance table. Does any one source account for most of the total variability in hours of relief in the study? Explain.

c. Test whether or not the two factors interact; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test whether or not main effects for the two ingredients are present. Use $\alpha = .05$ in each case and state the alternatives, decision rule, and conclusion. What is the P-value of each test? Is it meaningful here to test for main factor effects? Explain.

e. Obtain an upper bound on the family level of significance for the tests in parts (c) and (d); use the Kimball inequality (19.53).

f. Do the results in parts (c) and (d) confirm your graphic analysis in part (a)?

19.16. **Disk drive service.** The staff of a service center for electronic equipment includes three technicians who specialize in repairing three widely used makes of disk drives for desktop computers. It was desired to study the effects of technician (factor A) and make of disk drive (factor B) on the service time. The data that follow show the number of minutes required to

complete the repair job in a study where each technician was randomly assigned to five jobs on each make of disk drive.

Factor A (technician)	Factor B (make of drive) $j = 1$ Make 1	$j = 2$ Make 2	$j = 3$ Make 3
$i = 1$ Technician 1	62	57	59
	48	45	53
	...	...	...
	69	44	47
$i = 2$ Technician 2	51	61	55
	57	58	58
	...	...	...
	39	51	49
$i = 3$ Technician 3	59	58	47
	65	63	56
	...	...	...
	70	60	50

a. Obtain the fitted values for ANOVA model (19.23).

b. Obtain the residuals.

c. Plot the residuals against the fitted values. What departures from ANOVA model (19.23) can be studied from this plot? What are your findings?

d. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

e. The observations for each treatment were obtained in the order shown. Prepare residual sequence plots and analyze them. What are your findings?

19.17. Refer to **Disk drive service** Problem 19.16. Assume that ANOVA model (19.23) is applicable.

a. Prepare an estimated treatment means plot. Does your graph suggest that any factor effects are present? Explain.

b. Obtain the analysis of variance table. Does any one source account for most of the total variability? Explain.

c. Test whether or not the two factors interact; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test whether or not main effects for technician and make of drive are present. Use $\alpha = .01$ in each case and state the alternatives, decision rule, and conclusion. What is the P-value of each test? Is it meaningful here to test for main factor effects? Explain.

e. Obtain an upper bound on the family level of significance for the tests in parts (c) and (d); use the Kimball inequality (19.53).

f. Do the results in parts (c) and (d) confirm your graphic analysis in part (a)?

19.18. **Kidney failure hospitalization.** Kidney failure patients are commonly treated on dialysis machines that filter toxic substances from the blood. The appropriate "dose" for effective treatment depends, among other things, on duration of treatment and weight gain between treatments as a result of fluid buildup. To study the effects of these two factors on the number of days hospitalized (attributable to the disease) during a year, a random sample of 10 patients per group who had undergone treatment at a large dialysis facility was obtained. Treatment

duration (factor A) was categorized into two groups: short duration (average dialysis time for the year under four hours) and long duration (average dialysis time for the year equal to or greater than four hours). Average weight gain between treatments (factor B) during the year was categorized into three groups: slight, moderate, and substantial. The data on number of days hospitalized follow.

Factor A (duration)		Factor B (weight gain) $j = 1$ Mild		$j = 2$ Moderate		$j = 3$ Substantial	
$i = 1$	Short	0	2	2	4	15	16
		2	0	4	3	10	7
		...	...		...	...	...
		0	8	15	20	25	27
$i = 2$	Long	0	2	5	1	10	15
		1	7	3	3	8	4
		..			...	.	
		4	3	1	9	7	1

The transformed data $Y' = \log_{10}(Y + 1)$ are to be used for the analysis.

a. Obtain the fitted values and residuals for ANOVA model (19.23) for the transformed data.

b. Prepare aligned residual dot plots for the treatments. What departures from ANOVA model (19.23) can be studied from these plots? What are your findings?

c. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

19.19. Refer to **Kidney failure hospitalization** Problem 19.18. Assume that ANOVA model (19.23) is appropriate for the transformed response variable.

a. Prepare an estimated treatment means plot. Does your graph suggest that any factor effects are present? Explain.

b. Obtain the analysis of variance table. Does any one source account for most of the total variability? Explain.

c. Test whether or not the two factors interact; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test whether or not main effects for duration and weight gain are present. Use $\alpha = .05$ in each case and state the alternatives, decision rule, and conclusion. What is the P-value of each test? Is it meaningful here to test for main factor effects? Explain.

e. Obtain an upper bound on the family level of significance for the tests in parts (c) and (d); use the Kimball inequality (19.53).

f. Do the results in parts (c) and (d) confirm your graphic analysis in part (a)?

*19.20. **Programmer requirements.** A computer software firm was encountering difficulties in forecasting the programmer requirements for large-scale programming projects. As part of a study to remedy the difficulties, 24 programmers, classified into equal groups by type of experience (factor A) and amount of experience (factor B), were asked to predict the number of programmer-days required to complete a large project about to be initiated. After this project

was completed, the prediction errors (actual minus predicted programmer-days) were determined. The data on prediction errors follow.

Factor A (type of experience)		Factor B (years of experience) $j = 1$ Under 5	$j = 2$ 5–under 10	$j = 3$ 10 or more
$i = 1$	Small systems only	240	110	56
		206	118	60
		217	103	68
		225	95	58
$i = 2$	Small and large systems	71	47	37
		53	52	33
		68	31	40
		57	49	45

a. Obtain the fitted values for ANOVA model (19.23).

b. Obtain the residuals.

c. Prepare aligned residual dot plots for the treatments. What departures from ANOVA model (19.23) can be studied from these plots? What are your findings?

d. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

*19.21. Refer to **Programmer requirements** Problem 19.20. Assume that ANOVA model (19.23) is applicable.

a. Prepare an estimated treatment means plot. Does your graph suggest that any factor effects are present? Explain.

b. Obtain the analysis of variance table. Does any one source account for most of the total variability? Explain.

c. Test whether or not the two factors interact; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test whether or not main effects for type of experience and years of experience are present. Use $\alpha = .01$ in each case and state the alternatives, decision rule, and conclusion. What is the P-value of each test? Is it meaningful here to test for main factor effects? Explain.

e. Obtain an upper bound on the family level of significance for the tests in parts (c) and (d); use the Kimball inequality (19.53).

f. Do the results in parts (c) and (d) confirm your graphic analysis in part (a)?

19.22. How does the randomization of treatment assignments in a two-factor study differ when both factors are experimental factors and when only one factor is an experimental factor?

19.23. Refer to **Eye contact effect** Problem 19.12.

a. Explain how you would make the assignments of personnel officers to treatments in this two-factor study. Make all appropriate randomizations.

b. Did you randomize the officers to the factor levels of each factor?

*19.24. Refer to **Hay fever relief** Problem 19.14.

a. Explain how you would make the assignments of volunteers to treatments in this study. Make all appropriate randomizations.

b. Did you randomize the volunteers to the factor levels of each factor?

19.25. Refer to **Disk drive service** Problem 19.16.

a. Is any randomization of treatment assignments called for in this study? Is any randomization utilized? Explain.

b. Would you consider this study to be experimental in nature? Discuss.

19.26. Why is it suggested in the flowchart in Figure 19.11 that a test for interactions should be conducted before tests for main factor effects? Explain.

*19.27. A two-factor study was conducted with $a = 5$, $b = 5$, and $n = 4$. No interactions between factors A and B were noted, and the analyst now wishes to estimate all pairwise comparisons among the factor A level means and all pairwise comparisons among the factor B level means. The family confidence coefficient for the joint set of interval estimates is to be 90 percent.

a. Is it more efficient to use the Bonferroni procedure for the entire family or to use the Tukey procedure for each family of factor level mean comparisons and then to join the two families by means of the Bonferroni procedure?

b. Would your answer differ if each factor had three levels, everything else remaining the same?

19.28. A two-factor study was conducted with $a = 6$, $b = 6$, and $n = 10$. No interactions between factors A and B were found, and it is now desired to estimate five contrasts of factor A level means and four contrasts of factor B level means. The family confidence coefficient for the joint set of estimates is to be 95 percent. Which of the three procedures at the bottom of page 852 and the top of page 853 will be most efficient here?

19.29. Refer to the Castle Bakery example at the top of page 855, where two pairwise comparison estimates were made by means of the Tukey procedure. Why would it not be appropriate to use the Bonferroni procedure here? Discuss.

*19.30. Refer to **Cash offers** Problems 19.10 and 19.11.

a. Estimate μ_{11} with a 95 percent confidence interval. Interpret your interval estimate.

b. Prepare a bar graph of the estimated factor B level means. What does this plot suggest about the equality of the factor B level means?

c. Estimate $D = \mu_{.1} - \mu_{.2}$ by means of a 95 percent confidence interval. Is your confidence interval consistent with the test result in Problem 19.11d? Is your confidence interval consistent with your finding in part (b)? Explain.

d. Prepare a bar graph of the estimated factor A level means. What does this plot suggest about the factor A main effects?

e. Obtain all pairwise comparisons among the factor A level means; use the Tukey procedure with a 90 percent family confidence coefficient. Present your findings graphically and summarize your results. Are your conclusions consistent with those in part (d)?

f. Is the Tukey procedure used in part (e) the most efficient one that could be used here? Explain.

g. Estimate the contrast:

$$L = \frac{\mu_{1\cdot} + \mu_{3\cdot}}{2} - \mu_{2\cdot}$$

with a 95 percent confidence interval. Interpret your interval estimate.

h. Suppose that in the population of female owners, 30 percent are young, 60 percent are middle-aged, and 10 percent are elderly. Obtain a 95 percent confidence interval for the mean cash offer in the population of female owners.

19.31. Refer to **Eye contact effect** Problems 19.12 and 19.13.

a. Estimate μ_{21} with a 99 percent confidence interval. Interpret your interval estimate.

b. Estimate $\mu_{1\cdot}$ with a 99 percent confidence interval. Interpret your interval estimate.

c. Prepare a bar graph of the estimated factor B level means. What does this plot suggest about the factor B main effects?

d. Obtain confidence intervals for $\mu_{\cdot 1}$ and $\mu_{\cdot 2}$, each with a 99 percent confidence coefficient. Interpret your interval estimates. What is the family confidence coefficient for the set of two estimates?

e. Prepare a bar graph of the estimated factor A level means. What does this plot suggest about the factor A main effects?

f. Obtain confidence intervals for $D_1 = \mu_{2\cdot} - \mu_{1\cdot}$ and $D_2 = \mu_{\cdot 2} - \mu_{\cdot 1}$; use the Bonferroni procedure and a 95 percent family confidence coefficient. Summarize your findings. Are your findings consistent with those in parts (c) and (e)?

g. Is the Bonferroni procedure used in part (f) the most efficient one that could be used here? Explain.

*19.32. Refer to **Hay fever relief** Problems 19.14 and 19.15.

a. Estimate μ_{23} with a 95 percent confidence interval. Interpret your interval estimate.

b. Estimate $D = \mu_{12} - \mu_{11}$ with a 95 percent confidence interval. Interpret your interval estimate.

c. The analyst decided to study the nature of the interacting factor effects by means of the following contrasts:

$$L_1 = \frac{\mu_{12} + \mu_{13}}{2} - \mu_{11} \qquad L_4 = L_2 - L_1$$
$$L_2 = \frac{\mu_{22} + \mu_{23}}{2} - \mu_{21} \qquad L_5 = L_3 - L_1$$
$$L_3 = \frac{\mu_{32} + \mu_{33}}{2} - \mu_{31} \qquad L_6 = L_3 - L_2$$

Obtain confidence intervals for these contrasts; use the Scheffé multiple comparison procedure with a 90 percent family confidence coefficient. Interpret your findings.

d. The analyst also wished to identify the treatment(s) yielding the longest mean relief. Using the Tukey testing procedure with family significance level $\alpha = .10$, identify the treatment(s) providing the longest mean relief.

e. To examine whether a transformation of the data would make the interactions unimportant, plot separately the transformed estimated treatment means for the reciprocal and square root transformations. Would either of these transformations have made the interaction effects unimportant? Explain.

19.33. Refer to **Disk drive service** Problems 19.16 and 19.17.

a. Estimate μ_{11} with a 99 percent confidence interval. Interpret your interval estimate.

b. Estimate $D = \mu_{22} - \mu_{21}$ with a 99 percent confidence interval. Interpret your interval estimate.

c. The nature of the interaction effects is to be studied by making, for each technician, all three pairwise comparisons among the disk drive makes in order to identify, if possible, the make of disk drive for which the technician's mean service time is lowest. The family confidence coefficient for each set of three pairwise comparisons is to be 95 percent. Use the Bonferroni procedure to make all required pairwise comparisons. Summarize your findings.

d. The service center currently services 30 disk drives of each of the three makes per week, with each technician servicing 10 machines of each make. Estimate the expected total amount of service time required per week to service the 90 disk drives; use a 99 percent confidence interval.

e. How much time could be saved per week, on the average, if technician 1 services only make 2, technician 2 services only make 1, and technician 3 services only make 3? Use a 99 percent confidence interval.

f. To examine whether a transformation of the data would make the interactions unimportant, plot separately the transformed estimated treatment means for the reciprocal and logarithmic transformations. Would either of these transformations have made the interaction effects unimportant? Explain.

19.34. Refer to **Kidney failure hospitalization** Problems 19.18 and 19.19. Continue to work with the transformed observations $Y' = \log_{10}(Y + 1)$.

a. Estimate μ_{22} with a 95 percent confidence interval. Interpret your interval estimate.

b. Estimate $D = \mu_{23} - \mu_{21}$ with a 95 percent confidence interval. Interpret your interval estimate.

c. Prepare separate bar graphs of the estimated factor A and factor B level means. What do these plots suggest about the factor main effects?

d. The researcher wishes to study the main effects of each of the two factors by making all pairwise comparisons of factor level means with a 90 percent family confidence coefficient for the entire set of comparisons. Which multiple comparison procedure is most efficient here?

e. Using the Bonferroni procedure, make all pairwise comparisons called for in part (d). State your findings and prepare a graphic summary. Are your findings consistent with those in part (c)?

f. It is known from past experience that 30 percent of patients have mild weight gains, 40 percent have moderate weight gains, and 30 percent have severe weight gains, and that these proportions are the same for the two duration groups. Estimate the mean number of days hospitalized (in transformed units) in the entire population with a 95 percent confidence interval. Convert your confidence limits to the original units. Does it appear that the mean number of days is less than 7?

*19.35. Refer to **Programmer requirements** Problems 19.20 and 19.21.

a. Estimate μ_{23} with a 99 percent confidence interval. Interpret your interval estimate.

b. Estimate $D = \mu_{12} - \mu_{13}$ with a 99 percent confidence interval. Interpret your interval estimate.

c. The nature of the interaction effects is to be studied by comparing the effect of type of experience for each years-of-experience group. Specifically, the following comparisons are to be estimated:

$$D_1 = \mu_{11} - \mu_{21} \qquad L_1 = D_1 - D_2$$
$$D_2 = \mu_{12} - \mu_{22} \qquad L_2 = D_1 - D_3$$
$$D_3 = \mu_{13} - \mu_{23} \qquad L_3 = D_2 - D_3$$

The family confidence coefficient is to be 95 percent. Which multiple comparison procedure is most efficient here?

d. Use the most efficient procedure to estimate the comparisons specified in part (c). State your findings.

e. Use the Tukey testing procedure with family significance level $\alpha = .05$ to identify the type of experience-years of experience group(s) with the smallest mean prediction errors.

f. For each group identified in part (e), obtain a confidence interval for the mean prediction error. Use the Bonferroni procedure with a 95 percent family confidence coefficient. Does any group have a mean prediction error that could be zero? Explain.

g. To examine whether a transformation of the data would make the interactions unimportant, plot separately the transformed estimated treatment means for the reciprocal and logarithmic transformations. Would either of these transformations have made the interaction effects unimportant? Explain.

19.36. Refer to **Brand preference** Problem 6.5. Suppose the market researcher first wished to employ analysis of variance model (19.23) to determine whether or not moisture content (factor A) and sweetness (factor B) affect the degree of brand liking.

a. State the analysis of variance model for this case.

b. Obtain the analysis of variance table.

c. Test whether or not the two factors interact; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

d. Study possible curvilinearity of the moisture content effect by estimating the following contrast:

$$L = (\mu_{4\cdot} - \mu_{3\cdot}) - (\mu_{2\cdot} - \mu_{1\cdot})$$

Use a 95 percent confidence interval. What do you conclude?

e. Test whether or not sweetness affects brand liking; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

19.37. A market research manager is planning to study the effects of duration of advertising (factor A) and price level (factor B) on sales. Each factor has three levels. No important interactions are expected, and the primary analysis is to consist of pairwise comparisons of factor level means for each factor. Equal sample sizes are to be used for each treatment. The precision of each comparison is to be ± 3 thousand dollars. The family confidence coefficient for the joint set of comparisons is to be 90 percent, the Tukey procedure is to be used in making the comparisons for each factor, and the Bonferroni procedure is then to be used to join the two sets of comparisons. Assume that $\sigma = 7$ thousand dollars is a reasonable planning value for the error standard deviation. What sample sizes do you recommend?

*19.38. Refer to **Cash offers** Problem 19.10. Suppose that the sample sizes have not yet been determined but it has been decided to use the same number of "owners" in each age-gender group. What are the required sample sizes if: (1) differences in the age factor level means are to be detected with probability .90 or more when the range of the factor level means is 3 (hundred dollars), and (2) the α risk is to be controlled at .05? Assume that a reasonable planning value for the error standard deviation is $\sigma = 1.5$ (hundred dollars).

19.39. Refer to **Eye contact effect** Problem 19.12. Suppose that the sample sizes have not yet been determined but it has been decided to use equal sample sizes for each treatment. Primary interest is in the two comparisons $L_1 = \mu_{1\cdot} - \mu_{2\cdot}$ and $L_2 = \mu_{\cdot 1} - \mu_{\cdot 2}$. What are the required sample sizes if each of these comparisons is to be estimated with precision not to exceed ± 1.2 with a 95 percent family confidence coefficient, using the most efficient multiple comparison procedure? Assume that a reasonable planning value for the error standard deviation is $\sigma = 2.4$.

*19.40. Refer to **Hay fever relief** Problem 19.14. Suppose that the sample sizes have not yet been determined but it has been decided to use equal sample sizes for each treatment. The chief

objective is to identify the dosage combination that yields the longest mean relief. The probability should be at least .99 that the correct dosage combination is identified when the mean relief duration for the second best combination differs by .5 hour or more. What are the required sample sizes? Assume that a reasonable planning value for the error standard deviation is $\sigma = .29$ hour.

19.41. Refer to **Kidney failure hospitalization** Problem 19.18. Suppose that the sample sizes have not yet been determined but it has been decided to use equal sample sizes for each treatment. The chief objective is to estimate the pairwise comparisons:

$$L_1 = \mu_{1\cdot} - \mu_{2\cdot} \qquad L_3 = \mu_{\cdot 1} - \mu_{\cdot 3}$$
$$L_2 = \mu_{\cdot 1} - \mu_{\cdot 2} \qquad L_4 = \mu_{\cdot 2} - \mu_{\cdot 3}$$

What are the required sample sizes if the precision of each of the estimates should not exceed $\pm .20$ (in transformed units), using the Bonferroni procedure with a family confidence coefficient of 90 percent for the joint set of comparisons? A reasonable planning value for the error standard deviation is $\sigma = .32$ (in transformed units).

*19.42. Refer to **Programmer requirements** Problem 19.20. Suppose that the sample sizes have not yet been determined but it has been decided to use equal sample sizes for each treatment. Primary interest is in identifying the type of experience-years of experience combination for which the mean prediction error is smallest. The probability should be at least .95 that the correct combination is identified when the mean prediction error for the second best combination differs by 8.0 programmer-days or more. Assume that a reasonable planning value for the error standard deviation is $\sigma = 9.1$ days. What are the required sample sizes?

Exercises

19.43. Derive (19.7a) from (19.7).

19.44. Prove the result in (19.9b).

19.45. (Calculus needed.) State the likelihood function for ANOVA model (19.15) when $a = 2$, $b = 2$, and $n = 2$. Find the maximum likelihood estimators.

19.46. (Calculus needed.) Derive (19.29).

19.47. Derive (19.39) from (19.38).

19.48. Show that the point estimator (19.67) is unbiased. Find the variance of this estimator.

19.49. Find the variance of the estimator (19.93a).

19.50. Consider a two-factor study with $a = 2$ and $b = 2$. Show that the interactions $(\alpha\beta)_{12}$ and $(\alpha\beta)_{21}$ are equal.

Projects

19.51. Refer to the **SENIC** data set in Appendix C.1. The following hospitals are to be considered in a study of the effects of region (factor A: variable 9) and average age of patients (factor B: variable 3) on the mean length of hospital stay of patients (variable 2):

1–44	46	48	51	53	57	58	60	63	66	74
76	79	80	83	84	88	94	101	103	111	

For purposes of this ANOVA study, average age is to be classified into two categories: less than or equal to 53.9 years, 54.0 years or more.

a. Assemble the required data and obtain the fitted values for ANOVA model (19.23).

b. Obtain the residuals.

c. Plot the residuals against the fitted values. What departures from ANOVA model (19.23) can be studied from this plot? What are your findings?

d. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

52. Refer to the **SENIC** data set in Appendix C.1 and Project 19.51. Assume that ANOVA model (19.23) is applicable.

a. Prepare an estimated treatment means plot. Does it appear that any factor effects are present? Explain.

b. Obtain the analysis of variance table. Does any one source account for most of the total variability in the study? Explain.

c. Test whether or not interaction effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test whether or not region and age main effects are present. In each case, use $\alpha = .05$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test? Is it meaningful here to test for main factor effects? Explain.

e. Obtain an upper bound on the family level of significance for the tests in parts (c) and (d); use the Kimball inequality (19.53).

f. Do the results in parts (c) and (d) confirm your graphic analysis in part (a)?

53. Refer to the **CDI** data set in Appendix C.2. The following metropolitan areas are to be considered in a study of the effects of region (factor A: variable 17) and percent below poverty level (factor B: variable 13) on the crime rate (variable 10 ÷ variable 5):

1–5	7	10–17	19–29	32–34	36–42	44	46	49
51–52	54	57	75	84	87	94	136	151
164	178	182	202	218	410	421	434	

For purposes of this ANOVA study, percent of population below poverty level is to be classified into two categories: less than 8 percent, 8 percent or more.

a. Assemble the required data and obtain the fitted values for ANOVA model (19.23).

b. Obtain the residuals.

c. Prepare aligned residual dot plots for the treatments. What departures from ANOVA model (19.23) can be studied from these plots? What are your findings?

d. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

54. Refer to the **CDI** data set in Appendix C.2 and Project 19.53. Assume that ANOVA model (19.23) is applicable.

a. Prepare an estimated treatment means plot. Does it appear that any factor effects are present? Explain.

b. Set up the analysis of variance table. Does any one source account for most of the total variability in the study? Explain.

c. Test whether or not interaction effects are present; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test whether or not region and percent of population below poverty level main effects are present. In each case, use $\alpha = .01$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test? Is it meaningful here to test for main factor effects? Explain.

e. Obtain an upper bound on the family level of significance for the tests in parts (c) and (d); use the Kimball inequality (19.53).

f. Do the results in parts (c) and (d) confirm your graphic analysis in part (a)?

19.55. Refer to the **Market share** data set in Appendix C.3. A balanced ANOVA study of the effects of discount price (factor A: variable 5) and package promotion (factor B: variable 6) on the average monthly market share (variable 2) is to be conducted. Order the observations in the four factor-level combination cells from smallest to largest observation number and retain the first 7 observations in each cell for a total of 28 observations. (This process omits cases with identification numbers (variable 1) equal to 24, 25, 27, 28, 30, 33, 34, and 36.)

a. Assemble the required data and obtain the fitted values for ANOVA model (19.23).

b. Obtain the residuals.

c. Plot the residuals against the fitted values. What departures from ANOVA model (19.23) can be studied from this plot? What are your findings?

d. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

19.56. Refer to the **Market share** data set in Appendix C.3 and Project 19.55. Assume that ANOVA model (19.23) is applicable.

a. Prepare an estimated treatment means plot. Does it appear that any factor effects are present? Explain.

b. Obtain the analysis of variance table. Does any one source account for most of the total variability in the study? Explain.

c. Test whether or not interaction effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test whether or not discount price and package promotion main effects are present. In each case, use $\alpha = .05$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test? Is it meaningful here to test for main factor effects? Explain.

e. Obtain an upper bound on the family level of significance for the tests in parts (c) and (d); use the Kimball inequality (19.53).

f. Do the results in parts (c) and (d) confirm your graphic analysis in part (a)?

19.57. Refer to the **SENIC** data set in Appendix C.1 and Projects 19.51 and 19.52.

a. Prepare a bar graph of the estimated factor level means $\bar{Y}_{i..}$. What does this plot suggest regarding the region main effects?

b. Analyze the effects of region on mean length of hospital stay by making all pairwise comparisons between regions; use the Tukey procedure and a 90 percent family confidence coefficient. State your findings and present a graphic summary. Are your findings consistent with those in part (a)?

19.58. Refer to the **CDI** data set in Appendix C.2 and Projects 19.53 and 19.54.

a. Prepare a bar graph of the estimated factor level means $\bar{Y}_{i..}$. What does this plot suggest regarding the region main effects?

b. Analyze the effects of region on crime rate by making all pairwise comparisons between regions; use the Tukey procedure and a 95 percent family confidence coefficient. State your findings and present a graphic summary. Are your findings consistent with those in part (a)?

ase
dies

19.59. Refer to the **Real estate sales** data set in Appendix C.7. Carry out a balanced two-way analysis of variance of this data set where the response of interest is sales price (variable 2) and the two crossed factors are quality (variable 10) and style (variable 11). Style is recoded as either 1 or not 1. Order the observations in the six factor-level-combination cells from smallest to largest observation number and retain the first 25 observations in each cell for a total of 150 observations. The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

19.60. Refer to the **Ischemic heart disease** data set in Appendix C.9. Carry out a balanced two-way analysis of variance of this data set where the response of interest is total cost (variable 2) and the two crossed factors are number of interventions (variable 5) and number of comorbidities (variable 9). Recode the number of interventions into six categories: 0, 1, 2, 3–4, 5–7, and greater than or equal to 8. Recode the number of comorbidities into two categories: 0–1, and greater than or equal to 2. Order the observations in the twelve factor-level-combination cells from smallest to largest observation number and retain the first 43 observations in each cell for a total of 516 observations. The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

Chapter 20

Two-Factor Studies—One Case per Treatment

In many studies, constraints on cost, time, and materials severely limit the number of observations that can be obtained. For example, a process engineer in a manufacturing company may have only a limited time to experiment with the production line. If the line is available for one day and only eight batches of product can be produced in a day, the experiment may have to be limited to eight observations. If the study involves one factor at four levels and a second factor at two levels so that there are eight factor level combinations, only one replication of the experiment is then possible for each treatment.

Another reason why some studies contain only one case per treatment is that the response of interest is a single aggregate measure of performance. For example, in a marketing research study of alternative package designs, evaluation of each alternative may require a separate market test. The response of interest is the observed market share, and this results in a single response for each treatment combination.

A modification of the ANOVA model is required for the analysis of two-factor studies containing only one replication per treatment because no degrees of freedom are available for estimation of the experimental error with the standard two-factor ANOVA model (19.23). In this chapter, we describe a modification of the ANOVA model that permits the two-factor analysis of variance to be conducted with only one case per treatment. This modification requires the assumption that the two factors do not interact. We then discuss inference procedures with this additive model. We conclude the chapter by considering a test for examining the reasonableness of the assumption of additivity of the two factors—the Tukey test. This test is important not only when there is just a single case for each treatment in a two-factor study, but it is also useful for a variety of experimental designs to be discussed in later chapters.

20.1 No-Interaction Model

When there is only one case for each treatment, we no longer can work with two-factor ANOVA model (19.23) because no estimate of the error variance σ^2 will be available. Recall from (19.37c) that *SSE* is a sum of squares made up of components measuring the variability within each treatment, $\sum_k (Y_{ijk} - \bar{Y}_{ij.})^2$. With only one case per treatment, there is no variability within a treatment, and *SSE* will then always be zero.

A way out of this difficulty is to change the model. Formula (19.42d) indicates that if the two factors do not interact so that $(\alpha\beta)_{ij} \equiv 0$, the interaction mean square $MSAB$ has expectation σ^2. Thus, if it is possible to assume that the two factors do not interact, we may use $MSAB$ as the estimator of the error variance σ^2 and proceed with the analysis of factor effects as usual. If it is unreasonable to assume that the two factors do not interact, transformations may be tried to remove the interaction effects. We shall say more about this in the next section.

Model

The two-factor ANOVA model with fixed factor levels in (19.23), when all interactions are zero so that $(\alpha\beta)_{ij} \equiv 0$, becomes for $n = 1$, the case considered here:

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + \varepsilon_{ij} \tag{20.1}$$

Note that the third subscript has been dropped from the Y and ε terms because there is now only one case per treatment.

Analysis of Variance

The factor effects sums of squares SSA and SSB are calculated as before from (19.39a) and (19.39b), respectively, with $n = 1$. The interaction sum of squares in (19.39c) with $n = 1$ now is expressed as follows:

$$SSAB = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2 \qquad n = 1 \tag{20.2}$$

Note that $SSAB$ in (20.2) is identical to $SSAB$ in (19.39c) with $n = 1$; the third subscript has been dropped because there is only one case per treatment, and the mean $\bar{Y}_{ij\cdot}$ is replaced by the observation Y_{ij} for the same reason. The number of degrees of freedom associated with $SSAB$ in (20.2) is the same as before, namely, $(a-1)(b-1)$. The analysis of variance table for the case $n = 1$ for no-interaction model (20.1) is shown in Table 20.1.

Inference Procedures

No new problems arise in the tests for factor A and factor B main effects, nor in estimating these effects. Since the expected value of $MSAB$ is σ^2 for no-interaction model (20.1), as

TABLE 20.1 ANOVA Table for No-Interaction Two-Factor Model (20.1) with Fixed Factor Levels, $n = 1$.

Source of Variation	SS	df	MS	E{MS}
Factor A	$SSA = b\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$a-1$	$MSA = \dfrac{SSA}{a-1}$	$\sigma^2 + b\dfrac{\sum(\mu_{i\cdot} - \mu_{\cdot\cdot})^2}{a-1}$
Factor B	$SSB = a\sum(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2$	$b-1$	$MSB = \dfrac{SSB}{b-1}$	$\sigma^2 + a\dfrac{\sum(\mu_{\cdot j} - \mu_{\cdot\cdot})^2}{b-1}$
Error	$SSAB = \sum\sum(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2$	$(a-1)(b-1)$	$MSAB = \dfrac{SSAB}{(a-1)(b-1)}$	σ^2
Total	$SSTO = \sum\sum(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	$ab-1$		

shown in the last column of Table 20.1, the F^* test statistics for testing factor A and factor B main effects will now utilize $MSAB$ in the denominator, instead of MSE as before:

$$\text{Factor } A \text{ main effects:} \qquad F^* = \frac{MSA}{MSAB} \tag{20.3a}$$

$$\text{Factor } B \text{ main effects:} \qquad F^* = \frac{MSB}{MSAB} \tag{20.3b}$$

Similarly, for estimating comparisons of factor A and factor B level means, we simply replace MSE in all of the earlier results with $MSAB$ as the estimator of the error variance σ^2 and modify the degrees of freedom accordingly.

A special problem exists in estimating treatment means. We shall explain how to handle this problem after presenting an example.

Example

An analyst in an insurance commissioner's office studied the premiums for automobile insurance charged by an insurance company in six cities. The six cities were selected to represent different regions of the state and different sizes of cities. Table 20.2a shows the amounts of three-month premiums charged by the automobile insurance firm for a specific type and amount of coverage in a given risk category for each of the six cities, classified by size of city (factor A) and geographic region (factor B). Note there is only one observation per cell, namely, the amount of the premium charged in the city for each factor level combination. The analyst wished to evaluate the effects of city size and geographic region on the amount of the premium.

Figure 20.1 contains a plot of the observations Y_{ij}. Note since $n = 1$ here that the observations Y_{ij} constitute estimates of the treatment means μ_{ij}. It appears from Figure 20.1 that there could be a slight interaction between region and size of city in their effects on the

TABLE 20.2 Two-Factor Study with $n = 1$—Insurance Premium Example.

(a) Premiums for Automobile Insurance Policy (in dollars)

Size of City (factor A)	Region (factor B): East ($j = 1$)	West ($j = 2$)	Average
Small ($i = 1$)	140	100	120
Medium ($i = 2$)	210	180	195
Large ($i = 3$)	220	200	210
Average	190	160	175

(b) ANOVA Table

Source of Variation	SS	df	MS
Size of city (A)	9,300	2	4,650
Region (B)	1,350	1	1,350
Error	100	2	50
Total	10,750	5	

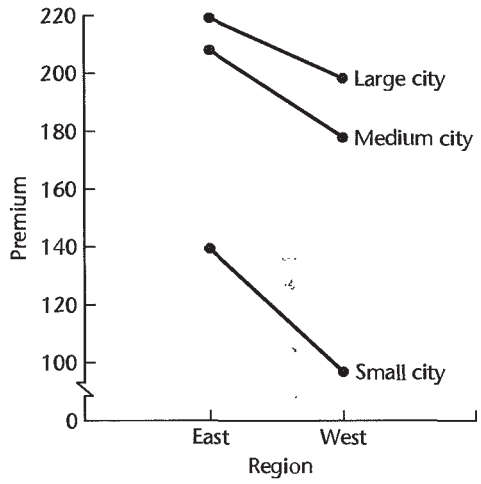


FIGURE 20.1 Plot of Observations Y_{ij}—Insurance Premium Example.

premium. However, since there is only one observation per treatment, the moderate lack of parallelism in the response lines could simply be the result of random effects within each treatment cell. The analyst conducted the Tukey test for interactions (to be discussed in Section 20.2), which indicated that no interaction effects are present. Hence, the analyst adopted the no-interaction model (20.1).

The analyst obtained the required sums of squares as follows, using (19.37a) and (19.39) for $n = 1$:

$$SSA = 2[(120 - 175)^2 + (195 - 175)^2 + (210 - 175)^2] = 9{,}300$$
$$SSB = 3[(190 - 175)^2 + (160 - 175)^2] = 1{,}350$$
$$SSAB = (140 - 120 - 190 + 175)^2 + \cdots + (200 - 210 - 160 + 175)^2 = 100$$
$$SSTO = (140 - 175)^2 + \cdots + (200 - 175)^2 = 10{,}750$$

The ANOVA table is given in Table 20.2b. For the test of city size (factor A) effects, the alternative conclusions are:

$$H_0\colon \alpha_1 = \alpha_2 = \alpha_3 = 0$$
$$H_a\colon \text{not all } \alpha_i \text{ equal zero}$$

The F^* test statistic is given by (20.3a):

$$F^* = \frac{MSA}{MSAB} = \frac{4{,}650}{50} = 93$$

and the decision rule for $\alpha = .05$ is [remember that the denominator of F^* here involves $(a-1)(b-1)$ degrees of freedom]:

$$\text{If } F^* \le F[1-\alpha; a-1, (a-1)(b-1)] = F(.95; 2, 2) = 19.0, \text{ conclude } H_0$$
$$\text{If } F^* > F[1-\alpha; a-1, (a-1)(b-1)] = F(.95; 2, 2) = 19.0, \text{ conclude } H_a$$

Since $F^* = 93 > 19.0$, we conclude H_a, that city size effects are present. The P-value of the test is .011.

The test for geographic region (factor B) effects proceeds similarly, the alternative conclusions being:

$$H_0: \beta_1 = \beta_2 = 0$$
$$H_a: \text{not all } \beta_j \text{ equal zero}$$

For $\alpha = .05$ the decision rule is:

$$\text{If } F^* \leq F(.95; 1, 2) = 18.5, \text{ conclude } H_0$$
$$\text{If } F^* > F(.95; 1, 2) = 18.5, \text{ conclude } H_a$$

Test statistic (20.3b) here is:

$$F^* = \frac{MSB}{MSAB} = \frac{1{,}350}{50} = 27$$

Since $F^* = 27 > 18.5$, we conclude H_a, that geographic region effects are present. The P-value of this test is .035.

The analysis of the magnitudes of the geographic region and city size main effects involves no new problems. The analyst employed three pairwise comparisons of the factor level means $\mu_{i\cdot}$ for city size effects and a pairwise comparison of the geographic region factor level means $\mu_{\cdot j}$. The methods described in Section 19.8 are entirely applicable here; the error variance σ^2 is now estimated by $MSAB$, and the degrees of freedom associated with the estimate of the error variance now are $(a - 1)(b - 1)$. Since no new issues are involved in the analysis, we do not present further details.

Estimation of Treatment Mean

Occasionally when no-interaction model (20.1) is employed with $n = 1$, there is interest in estimating a treatment mean μ_{ij}. We could estimate treatment mean μ_{ij} in the usual fashion with the sample mean $\bar{Y}_{ij\cdot}$, here simply the single observation Y_{ij}. However, we can obtain an improved estimate by making use of the model assumption of no interactions. We know from (19.7a) that when the factor effects are additive, the treatment mean μ_{ij} can be expressed as follows:

$$\mu_{ij} = \mu_{i\cdot} + \mu_{\cdot j} - \mu_{\cdot\cdot} \tag{20.4}$$

Hence, we can estimate μ_{ij} for additive model (20.1) by substituting the estimated values $\hat{\mu}_{i\cdot} = \bar{Y}_{i\cdot}$, $\hat{\mu}_{\cdot j} = \bar{Y}_{\cdot j}$, and $\hat{\mu}_{\cdot\cdot} = \bar{Y}_{\cdot\cdot}$ into (20.4):

$$\hat{\mu}_{ij} = \bar{Y}_{i\cdot} + \bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot} \tag{20.5}$$

The estimator of the treatment mean μ_{ij} in (20.5) is an improved estimator because it has minimum variance in the class of unbiased linear estimators according to an extension of the Gauss-Markov theorem (1.11).

Example

For the insurance premium example in Table 20.2a, we shall use (20.5) to obtain improved estimates of the treatment means μ_{ij}. We obtain, for instance:

$$\hat{\mu}_{11} = 120 + 190 - 175 = 135$$
$$\hat{\mu}_{12} = 120 + 160 - 175 = 105$$

The other treatment mean estimates are:

$$\hat{\mu}_{21} = 210 \qquad \hat{\mu}_{22} = 180 \qquad \hat{\mu}_{31} = 225 \qquad \hat{\mu}_{32} = 195$$

Note that these improved estimates differ only slightly from the simpler estimates Y_{ij} in Table 20.2a.

Precision of Estimated Treatment Mean. To set up a confidence interval for a treatment mean μ_{ij}, we require the estimated variance of $\hat{\mu}_{ij}$ in (20.5). One simple method of estimating this variance is by means of the regression model equivalent to ANOVA model (20.1). For the insurance premium example, this equivalent regression model is:

$$Y_{ij} = \mu_{..} + \alpha_1 X_{ij1} + \alpha_2 X_{ij2} + \beta_1 X_{ij3} + \varepsilon_{ij}$$

where:

$$X_1 = \begin{cases} 1 & \text{if small city} \\ -1 & \text{if large city} \\ 0 & \text{if medium city} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if medium city} \\ -1 & \text{if large city} \\ 0 & \text{if small city} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{if region East} \\ -1 & \text{if region West} \end{cases}$$

Note that the fitted value for observation Y_{ij} will be:

$$\hat{Y}_{ij} = \bar{Y}_{..} + \hat{\alpha}_i + \hat{\beta}_j$$

which is identical to $\hat{\mu}_{ij}$ in (20.5):

$$\hat{Y}_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..}) = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} = \hat{\mu}_{ij}$$

Hence, the estimated variance of $\hat{Y}_{ij}$ is also the estimated variance of $\hat{\mu}_{ij}$. The estimated variance $s^2\{\hat{Y}_{ij}\}$ is furnished by most computer regression packages or can be calculated by means of (6.58).

Comments

1. The analysis of two-factor studies with $n = 1$ just outlined depends on the assumption that the two factors do not interact. If this analysis is used when in fact interactions are present, the result is that the actual level of significance for testing factor A and factor B main effects is below the specified level and the actual power of the tests is lower than the expected power. Correspondingly, confidence intervals for contrasts of factor level means will tend to be too wide. This means that when interactions are present, the analysis is more likely to fail to disclose real effects than anticipated. However, when the analysis based on the no-interaction model does indicate the presence of factor A or factor B main effects, they may be taken as real effects even though interactions are actually present.

2. Sometimes, the case $n = 1$ is encountered when the observations Y_{ij} are proportions. For instance, the data may consist of the proportion of employees in a plant absent during the past week, with the plants classified by size and geographic area. As noted earlier, the arcsine transformation can be used for such data to stabilize the error variance. The transformed data then can be analyzed using no-interaction model (20.1), provided that each proportion is based on roughly the same number of

cases. If the number of cases differs greatly, weighted least squares or logistic regression should be utilized. ■

20.2 Tukey Test for Additivity

We describe now the Tukey test that may be used for examining, when $n = 1$, whether or not the two factors in a two-factor study interact. This test is also useful for a variety of experimental designs to be discussed in later chapters.

Development of Test Statistic

As noted in Section 20.1, we considered no-interaction model (20.1) when $n = 1$ to enable us to obtain an estimate of the error variance in this case. It would have been possible, however, to impose less severe restrictions on the interaction effects $(\alpha\beta)_{ij}$ and include the restricted interaction effects in the ANOVA model. Suppose we assume that:

$$(\alpha\beta)_{ij} = D\alpha_i\beta_j \tag{20.6}$$

where D is some constant. One motivation for this restriction is that if $(\alpha\beta)_{ij}$ is any second-degree polynomial function of α_i and β_j, then it must be of the form (20.6) because of the restrictions on the α_i, β_j, and $(\alpha\beta)_{ij}$ in (19.23) that the sums over each subscript be zero.

Using (20.6) in a regular two-factor ANOVA model with interactions for the case $n = 1$, we obtain:

$$Y_{ij} = \mu_{..} + \alpha_i + \beta_j + D\alpha_i\beta_j + \varepsilon_{ij} \tag{20.7}$$

where each term has the usual meaning. Remember there is no third subscript here because $n = 1$. The interaction sum of squares $\sum\sum D^2\alpha_i^2\beta_j^2$ now needs to be obtained. Assuming the other parameters are known, the least squares and maximum likelihood estimator of D turns out to be:

$$\hat{D} = \frac{\sum_i\sum_j \alpha_i\beta_j Y_{ij}}{\sum_i \alpha_i^2 \sum_j \beta_j^2} \tag{20.8}$$

The usual estimator of α_i is $\bar{Y}_{i.} - \bar{Y}_{..}$ and that of β_j is $\bar{Y}_{.j} - \bar{Y}_{..}$. Replacing the parameters in $\hat{D}$ by these estimators, we obtain:

$$\hat{D} = \frac{\sum_i\sum_j(\bar{Y}_{i.} - \bar{Y}_{..})(\bar{Y}_{.j} - \bar{Y}_{..})Y_{ij}}{\sum_i(\bar{Y}_{i.} - \bar{Y}_{..})^2\sum_j(\bar{Y}_{.j} - \bar{Y}_{..})^2} \tag{20.8a}$$

The sample counterpart of the interaction sum of squares $\sum\sum D^2\alpha_i^2\beta_j^2$ will be denoted by $SSAB^*$ to remind us that this interaction sum of squares is for the special form of interaction in model (20.7). Substituting the sample estimates into $\sum\sum D^2\alpha_i^2\beta_j^2$, we obtain the interaction sum of squares:

$$\begin{aligned} SSAB^* &= \sum_i\sum_j \hat{D}^2(\bar{Y}_{i.} - \bar{Y}_{..})^2(\bar{Y}_{.j} - \bar{Y}_{..})^2 \\ &= \frac{\left[\sum_i\sum_j(\bar{Y}_{i.} - \bar{Y}_{..})(\bar{Y}_{.j} - \bar{Y}_{..})Y_{ij}\right]^2}{\sum_i(\bar{Y}_{i.} - \bar{Y}_{..})^2\sum_j(\bar{Y}_{.j} - \bar{Y}_{..})^2} \end{aligned} \tag{20.9}$$

The analysis of variance decomposition for the special interaction model (20.7) therefore is:

$$SSTO = SSA + SSB + SSAB^* + SSRem^* \tag{20.10}$$

where $SSRem^*$ is the *remainder sum of squares:*

$$SSRem^* = SSTO - SSA - SSB - SSAB^* \tag{20.10a}$$

It can be shown that if $D = 0$—that is, if no interactions of the type $D\alpha_i\beta_j$ exist—$SSAB^*$ and $SSRem^*$ are independently distributed as chi-square random variables with 1 and $ab - a - b$ degrees of freedom, respectively. Hence, if $D = 0$, the test statistic:

$$F^* = \frac{SSAB^*}{1} \div \frac{SSRem^*}{ab - a - b} \tag{20.11}$$

is distributed as $F(1, ab - a - b)$.

Thus, for testing:

$$\begin{aligned} &H_0\colon D = 0 && \text{(no interactions present)} \\ &H_a\colon D \neq 0 && \text{(interactions } D\alpha_i\beta_j \text{ present)} \end{aligned} \tag{20.12a}$$

we use test statistic F^* defined in (20.11). Large values of F^* lead to conclusion H_a. The appropriate decision rule for controlling the risk of a Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F(1-\alpha; 1, ab - a - b), \text{ conclude } H_0 \\ &\text{If } F^* > F(1-\alpha; 1, ab - a - b), \text{ conclude } H_a \end{aligned} \tag{20.12b}$$

The power of this test has been studied, and it appears that if interactions of approximately the type postulated in (20.6) are present and the factor A and factor B main effects are large, the test is effective in detecting the interactions. The test is usually called the *Tukey one degree of freedom test.* This test also may be used for testing for the presence of general interactions.

Example

We shall conduct the Tukey test for the insurance premium example. The data are presented in Table 20.2a. First, we obtain the elements of $SSAB^*$:

$$\begin{aligned} \sum\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})Y_{ij} &= (120 - 175)(190 - 175)(140) + \cdots \\ &\quad + (210 - 175)(160 - 175)(200) = -13{,}500 \\ \sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 &= \frac{SSA}{2} = \frac{9{,}300}{2} = 4{,}650 \\ \sum(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 &= \frac{SSB}{3} = \frac{1{,}350}{3} = 450 \end{aligned}$$

Hence, the special interaction sum of squares is:

$$SSAB^* = \frac{(-13{,}500)^2}{4{,}650(450)} = 87.1$$

Using the ANOVA sums of squares in Table 20.2b, we have by (20.10a):

$$SSRem^* = 10{,}750 - 9{,}300 - 1{,}350 - 87.1 = 12.9$$

Finally, we obtain the test statistic by (20.11):

$$F^* = \frac{87.1}{1} \div \frac{12.9}{3(2) - 3 - 2} = 6.8$$

For $\alpha = .10$, we require $F(.90; 1, 1) = 39.9$. Since $F^* = 6.8 \leq 39.9$, we conclude that region and size of city do not interact. The P-value of this test is .23. Use of the no-interaction model for the data in Table 20.2a therefore appears to be reasonable.

Remedial Actions if Interaction Effects Are Present

When the Tukey test indicates the presence of interaction effects in an analysis of variance application where $n = 1$, efforts should be made to remove the interactions so that the analysis described in Section 20.1 can be utilized. As we described in Chapter 19, transformations of the data can often be used to remove interaction effects or to make them unimportant.

One possibility is to try simple transformations of the response variable, such as a square root or a logarithmic transformation. Another possibility is to search in the family of power transformations on Y described in Chapter 3 in connection with the Box-Cox transformations. The procedure is to make transformations on Y according to (3.36) for selected values of λ. For each value of λ, the Tukey test statistic (20.11) is then obtained. If a λ value leads to a nonsignificant F^* test statistic, a transformation will then have been found that removes the interaction effect. Frequently, a range of λ values will yield nonsignificant test statistics, in which case a simple λ value in this range, such as $\lambda = .5$, may be chosen.

If no transformation can be found to make the interactions unimportant, an approximate method of analysis can be employed; see, for instance, Reference 20.1.

Comment

If one or both factors are quantitative, a test for interactions effects can be obtained by regression methods. For example, consider a study in which both factors are quantitative, each has three levels, and $n = 1$ so that $n_T = 9$. Let X_{ij1} denote the value of the first factor for the treatment for which factor A is at the ith level and factor B is at the jth level. X_{ij2} is defined similarly for the second factor. Second-order regression model (8.7) may then be used:

$$Y_{ij} = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij1}^2 + \beta_4 x_{ij2}^2 + \beta_5 x_{ij1} x_{ij2} + \varepsilon_{ij}$$

where:

$$x_{ij1} = X_{ij1} - \overline{X}_1$$
$$x_{ij2} = X_{ij2} - \overline{X}_2$$

With this model, there would be $n_T - p = 9 - 6 = 3$ degrees of freedom for estimating the error variance σ^2, and the test for the presence of an interaction effect would be the usual test in (6.51) for testing whether $\beta_5 = 0$.

Still other tests for interactions could be conducted since additional cross-product terms could be incorporated into the regression model. However, this would not be desirable here since the number of degrees of freedom available for estimating the error variance σ^2 is already very small. ■

ed erence

20.1. Johnson, D. E., and F. A. Graybill. "Estimation of σ^2 in a Two-Way Classification Model with Interaction," *Journal of the American Statistical Association* 67 (1972), pp. 388–94.

oblems

20.1. Suppose that two-factor analysis of variance model (19.23) were to be employed with $n = 1$ for each factor level combination. How many degrees of freedom would be associated with *SSE* in (19.37c)? What does this imply?

*20.2. **Coin-operated terminals.** A university computer service conducted an experiment in which one coin-operated computer graphics terminal was placed at each of four different locations on the campus last semester during the midterm week and again during the final week of classes. The data that follow show the number of hours each terminal was *not* in use during the week at the four locations (factor A) and for the two different weeks (factor B).

	Factor B (week)	
Factor A (location)	$j = 1$ Midterm	$j = 2$ Final
$i = 1$	16.5	21.4
$i = 2$	11.8	17.3
$i = 3$	12.3	16.9
$i = 4$	16.6	21.0

Assume that no-interaction ANOVA model (20.1) is appropriate.

a. Plot the data in the format of Figure 20.1. Does it appear that interaction effects are present? Does it appear that factor A and factor B main effects are present? Discuss.

b. Conduct separate tests for location and week main effects. In each test, use level of significance $\alpha = .05$ and state the alternatives, decision rule, and conclusion. Give an upper bound for the family level of significance; use the Kimball inequality (19.53). What is the P-value for each test?

c. Make all pairwise comparisons among location means and estimate the difference between the means for the two weeks; use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

*20.3. Refer to **Coin-operated terminals** Problem 20.2. It is desired to estimate μ_{32}.

a. Obtain a point estimate of μ_{32} using (20.5).

b. Obtain the estimated variance of $\hat{\mu}_{32}$ by fitting the equivalent regression model.

c. Construct a 95 percent confidence interval for μ_{32}. Interpret your interval estimate. Is your interval estimate applicable if next year two graphics terminals will be placed at location 3? Explain.

*20.4. Refer to **Coin-operated terminals** Problem 20.2. Conduct the Tukey test for additivity; use $\alpha = .025$. State the alternatives, decision rule, and conclusion. If the additive model is not appropriate, what might you do?

20.5. **Brainstorming.** A researcher investigated whether brainstorming is more effective for larger groups than for smaller ones by setting up four groups of agribusiness executives, the group sizes being two, three, four, and five, respectively. He also set up four groups of agribusiness scientists, the group sizes being the same as for the agribusiness executives. The researcher gave each group the same problem: "How can Canada increase the value of its agricultural

exports?" Each group was allowed 30 minutes to generate ideas. The variable of interest was the number of different ideas proposed by the group. The results, classified by type of group (factor A) and size of group (factor B), were:

	Factor A (type of group)	Factor B (size of group) $j = 1$ Two	$j = 2$ Three	$j = 3$ Four	$j = 4$ Five
$i = 1$	Agribusiness executives	18	22	31	32
$i = 2$	Agribusiness scientists	15	23	29	33

Assume that no-interaction ANOVA model (20.1) is appropriate.

a. Plot the data in the format of Figure 20.1. Does it appear that interaction effects are present? Does it appear that factor A and factor B main effects are present? Discuss.

b. Conduct separate tests for type of group and size of group main effects. In each test, use level of significance $\alpha = .01$ and state the alternatives, decision rule, and conclusion. Give an upper bound for the family level of significance; use the Kimball inequality (19.53). What is the P-value for each test?

c. Obtain confidence intervals for $D_1 = \mu_{\cdot 2} - \mu_{\cdot 1}$, $D_2 = \mu_{\cdot 3} - \mu_{\cdot 2}$, and $D_3 = \mu_{\cdot 4} - \mu_{\cdot 3}$; use the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

d. Is the Bonferroni procedure used in part (c) the most efficient one here? Explain.

20.6. Refer to **Brainstorming** Problem 20.5. It is desired to estimate μ_{14}.

a. Obtain a point estimate of μ_{14} using (20.5).

b. Obtain the estimated variance of $\hat{\mu}_{14}$ by fitting the equivalent regression model.

c. Construct a 99 percent confidence interval for μ_{14}. Interpret your interval estimate. Is your interval estimate applicable if the two factors interact?

20.7. Refer to **Brainstorming** Problem 20.5. Conduct the Tukey test for additivity: use $\alpha = .01$. State the alternatives, decision rule, and conclusion. If the additive model is not appropriate, what might you do?

20.8. **Soybean sausage.** A food technologist, testing storage capabilities for a newly developed type of imitation sausage made from soybeans, conducted an experiment to test the effects of humidity level (factor A) and temperature level (factor B) in the freezer compartment on color change in the sausage. Three humidity levels and four temperature levels were considered. Five hundred sausages were stored at each of the 12 humidity-temperature combinations for 90 days. At the end of the storage period, the researcher determined the proportion of sausages for each humidity-temperature combination that exhibited color changes. The researcher transformed the data by means of the arcsine transformation (18.24) to stabilize the variances. The transformed data $Y' = 2 \arcsin \sqrt{Y}$ follow.

Factor A (humidity level)	Factor B (temperature level) $j = 1$	$j = 2$	$j = 3$	$j = 4$
$i = 1$	13.9	14.2	20.5	24.8
$i = 2$	15.7	16.3	21.7	23.6
$i = 3$	15.1	15.4	19.9	26.1

Assume that no-interaction ANOVA model (20.1) is appropriate.

a. Plot the data in the format of Figure 20.1. Does it appear that interaction effects are present? Does it appear that factor A and factor B main effects are present? Discuss.

b. Conduct separate tests for humidity and temperature main effects. In each test, use level of significance $\alpha = .025$ and state the alternatives, decision rule, and conclusion. What is the P-value for each test?

c. Obtain confidence intervals for $D_1 = \mu_{\cdot 2} - \mu_{\cdot 1}$, $D_2 = \mu_{\cdot 3} - \mu_{\cdot 2}$, and $D_3 = \mu_{\cdot 4} - \mu_{\cdot 3}$; use the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

d. Is the Bonferroni procedure used in part (c) the most efficient one here? Explain.

20.9. Refer to **Soybean sausage** Problem 20.8. It is desired to estimate μ_{23}.

a. Obtain a point estimate of μ_{23} using (20.5).

b. Obtain the estimated variance of $\hat{\mu}_{23}$ by fitting the equivalent regression model.

c. Construct a 98 percent confidence interval for μ_{23} and transform it back to the original units. Interpret your interval estimate. Is your interval estimate applicable if the two factors interact?

20.10. Refer to **Soybean sausage** Problem 20.8. Conduct the Tukey test for additivity; use $\alpha = .005$. State the alternatives, decision rule, and conclusion. If the additive model is not appropriate, what might you do?

Exercises

20.11. Modify formulas (19.39a) and (19.39b) to apply to ANOVA model (20.1), where $n = 1$.

20.12. Show that (20.7) is the only second-degree polynomial function of α_i and β_j such that $\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$.

Case Study

20.13. Refer to **Soybean sausage** Problem 20.8. Assume that the humidity levels and temperature levels employed are equally spaced—that is, actual humidity increases linearly with i, and actual temperature increases linearly with j so that i and j are coded levels of humidity and temperature. Use techniques discussed in Chapter 8 to develop a polynomial regression model to predict the transformed percentage of sausages exhibiting color change as a function of coded humidity and temperature levels. Your model should consider, at most, second-order terms in coded humidity level, and third-order terms in coded temperature level. What does your model suggest concerning the presence or absence of interactions? Use appropriate graphics to summarize your fitted regression model.

Chapter

Randomized Complete Block Designs

In Chapter 15, we introduced the concept of blocking. We noted there that when the available experimental units are not homogeneous, grouping the experimental units into blocks of homogeneous units will reduce the experimental error variance and also increase the range of validity for inferences about the treatment effects.

In this chapter, we shall take up the design and analysis of randomized complete block experiments in detail. We discuss when and how to conduct a randomized complete block design, the analysis of a randomized complete block design, and planning of sample sizes for blocked experiments.

For complete block designs, each block consists of one complete replication of the set of treatments. When the number of experimental units available in a block is less than the number of treatments, incomplete block designs may at times be useful. We shall consider incomplete block designs in Chapters 28 and 29.

21.1 Elements of Randomized Complete Block Designs

Description of Designs

In a *randomized complete block design*, the experimental units are first sorted into homogeneous groups, called *blocks*, and all treatment combinations are then assigned at random to experimental units within the blocks. Note that this requires a series of separate, restricted randomizations—one for each block. In effect, separate experiments are conducted within each block, which leads to greater homogeneity of experimental units, reduced experimental error, and more precise estimates of treatment effects. We illustrate the use of randomized block designs by considering three examples.

1. In an experiment on the effects of four levels of newspaper advertising saturation on sales volume, the experimental unit is a city, and 16 cities are available for the study. Size of city usually is highly correlated with the response variable, sales volume. Hence, it is desirable to block the 16 cities into four groups of four cities each, according to population size. Thus, the four largest cities will constitute block 1, and so on. Within each block, the four treatments are then assigned at random to the four cities, and the four randomizations, one for each block, are conducted independently.

2. In an experiment on the effects of three different incentive pay schemes on employee productivity of electronic assemblies, the experimental unit is an employee, and 30 employees are available for the study. Since productivity here is highly correlated with manual dexterity, it is desirable to block the 30 employees into 10 groups of three according to their manual dexterity. Thus, the three employees with the highest manual dexterity ratings are grouped into one block, and so on for the other employees. Within each block, the three incentive pay schemes are then assigned randomly to the three employees.

3. A chemist is studying the reaction rate of five chemical agents. Only five agents can be analyzed effectively per day. Since day-to-day differences may affect the reaction rate, each day is used as a block, and all five chemical agents are tested each day in independently randomized orders.

As these examples imply, the key objective in blocking the experimental units is to make them as homogeneous as possible within blocks with respect to the response variable under study, and to make the different blocks as heterogeneous as possible with respect to the response variable. As noted earlier, the design in which each treatment is included once in each block is called a randomized complete block design. Often, we shall drop the term "complete" because the context makes it clear that all treatments are included in each block.

Comments

1. In a complete block design, each block constitutes a replication of the experiment. For that reason, it is highly desirable that the experimental units within a block be processed together whenever this will help to reduce experimental error variability. As an example, an experimenter may tend to make changes in experimental techniques over time (e.g., in the administration of the experiment to subjects) without being aware of it. Consecutive processing of the experimental units block by block will tend to exclude such sources of variation from the within-blocks analysis and thereby make the experimental results more precise.

2. In factorial experiments, some of the factors of interest may be characteristics of the experimental units, such as gender, age, and amount of experience on the job. Even though these factors are not introduced to reduce experimental error variability but rather are included for their intrinsic interest, we shall nevertheless consider such experiments to be randomized block designs because the randomization of the experimental factors to the experimental units is restricted by the nature of the observational factors considered. ■

Criteria for Blocking

As noted earlier, the purpose of blocking is to sort experimental units into groups within each of which the elements are homogeneous with respect to the response variable, such that the differences between groups are as great as possible. To help recognize some of the characteristics of experimental units that are useful criteria for blocking, we need a precise definition of an experimental unit. Any aspect of the experimental setting that changes from treatment application to treatment application—excluding the treatment changes themselves—is a characteristic of the experimental unit. For example, suppose the treatment in a taste-testing experiment consists of a vegetable containing a particular additive. The experimental unit might then be defined as a homemaker of a given age, evaluated by a given observer on a specified day at a particular time, and served food from a given batch of cooked vegetable. Still other elements of the experimental setting might be included in the definition of the experimental unit, and should be if they contribute to variability in the responses.

A full definition of the experimental unit such as the one just given suggests two types of blocking criteria:

1. Characteristics associated with the unit—for persons: gender, age, income, intelligence, education, job experience, attitudes, etc.; for geographic areas: population size, average income, etc.
2. Characteristics associated with the experimental setting—observer, time of processing, machine, batch of material, measuring instrument, etc.

Use of time as a blocking variable frequently captures a number of different sources of variability, such as learning by observer, changes in equipment, and drifts in environmental conditions (e.g., weather). Blocking by observers often eliminates a substantial amount of interobserver variability; similarly, blocking by batches of material frequently is very effective. There is no need to use only a single blocking criterion: several may be employed if the experimental error can be further reduced by doing so.

The design of an effective randomized block experiment requires the ability to anticipate potential sources of variation—the blocking variables—in advance of experimentation. These variables are then held constant within blocks as the experiment is conducted in order to reduce the experimental error variability. Often, past experience in the subject matter field enables the experimenter to select good blocking variables. If some experiments have been run in the past in which blocking has been employed, these results can be analyzed to determine the effectiveness of the blocking variables. In the absence of any information on the effectiveness of potential blocking variables, uniformity trials can be run where all experimental units are assigned the same treatment. From these trials, information can be obtained on the effectiveness of different blocking variables.

Comment

As noted in Chapter 15, when subjects are used as a blocking variable, the resulting design is sometimes called a *repeated measures design*. Since these designs involve some special problems, we will discuss them separately in Chapter 27. ■

Advantages and Disadvantages

The advantages of a randomized complete block design are:

1. It can, with effective grouping, provide substantially more precise results than a completely randomized design of comparable size.
2. It can accommodate any number of treatments and replications.
3. Different treatments need not have equal sample sizes. For instance, if the control is to have twice as large a sample size as each of three treatments, blocks of size five would be used; three units in a block are then assigned at random to the three treatments and two to the control.
4. The statistical analysis is relatively simple.
5. If an entire treatment or a block needs to be dropped from the analysis for some reason, such as spoiled results, the analysis is not complicated thereby.
6. Variability in experimental units can be deliberately introduced to widen the range of validity of the experimental results without sacrificing the precision of the results.

Disadvantages include:

1. If observations are missing within a block, a more complex analysis is required.
2. The degrees of freedom for experimental error are not as large as with a completely randomized design. One degree of freedom is lost for each block after the first.
3. More assumptions are required for the model (e.g., no interactions between treatments and blocks, constant variance from block to block) than for a completely randomized design model.
4. Because the blocking variable is an observational factor and not an experimental factor, cause-and-effect inferences concerning the relationship between the blocking variable and the response variable is problematic. This is not a serious disadvantage, because investigators usually are not concerned with estimating or making inference about block effects. Blocking is primarily a device for reducing experimental variation and thereby increasing the precision of the estimates of the treatment effects.

How to Randomize

The randomization procedure for a randomized block design is straightforward. Within each block a random permutation is used to assign treatments to experimental units, just as in a completely randomized design. Independent permutations are selected for the several blocks.

Illustration

In an experiment on decision making, executives were exposed to one of three methods of quantifying the maximum risk premium they would be willing to pay to avoid uncertainty in a business decision. The three methods are the utility method, the worry method, and the comparison method. After using the assigned method, the subjects were asked to state their degree of confidence in the method of quantifying the risk premium on a scale from 0 (no confidence) to 20 (highest confidence).

Fifteen subjects were used in the study. They were grouped into five blocks of three executives, according to age. Block 1 contained the three oldest executives, and so on. The design layout, after five independent random permutations of three were employed, is shown in Figure 21.1. Table 21.1 contains the results of the experiment, and Figure 21.2

FIGURE 21.1 Layout for Randomized Complete Block Design—Risk Premium Example.

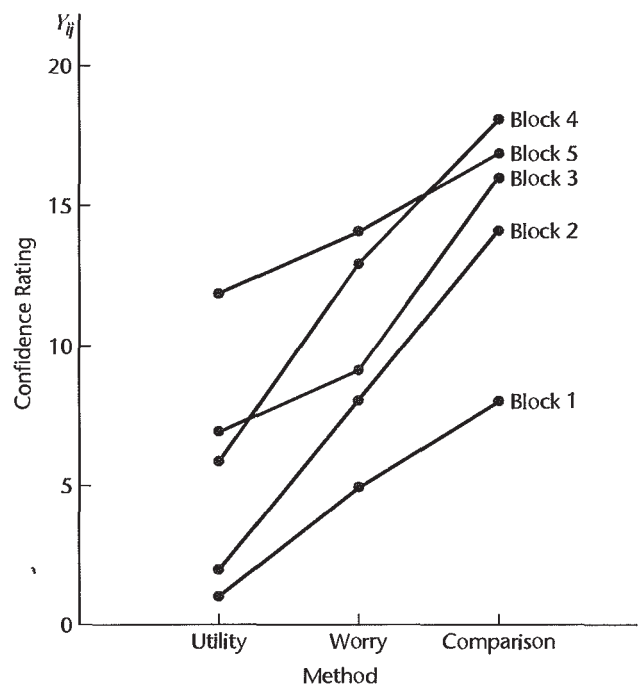
	Experimental Unit		
	1	2	3
Block 1 (oldest executives)	C	W	U
2	C	U	W
3	U	W	C
4	W	U	C
5 (youngest executives)	W	C	U

C : Comparison method
W : Worry method
U : Utility method

TABLE 21.1
Data on Confidence Ratings (ratings on scale from 0 to 20)—Risk Premium Example.

Block i	Method (j) Utility	Worry	Comparison	Average
1 (oldest)	1	5	8	4.7
2	2	8	14	8.0
3	7	9	16	10.7
4	6	13	18	12.3
5 (youngest)	12	14	17	14.3
Average	5.6	9.8	14.6	10.0

FIGURE 21.2
Plot of Confidence Ratings by Blocks—Risk Premium Example.



presents graphically the confidence ratings for each method by block. It appears from Figure 21.2 that there is much variation between blocks, but that in all blocks the comparison method has the highest confidence rating and the utility method the lowest. It also appears that there are no important interaction effects between blocks and treatments on the responses; the response curves do not seem to deviate too much from being parallel. We discuss next a widely used model for randomized complete block designs and the analysis of variance for this model before undertaking a formal analysis of the results in our example.

21.2 Model for Randomized Complete Block Designs

Table 21.1 is similar in appearance to Table 20.2a, which shows the data for a two-factor study with one observation in each cell. In fact, a randomized complete block design may be viewed as corresponding to a two-factor study (blocks and treatments are the factors), with one observation in each cell. As we noted in Section 20.1, the assumption of no interactions between the two factors permits an analysis of factor effects when there is only one observation in each cell and the factors have fixed effects.

The model for a randomized complete block design containing the assumption of no interaction effects, when both the block and treatment effects are fixed and there are n_b blocks (replications) and r treatments, is as follows:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \varepsilon_{ij} \tag{21.1}$$

where:

$\mu_{..}$ is a constant
ρ_i are constants for the block (row) effects, subject to the restriction $\sum \rho_i = 0$
τ_j are constants for the treatment effects, subject to the restriction $\sum \tau_j = 0$
ε_{ij} are independent $N(0, \sigma^2)$
$i = 1, \ldots, n_b; j = 1, \ldots, r$

The responses Y_{ij} with randomized block model (21.1) are independent and normally distributed, with mean:

$$E\{Y_{ij}\} = \mu_{..} + \rho_i + \tau_j \tag{21.2a}$$

and constant variance:

$$\sigma^2\{Y_{ij}\} = \sigma^2 \tag{21.2b}$$

Randomized block model (21.1) is identical to the two-factor, no-interaction model (20.1), except that we now use ρ_i for the block effect, τ_j for the treatment effect, and n_b to designate the total number of blocks. Note that Y_{ij} here stands for the response for the jth treatment in the ith block.

Comments

1. When the experimental units are grouped according to specified categories, such as into particular age groups, income groups, and order-of-processing groups, the block effects ρ_i are usually considered to be fixed. Sometimes the block effects are viewed as random. For instance, when observers or subjects are used as blocks, the particular observers or subjects in the study may be considered to be a sample from a population of observers or subjects. The case of random block effects will be taken up in Chapter 25.

2. If the treatment effects are random, the only changes in model (21.1) are that the τ_j now represent independent normal variables with expectation zero and variance σ_τ^2, and that the τ_j are independent of the ε_{ij}. Random treatment effects are also considered in Chapter 25.

3. The additive model (21.1) implies that the expected values of observations in different blocks for the same treatment may differ (e.g., older executives may tend to have lower confidence ratings for any of the methods of quantifying the risk premium than younger executives), but the treatment

effects (e.g., how much higher the confidence rating for one method is over that for another) are the same for all blocks. We shall consider the possibility of interactions between blocks and treatments later in Section 21.7. ■

21.3 Analysis of Variance and Tests

Fitting of Randomized Complete Block Model

The least squares and maximum likelihood estimators of the parameters in randomized block model (21.1) are obtained in the customary fashion and again are the same. Employing our usual notation, they are:

Parameter	Estimator	
$\mu_{..}$	$\hat{\mu}_{..} = \bar{Y}_{..}$	(21.3a)
ρ_i	$\hat{\rho}_i = \bar{Y}_{i.} - \bar{Y}_{..}$	(21.3b)
τ_j	$\hat{\tau}_j = \bar{Y}_{.j} - \bar{Y}_{..}$	(21.3c)

The fitted values therefore are:

$$\hat{Y}_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..}) = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..} \tag{21.4}$$

and the residuals are:

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..} \tag{21.5}$$

Analysis of Variance

The analysis of variance for a randomized complete block design is identical to that for a two-factor, no-interaction model with one observation per cell, as described in Section 20.1:

$$SSBL = r \sum_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 \tag{21.6a}$$

$$SSTR = n_b \sum_j (\bar{Y}_{.j} - \bar{Y}_{..})^2 \tag{21.6b}$$

$$SSBL.TR = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2 = \sum_i \sum_j e_{ij}^2 \tag{21.6c}$$

Here, *SSBL* denotes the *sum of squares for blocks*, *SSTR* denotes, as usual, the treatment sum of squares, and *SSBL.TR* denotes the *interaction sum of squares between blocks and treatments* [note from (21.5) that this sum of squares here is the same as the sum of the squared residuals]; rn_b is the total number of experimental units in the study.

A summary of the analysis of variance, including the expected mean squares for fixed treatment effects, is given in Table 21.2. Note that since there are no interaction terms in the model, the expected mean squares contain only σ^2 and, as appropriate, the treatment or block main effects term. Also note from the $E\{MS\}$ columns in Table 21.2 that the appropriate denominator in the F^* test statistic for testing treatment effects is the interaction mean square, here denoted by *MSBL.TR*. This is the same as in Section 20.1 for the two-factor

TABLE 21.2 ANOVA Table for Randomized Complete Block Design, Block Effects Fixed.

Source of Variation	*SS*	*df*	*MS*	$E\{MS\}$
Blocks	*SSBL*	$n_b - 1$	*MSBL*	$\sigma^2 + r\frac{\sum \rho_i^2}{n_b - 1}$
Treatments	*SSTR*	$r - 1$	*MSTR*	$\sigma^2 + n_b\frac{\sum \tau_j^2}{r - 1}$
Error	*SSBL.TR*	$(n_b - 1)(r - 1)$	*MSBL.TR*	σ^2
Total	*SSTO*	$n_b r - 1$		

no-interaction model with $n_b = 1$. Hence, to test for treatment effects:

Fixed Treatment Effects

$$H_0\text{: all } \tau_j = 0$$
$$H_a\text{: not all } \tau_j \text{ equal zero} \tag{21.7a}$$

we use the same test statistic:

$$F^* = \frac{MSTR}{MSBL.TR} \tag{21.7b}$$

and the decision rule for controlling the Type I error at α is:

$$\text{If } F^* \leq F[1 - \alpha; r - 1, (n_b - 1)(r - 1)], \text{ conclude } H_0$$
$$\text{If } F^* > F[1 - \alpha; r - 1, (n_b - 1)(r - 1)], \text{ conclude } H_a \tag{21.7c}$$

Example

Table 21.3 contains the analysis of variance for the risk premium example in Table 21.1. The calculations are straightforward and were carried out by a computer package. To test for treatment effects:

$$H_0\text{: } \tau_1 = \tau_2 = \tau_3 = 0$$
$$H_a\text{: not all } \tau_j \text{ equal zero}$$

TABLE 21.3 ANOVA Table for Randomized Complete Block Design—Risk Premium Example of Table 21.1.

Source of Variation	*SS*	*df*	*MS*
Blocks	171.3	4	42.8
Methods for risk premium specification	202.8	2	101.4
Error	23.9	8	2.99
Total	398.0	14	

we use the results in Table 21.3:

$$F^* = \frac{MSTR}{MSBL.TR} = \frac{101.4}{2.99} = 33.9$$

For level of significance $\alpha = .01$, we require $F(.99; 2, 8) = 8.65$. Since $F^* = 33.9 > 8.65$, we conclude H_a, that the mean confidence ratings for the three methods differ. The P-value of the test is .0001.

Comments

1. Sometimes one may also wish to conduct a test for block effects:

$$H_0\text{: all } \rho_i = 0$$
$$H_a\text{: not all } \rho_i \text{ equal zero} \tag{21.8a}$$

Usually, however, the treatments are of primary interest, and blocks are chiefly the means for reducing the experimental error variability. Table 21.2 indicates that the test for fixed block effects uses the test statistic:

$$F^* = \frac{MSBL}{MSBL.TR} \tag{21.8b}$$

For the risk premium example, this test statistic is:

$$F^* = \frac{42.8}{2.99} = 14.3$$

For level of significance $\alpha = .01$, we require $F(.99; 4, 8) = 7.01$. Since $F^* = 14.3 > 7.01$, we conclude that the mean confidence ratings (averaged over treatments) differ for the various blocks.

Since blocks correspond to an observational factor, care needs to be used in interpreting the implications of block effects. In our risk premium example, for instance, the block effects might not be due to age, even though age was the grouping variable. Education could be the pivotal explanatory variable, the effect by age appearing if older executives have less formal education than younger ones.

2. If only two treatments are investigated in a randomized complete block design, it can be shown that the F test for treatment effects based on test statistic (21.7b) is equivalent to the two-sided t test for paired observations based on test statistic (A.69).

3. When the responses Y_{ij} in a randomized complete block design are far from normally distributed and transformations of the data are not successful to meet the robustness properties of the standard inference procedures, a nonparametric test of treatment effects may be useful. The nonparametric rank F test introduced in Section 18.7 for single-factor studies is easily adapted for use in studies based on randomized complete block designs. The r observations in each block are ranked from 1 to r in ascending order and the usual F^* test in (21.7b) for testing treatment effects in a randomized block design is carried out, but now based on the ranked data. We use F_R^* to denote the F^* test statistic when the test is based on the ranked data.

The rank F test statistic is equivalent to the statistic for the *Friedman test*, a widely used nonparametric rank procedure for testing the equality of treatment means in randomized complete block designs. The Friedman test is also based on the within-block ranks R_{ij} of the data. The Friedman test statistic is:

$$X_F^2 = SSTR \div \frac{SSTR + SSBL.TR}{n_b(r-1)}$$

which can be reduced to (when no ties are present):

$$X_F^2 = \left[\frac{12}{n_b r(r+1)} \sum_j R_{\cdot j}^2\right] - 3n_b(r+1)$$

Instead of using the F distribution, the Friedman test approximates the distribution of X_F^2 when H_0 holds by the chi-square distribution with $r-1$ degrees of freedom, provided that the number of blocks is not too small. The decision rule is therefore:

$$\text{If } X_F^2 \leq \chi^2(1-\alpha; r-1), \text{ conclude } H_0$$

$$\text{If } X_F^2 > \chi^2(1-\alpha; r-1), \text{ conclude } H_a$$

The rank F test statistic F_R^* and the X_F^2 test statistic are related as follows:

$$F_R^* = \frac{(n_b - 1)X_F^2}{n_b(r-1) - X_F^2}$$

■

21.4 Evaluation of Appropriateness of Randomized Complete Block Model

The importance of examining the appropriateness of a statistical model for a given set of data has been mentioned many times. Since the techniques of examination are similar, we shall make only a few points of special relevance to randomized complete block designs here.

Diagnostic Plots

Some of the chief ways in which the data may not fit randomized complete block model (21.1) are:

1. Unequal error variability by blocks
2. Unequal error variability by treatments
3. Time effects
4. Block-treatment interactions

Use of residual plots in connection with points 2 and 3 has been considered in Section 18.1 with reference to a completely randomized design. The discussion there applies also to the residuals of a randomized complete block design, given in (21.5):

$$e_{ij} = Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot}$$

We simply add here that if treatments do have unequal error variability in a randomized complete block design, the differences between any two treatments can always be estimated by working with the differences between the paired observations, $Y_{ij} - Y_{ij'}$, which are unaffected by any unequal treatment variances.

Unequal error variability by blocks can be studied by aligned residual dot plots for each block, as shown in Figure 21.3 for a randomized block study with 10 treatments run in three blocks. The residual dot plots in Figure 21.3 are suggestive of increasing error variability with increasing block number. If, for instance, the blocks were processed in block number order, some modifications in procedures may have taken place leading to larger experimental error variability over time. Tests concerning the equality of variances, such as those described in Section 18.2, may be employed for a more formal determination, provided that the sample sizes are reasonably large so that the residuals can be treated as if they were independent.

FIGURE 21.3 Residual Dot Plots Suggesting Unequal Error Variances by Blocks.

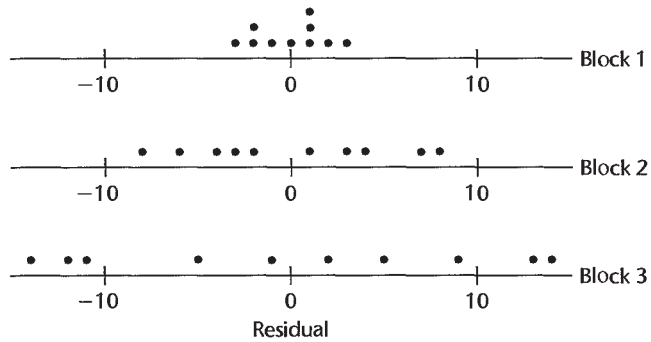
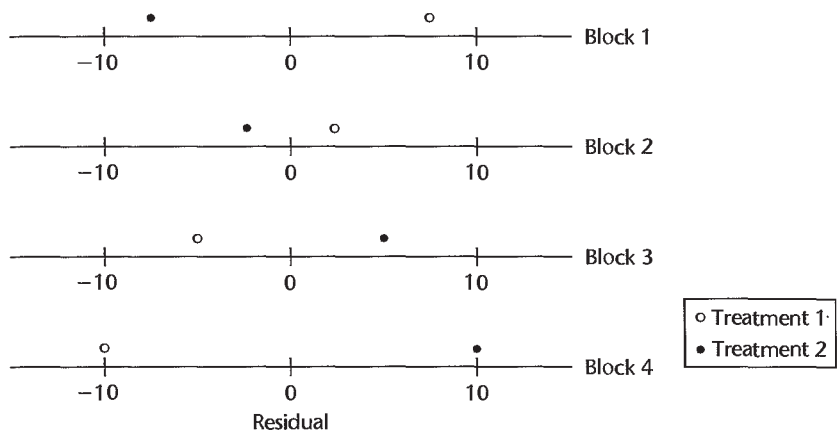


FIGURE 21.4 Residual Dot Plots Suggesting Block-Treatment Interactions.



Interactions between treatments and blocks are somewhat more difficult to detect from residual plots. Figure 21.4 contains the residuals for an experiment with two treatments run in four blocks. The reversal in pattern of the residuals is suggestive of an interaction effect. There are, however, many other possible types of interaction patterns that would appear very much different from that in Figure 21.4.

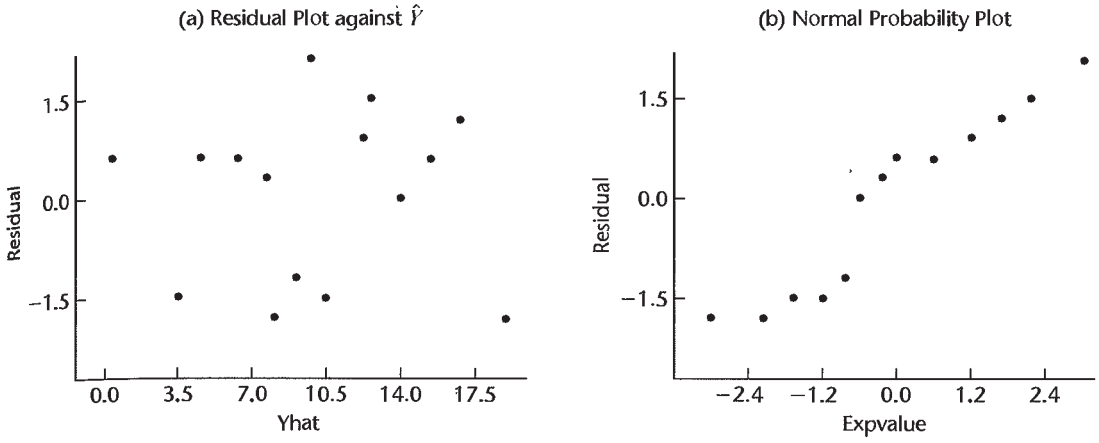
Another diagnostic plot that may be helpful to detect interaction effects is a plot of the residuals e_{ij} against the fitted values $\hat{Y}_{ij}$. A curvilinear pattern of the residuals in such a plot often suggests the presence of interaction effects between blocks and treatments. This plot also provides information about the constancy of the error variance.

Still another diagnostic plot for interactions, which is often more effective than a residual plot, is a plot of the responses Y_{ij} by blocks. Figure 21.2 illustrates this type of plot. A severe lack of parallelism in such a plot is a strong indication that blocks and treatments interact in their effects on the response.

Example

We already noted that the plot of responses by block in Figure 21.2 for the risk premium example does not exhibit a severe lack of parallelism, thus suggesting that blocks and treatments do not interact in any major fashion. Figure 21.5a, which presents a plot of the residuals against the fitted values, leads to a similar conclusion. There is no strong evidence

FIGURE 21.5 Diagnostic Residual Plots—Risk Premium Example.



of a curvilinear pattern here. In addition, Figure 21.5a does not indicate the existence of substantially unequal error variances.

Figure 21.5b contains a normal probability plot of the residuals. This plot does not suggest any strong departures from a normal error distribution. The coefficient of correlation between the ordered residuals and their expected values under normality is .959 and supports this conclusion. Residual dot plots for each treatment and for each block were also prepared (they are not shown here). They suggested that the error variances did not differ substantially between treatments and between blocks. These results, in addition to a formal test that found no interactions between block and treatment effects (to be discussed next), led the analyst to conclude that randomized block model (21.1) is appropriate for the data.

Tukey Test for Additivity

The Tukey test for additivity, discussed in Section 20.2, may be employed for a formal test of interaction effects between blocks and treatments for a randomized block design. The special interaction sum of squares in (20.9) will be denoted here by $SSBL.TR^*$.

Example

To test for interaction effects between blocks and treatments in the risk premium example, we calculate the special interaction sum of squares in (20.9) as follows, using the data in Tables 21.1 and 21.3:

$$\sum\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})Y_{ij} = 24.80$$

$$\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = \frac{SSBL}{r} = \frac{171.3}{3} = 57.10$$

$$\sum(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 = \frac{SSTR}{n_b} = \frac{202.8}{5} = 40.56$$

Hence:

$$SSBL.TR^* = \frac{(24.80)^2}{57.10(40.56)} = .27$$

Using the results from Table 21.3, we can now obtain the remainder sum of squares (20.10a) for the special interaction model (20.7):

$$\begin{aligned} SSRem^* &= SSTO - SSBL - SSTR - SSBL.TR^* \\ &= 398.0 - 171.3 - 202.8 - .27 \\ &= 23.63 \end{aligned}$$

Hence, test statistic (20.11) is:

$$\begin{aligned} F^* &= \frac{SSBL.TR^*}{1} \div \frac{SSRem^*}{rn_b - r - n_b} \\ &= \frac{.27}{1} \div \frac{23.63}{7} = .08 \end{aligned}$$

For level of significance $\alpha = .05$, we need $F(.95; 1, 7) = 5.59$. Since $F^* = .08 \leq 5.59$, we conclude that no block-treatment interaction effects are present. The P-value of this test is .79.

Comment

When interaction effects are present, transformations of the data should be attempted to remove at least the important interaction effects. The discussion in Section 20.2 is relevant to this point. ■

21.5 Analysis of Treatment Effects

Once the existence of fixed treatment effects has been established through the analysis of variance, the analysis of these effects proceeds as described in Chapter 17 for single-factor studies. Often, a useful preliminary view of the treatment effects can be obtained from a bar-interval plot of the estimated treatment means $\bar{Y}_{.j}$. The formal analysis of the treatment effects usually involves estimation of one or more contrasts of the treatment means $\mu_{.j}$, where $\mu_{.j}$ is the mean response for treatment j averaged over all blocks. The formulas in Chapter 17 for estimating contrasts of the treatment means apply here, with the treatment means now denoted by $\mu_{.j}$ and the estimated treatment means by $\bar{Y}_{.j}$. The appropriate mean square term to be used in the estimated variance of the contrast is $MSBL.TR$, obtained from (21.6c), since it is the denominator of the F^* statistic for testing fixed treatment effects. The multiples for the estimated standard deviation of the contrast are now as follows:

Single comparison	$t[1 - \alpha/2; (n_b - 1)(r - 1)]$	**(21.9a)**
Tukey procedure (for pairwise comparisons)	$T = \frac{1}{\sqrt{2}} q[1 - \alpha; r, (n_b - 1)(r - 1)]$	**(21.9b)**
Scheffé procedure	$S^2 = (r - 1)F[1 - \alpha; r - 1, (n_b - 1)(r - 1)]$	**(21.9c)**
Bonferroni procedure	$B = t[1 - \alpha/2g; (n_b - 1)(r - 1)]$	**(21.9d)**

Example

The researcher who conducted the risk premium study was satisfied, on the basis of the residual analyses and tests, that randomized complete block model (21.1) is appropriate for the experiment. To analyze the treatment effects formally, the researcher wished to obtain all

pairwise comparisons with a 95 percent family confidence coefficient, utilizing the Tukey procedure. Using (17.30b), with *MSE* replaced by *MSBL.TR* and the results in Table 21.3, we obtain:

$$s^2\{\hat{L}\} = MSBL.TR\left(\frac{1}{n_b} + \frac{1}{n_b}\right) = \frac{2MSBL.TR}{n_b} = \frac{2(2.99)}{5} = 1.20$$

Remember that each estimated treatment mean $\bar{Y}_j$ consists of n_b observations (one from each of n_b blocks). Using (21.9b), we find for a 95 percent family confidence coefficient:

$$T = \frac{1}{\sqrt{2}}q(.95; 3, 8) = \frac{1}{\sqrt{2}}(4.04) = 2.86$$

Hence:

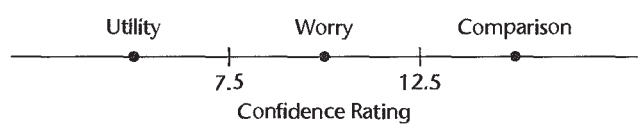
$$Ts\{\hat{L}\} = 2.86\sqrt{1.20} = 3.1$$

We now obtain for the pairwise comparisons using (17.30) and Table 21.1 for the $\bar{Y}_{\cdot j}$:

$$1.7 = (14.6 - 9.8) - 3.1 \leq \mu_{\cdot 3} - \mu_{\cdot 2} \leq (14.6 - 9.8) + 3.1 = 7.9$$
$$5.9 = (14.6 - 5.6) - 3.1 \leq \mu_{\cdot 3} - \mu_{\cdot 1} \leq (14.6 - 5.6) + 3.1 = 12.1$$
$$1.1 = (9.8 - 5.6) - 3.1 \leq \mu_{\cdot 2} - \mu_{\cdot 1} \leq (9.8 - 5.6) + 3.1 = 7.3$$

Here $\mu_{\cdot 1}$ is the mean confidence rating, averaged over all blocks, for the utility method, and $\mu_{\cdot 2}$ and $\mu_{\cdot 3}$ are the mean confidence ratings for the worry and comparison methods, respectively.

We conclude, just as Figure 21.2 suggests, that the comparison method has a higher mean confidence rating than the worry method, which in turn has a higher mean confidence rating than the utility method. The family confidence coefficient of .95 applies to this entire set of comparisons. A line plot of the estimated treatment means summarizes the results:



21.6 Use of More than One Blocking Variable

Sometimes, a substantial reduction in the experimental error variability can only be obtained by utilizing more than one variable for determining blocks. For instance, both age and gender might be needed for designating blocks:

Block	Characteristics of Experimental Units
1	Male, aged 20–29
2	Female, aged 20–29
3	Male, aged 30–39
4	Female, aged 30–39
etc.	etc.

As another example, both observer and day of treatment application may be helpful as blocking variables:

Block	Characteristics of Experimental Units
1	Observer 1, day 1
2	Observer 2, day 1
3	Observer 1, day 2
4	Observer 2, day 2
etc.	etc.

Unless the separate effects of each of the blocking variables need to be studied, no new problems arise when the blocks are defined by two or more variables. The n_b blocks are simply treated as ordinary blocks, and the usual block sum of squares is calculated.

When the effect of each of the blocking variables is to be isolated and the blocks are defined in a complete factorial fashion, the analysis simply treats each of the blocking variables as a factor and utilizes the methods developed in Chapter 19 for two-factor studies. For example, if twelve blocks are used when four observers and three days are employed for blocking, the analysis of variance would decompose the $12 - 1 = 8$ degrees of freedom for blocks into $4 - 1 = 3$ degrees of freedom for observer main effects, $3 - 1 = 2$ degrees of freedom for day main effects, and $3 \times 2 = 6$ degrees of freedom for observer $\times$ day interactions.

A problem that sometimes arises when two or more blocking variables are to be used is the large number of blocks called for. Suppose an experiment is to be conducted where the experimental units are stores. In order to reduce the experimental error variability to a reasonable level, it would be desirable to group the stores into six sales volume classes and also into six location classes (suburban shopping center, suburban other, etc.). Thirty-six blocks result from combining these two blocking variables. If six treatments were to be studied, 216 stores would be required for the experiment. Often, use of this many stores would be much too costly. Latin square designs, to be discussed in Chapter 28, permit in this type of study the use of a much smaller number of experimental units while still preserving the full benefits of error variance reduction by using both blocking variables in six classes each.

21.7 Use of More than One Replicate in Each Block

When block effects are fixed, use of an additive model in the presence of interactions between blocks and treatments has the effect of reducing the power of the test and increasing the width of interval estimates of treatment effects, thus making the experiment less sensitive. In addition, there are occasions when the nature of the interactions between blocks and treatments is of interest. It is possible to use a design that permits an interaction term in the model even when the block effects are fixed, and that allows the nature of the interaction effects to be investigated. This design is called a *generalized randomized block design*. It is the same as a randomized block design except that d experimental units are assigned to each treatment within a block. This generalized design increases the size of a block from r units for a randomized block design to dr units. The increase in block size will often have the effect of increasing experimental error variability when the total number of experimental

units is fixed. In the social sciences, however, increasing the size of the block moderately may cause little loss in efficiency. For instance, having one block of 10 persons aged 20–29 instead of two blocks of five persons of ages 20–24 and 25–29, respectively, will for many types of experiments involve little loss of efficiency.

As we shall demonstrate by an example, a generalized randomized block design is analyzed like an ordinary two-factor study where blocks are one factor. Hence, no new problems are encountered with a generalized randomized block design in testing for treatment effects or in estimating them. In particular, we can now calculate *MSE* and use it as an estimator of the error variance σ^2.

Example

Table 21.4 contains the data for a single-factor experiment in which the effects of distraction level (factor A: low distraction, high distraction) on the time required to complete a task were studied, using eight men and eight women. Four men were assigned at random to each of the $r = 2$ treatments, and independently four women were assigned at random to each treatment. Here gender is the blocking variable. Each block contains eight persons, with four randomly assigned to each treatment within the block. The layout in Table 21.4 corresponds to the layout in Table 19.7 for a two-factor study; to stress the correspondence, we have placed the blocks in columns rather than in rows as usual. Since blocks and distraction levels are considered to be fixed, we utilize the fixed effects two-factor model (19.23), with notation modified to fit the present context:

$$Y_{ijk} = \mu_{..} + \rho_i + \tau_j + (\rho\tau)_{ij} + \varepsilon_{ijk} \tag{21.10}$$

where:

$\mu_{..}$ is a constant
ρ_i, τ_j, are constants subject to the restrictions $\sum \rho_i = \sum \tau_j = 0$
$(\rho\tau)_{ij}$ are constants subject to the restrictions that the sums over any subscript are zero
ε_{ijk} are independent $N(0, \sigma^2)$
$i = 1, \ldots, n_b; j = 1, \ldots, r; k = 1, \ldots, d$

We shall refer to model (21.10) as the *generalized randomized block model.*

TABLE 21.4 Data on Completion Times for Generalized Randomized Block Design with $d = 4$—Task Completion Example.

	Block (gender)	
	Male	Female
Low Distraction:		
	12	3
	8	9
	7	5
	5	9
High Distraction:		
	14	11
	16	9
	15	10
	13	14

FIGURE 21.6 Portion of SAS GLM ANOVA Output for Data in Table 21.4—Task Completion Example ($n_b = 2, r = 2, d = 4$).

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	150.0000000	50.0000000	8.33	0.0029
Error	12	72.0000000	6.0000000		
Corrected Total	15	222.0000000			

R-Square	Coeff Var	Root MSE	y Mean
0.675676	24.49490	2.449490	10.00000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
Distraction	1	121.0000000	121.0000000	20.17	0.0007
Gender	1	25.0000000	25.0000000	4.17	0.0639
Dist*Gender	1	4.0000000	4.0000000	0.67	0.4301

The analysis of variance for generalized randomized block model (21.10) is the ordinary two-factor ANOVA of Table 19.8, with slight modifications in notation. The SAS GLM procedure was employed to obtain Figure 21.6 for the data in Table 21.4. We know from Table 19.8 that all test statistics use MSE in the denominator. These F^* statistics are shown in Figure 21.6. For $\alpha = .01$, we require $F(.99; 1, 12) = 9.33$ for each of the tests. It is evident from the results in Figure 21.6 (see also the P-values given there) that blocks (gender) do not interact with treatments (distraction level) and that high distraction level increases the time required to complete the task, compared to the low distraction level.

21.8 Factorial Treatments

Randomized complete block designs can also be used when the treatments have a factorial structure. For example, Figure 21.7 displays the layout for a randomized block design for a two-factor study, where each factor has two levels. Because the number of treatments is $r = ab = 4$, the block size here is four.

When factorial treatments are employed, the ANOVA model can be modified by showing the component factor effects in place of the treatment effect. For a two-factor study, we have:

$$Y_{ijk} = \mu_{...} + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \tag{21.11}$$

where the terms in the model have the usual meaning and (j, k) corresponds to the treatment mean $\mu_{.jk}$. In the analysis of variance, we proceed as always by decomposing the treatment sum of squares $SSTR$ into sums of squares for the factor main effects and interactions. This is shown in Table 21.5 for a two-factor study, the factors having a and b levels, respectively. The decomposition is done in the usual fashion, as explained in Section 19.4, utilizing the relation in (19.39):

$$SSTR = SSA + SSB + SSAB$$

FIGURE 21.7 Layout for a Two-factor Study in a Randomized Complete Block Design.

	A_1		A_2	
	B_1	B_2	B_1	B_2
Block 1	Y_{111}	Y_{112}	Y_{121}	Y_{122}
2	Y_{211}	Y_{212}	Y_{221}	Y_{222}
3	Y_{311}	Y_{312}	Y_{321}	Y_{322}

TABLE 21.5 ANOVA Table for a Two-Factor Study in a Randomized Complete Block Design—Randomized Block Model (21.11).

Source of Variation	*SS*	*df*	*MS*
Blocks	*SSBL*	$n_b - 1$	*MSBL*
Treatments	*SSTR*	$r - 1$	*MSTR*
Factor *A*	*SSA*	$a - 1$	*MSA*
Factor *B*	*SSB*	$b - 1$	*MSB*
AB interactions	*SSAB*	$(a-1)(b-1)$	*MSAB*
Error	*SSBL.TR*	$(n_b - 1)(r - 1)$	*MSBL.TR*
Total	*SSTO*	$n_b r - 1$	

Note: $r = ab$

Formulas (19.39a, b, c) are still appropriate for calculating the component sums of squares, remembering that (i, j) subscripts are there used to identify the treatments in terms of the factor level combinations. Tests for factor effects are conducted as usual, and no new problems are encountered in the estimation of fixed factor effects.

21.9 Planning Randomized Complete Block Experiments

The planning of sample sizes for a randomized complete block design is very similar to that for a completely randomized design. The needed number of blocks n_b may be determined either by specifying protection needed against making Type I and Type II errors or by specifying precision required for key contrasts of the treatment means. With either approach, it is necessary to assess in advance the magnitude of the experimental error variance σ^2.

Power Approach

Power of *F* Test. The power of the F test for treatment effects for a randomized complete block design involves the same noncentrality parameter as for a completely randomized design. Formula (16.88) gives the appropriate measure. Despite the same form of the noncentrality parameter, the two designs generally lead to different power levels even when based on the same sample sizes, for two reasons. First, the experimental error variance σ^2 will differ for the two designs. Second, the degrees of freedom associated with the denominator of the F^* statistic differ for the two designs.

Use of Table B.12. As when planning the sample sizes for a completely randomized design, an easy way to implement the power approach for planning the sample sizes for a randomized complete block design is to use Table B.12. This table may be used for planning randomized complete block designs provided that the number of treatments and blocks are not very small, specifically provided that $r(n_b - 1) \geq 20$. If this condition is not met, Table B.11 must be used iteratively to implement the power approach.

Example

In the risk premium example, suppose that the number of blocks had not yet been determined and the experimenter desired the following risk protections:

1. Type I error is to be controlled at $\alpha = .05$.
2. If any two treatment means differ by three or more rating points, i.e., if the minimum range of the treatment means is $\Delta = 3$, the risk of concluding that there are no treatment effects should not exceed $\beta = .20$.

The experimenter anticipates that the experimental error standard deviation when executives are grouped by age will be approximately $\sigma = 2$. Thus, the specifications can be summarized as follows:

$$r = 3 \qquad \alpha = .05 \qquad \Delta = 3$$
$$\beta = .20 \qquad 1 - \beta = .80 \qquad \sigma = 2$$

Using (16.91) we find:

$$\frac{\Delta}{\sigma} = \frac{3}{2} = 1.5$$

Entering Table B.12 for power $1 - \beta = .80$, $r = 3$, $\Delta/\sigma = 1.5$, and $\alpha = .05$, we find $n_b = 10$. Thus, the experimenter requires approximately 10 blocks of three executives each in order to obtain the desired protection against incorrect decisions.

Estimation Approach

For planning the number of blocks n_b by means of the estimation approach, we evaluate the anticipated standard deviations of key contrasts for different sample sizes until the desired precision is attained. Often, a multiple comparison procedure will be used for encompassing the different estimates under a family confidence coefficient.

Example

For the risk premium example, all pairwise comparisons are of interest. The desired width of the confidence intervals is ± 1.5. The Tukey procedure is to be used with a 95 percent family confidence coefficient. A planning value of $\sigma = 2$ is reasonable. Using $n_b = 10$ as a starting point, the anticipated variance of any pairwise difference is:

$$\sigma^2\{\hat{L}\} = \sigma^2\left(\frac{1}{n_b} + \frac{1}{n_b}\right) = (2)^2\left(\frac{1}{10} + \frac{1}{10}\right) = .8$$

or $\sigma\{\hat{L}\} = .89$. Further:

$$T = \frac{1}{\sqrt{2}}q[.95; r, (n_b - 1)(r - 1)] = \frac{1}{\sqrt{2}}q(.95; 3, 18) = \frac{1}{\sqrt{2}}(3.61) = 2.55$$

Thus, the anticipated half-width of the confidence interval is $T\sigma\{\hat{L}\} = 2.55(.89) = 2.3$. Since this precision is not adequate, a larger number of blocks should be tried next.

Continuing this iterative process, we find that $n_b = 21$ blocks are anticipated to meet the precision specification.

Efficiency of Blocking Variable

Once a randomized complete block experiment has been run, we often wish to estimate the efficiency of the blocking variable for guidance in future experimentation. This can be done readily. Let σ_b^2 stand for the experimental error variance for the randomized complete block design. Up to this point, we have used σ^2 for this error variance, but now that we will compare two designs we need to be more specific. Let σ_r^2 denote the experimental error variance for a completely randomized design. The relative efficiency of blocking, compared to a completely randomized design, is then defined as follows:

$$E = \frac{\sigma_r^2}{\sigma_b^2} \tag{21.12}$$

The measure E indicates how much larger the replications need be with a completely randomized design as compared to a randomized complete block design in order that the variance of any estimated treatment contrast be the same for the two designs.

We know that for the randomized block design, *MSBL.TR* is an unbiased estimator of σ_b^2. The question is how to estimate σ_r^2 from the data for the randomized block design. Since the same experimental units are involved in either case and there are assumed to be no interactions between treatments and blocks, it can be shown that an unbiased estimator of σ_r^2 is:

$$s_r^2 = \frac{(n_b - 1)MSBL + n_b(r-1)MSBL.TR}{n_b r - 1} \tag{21.13}$$

Hence, we estimate E as follows:

$$\hat{E} = \frac{s_r^2}{MSBL.TR} = \frac{(n_b - 1)MSBL + n_b(r-1)MSBL.TR}{(n_b r - 1)MSBL.TR} \tag{21.14}$$

Since the number of degrees of freedom for experimental error for a randomized block design is not as great as for a completely randomized design, E overstates the efficiency a little because it considers only the error variances. Several modified measures of efficiency have been suggested to take this overstatement into account. Unless the degrees of freedom for experimental error with both designs are very small, these modifications have little effect. One frequently used modification, applicable for assessing any design relative to another, is:

$$\hat{E}' = \frac{(df_2 + 1)(df_1 + 3)}{(df_2 + 3)(df_1 + 1)}\hat{E} \tag{21.15}$$

where df_1 denotes the degrees of freedom for the experimental error in the base design (completely randomized design, in our case) and df_2 denotes the degrees of freedom for the experimental error in the design whose efficiency is being assessed (randomized complete block design, in our case).

Example

We shall evaluate the efficiency of blocking by age of executives in the risk premium example. Placing the appropriate results from Table 21.3 in efficiency measure (21.13), we obtain:

$$\hat{E} = \frac{4(42.8) + 5(2)(2.99)}{14(2.99)} = 4.8$$

Thus, we would have required almost five times as many replications per treatment with a completely randomized design to achieve the same variance of any estimated contrast as is obtained with blocking by age. Clearly, blocking by age was highly effective here.

If we had used modified efficiency measure (21.14), we would have found:

$$\hat{E}' = \frac{(8+1)(12+3)}{(8+3)(12+1)}(4.8) = 4.5$$

This result does not differ greatly from that obtained by using (21.13).

Comment

The efficiency measure $\hat{E}$ in (21.13) equals 1 if $MSBL = MSBL.TR$; it is greater than 1 if $MSBL > MSBL.TR$; and it is less than 1 if $MSBL < MSBL.TR$. Since the test statistic for block effects in (21.8b) is $F^* = MSBL/MSBL.TR$, it follows that good blocking is achieved when this F^* value exceeds 1 by a considerable margin. ■

Problems

21.1. A student commented in a discussion group: "Random permutations are used to assign treatments to experimental units with a randomized block design just as with a completely randomized design. Hence, there is no basic difference between these two designs." Comment.

21.2. a. What might be some useful blocking variables for an experiment about the effects of different price levels on sales of a product, using stores as experimental units?

b. What might be some useful blocking variables for an experiment about the effects of different flight crew schedules on the morale of crews, using flight crews as experimental units?

c. What might be some useful blocking variables for an experiment about the effects of different drugs on the speed of a response to a stimulus, using laboratory animals as experimental units?

21.3. Five treatments are studied in an experiment with a randomized complete block design using four blocks. Obtain randomized assignments of treatments to experimental units.

21.4. Two treatments and a control are studied in an experiment with a randomized block design. Five blocks are employed, each containing four experimental units. In each block, each treatment is to be assigned to one experimental unit, and the control is to be assigned to two experimental units. Obtain randomized assignments of treatments to experimental units.

*21.5. **Auditor training.** An accounting firm, prior to introducing in the firm widespread training in statistical sampling for auditing, tested three training methods: (1) study at home with programmed training materials, (2) training sessions at local offices conducted by local staff, and (3) training sessions in Chicago conducted by national staff. Thirty auditors were grouped into 10 blocks of three, according to time elapsed since college graduation, and the auditors in each block were randomly assigned to the three training methods. At the end of the training, each auditor was asked to analyze a complex case involving statistical applications; a proficiency measure based on this analysis was obtained for each auditor. The results were

(block 1 consists of auditors graduated most recently, block 10 consists of those graduated most distantly):

Block i	Training Method (j) 1	2	3	Block i	Training Method (j) 1	2	3
1	73	81	92	6	73	75	86
2	76	78	89	7	68	72	88
3	75	76	87	8	64	74	82
4	74	77	90	9	65	73	81
5	76	71	88	10	62	69	78

a. Why do you think the blocking variable "time elapsed since college graduation" was employed?

b. Obtain the residuals for randomized block model (21.1) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings?

c. Plot the responses Y_{ij} by blocks in the format of Figure 21.2. What does this plot suggest about the appropriateness of the no-interaction assumption here?

d. Conduct the Tukey test for additivity of block and treatment effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

*21.6. Refer to **Auditor training** Problem 21.5. Assume that randomized block model (21.1) is appropriate.

a. Obtain the analysis of variance table.

b. Prepare a bar graph of the estimated treatment means. Does it appear that the treatment means differ substantially here?

c. Test whether or not the mean proficiency is the same for the three training methods. Use level of significance $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Make all pairwise comparisons between the training method means; use the Tukey procedure with a 90 percent family confidence coefficient. State your findings.

e. Test whether or not blocking effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

21.7. **Fat in diets.** A researcher studied the effects of three experimental diets with varying fat contents on the total lipid (fat) level in plasma. Total lipid level is a widely used predictor of coronary heart disease. Fifteen male subjects who were within 20 percent of their ideal body weight were grouped into five blocks according to age. Within each block, the three experimental diets were randomly assigned to the three subjects. Data on reduction in lipid level (in grams per liter) after the subjects were on the diet for a fixed period of time follow.

Block i		Fat Content of Diet $j = 1$ Extremely Low	$j = 2$ Fairly Low	$j = 3$ Moderately Low
1	Ages 15–24	.73	.67	.15
2	Ages 25–34	.86	.75	.21
3	Ages 35–44	.94	.81	.26
4	Ages 45–54	1.40	1.32	.75
5	Ages 55–64	1.62	1.41	.78

a. Why do you think that age of subject was used as a blocking variable?

b. Obtain the residuals for randomized block model (21.1) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings?

c. Plot the responses Y_{ij} by blocks in the format of Figure 21.2. What does this plot suggest about the appropriateness of the no-interaction assumption here?

d. Conduct the Tukey test for additivity of block and treatment effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

21.8. Refer to **Fat in diets** Problem 21.7. Assume that randomized block model (21.1) is appropriate.

a. Obtain the analysis of variance table.

b. Prepare a bar-interval graph of the estimated treatment means, using 95 percent confidence intervals. Does it appear that the treatment means differ substantially here?

c. Test whether or not the mean reductions in lipid level differ for the three diets; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Estimate $L_1 = \mu_{\cdot 1} - \mu_{\cdot 2}$ and $L_2 = \mu_{\cdot 2} - \mu_{\cdot 3}$ using the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

e. Test whether or not blocking effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. A standard diet was not used in this experiment as a control. What justification do you think the experimenters might give for not having a control treatment here for comparative purposes?

21.9. **Dental pain.** An anesthesiologist made a comparative study of the effects of acupuncture and codeine on postoperative dental pain in male subjects. The four treatments were: (1) placebo treatment—a sugar capsule and two inactive acupuncture points (A_1B_1), (2) codeine treatment only—a codeine capsule and two inactive acupuncture points (A_2B_1), (3) acupuncture treatment only—a sugar capsule and two active acupuncture points (A_1B_2), and (4) codeine and acupuncture treatment—a codeine capsule and two active acupuncture points (A_2B_2). Thirty-two subjects were grouped into eight blocks of four according to an initial evaluation of their level of pain tolerance. The subjects in each block were then randomly assigned to the four treatments. Pain relief scores were obtained for all subjects two hours after dental treatment. Data were collected on a double-blind basis. The data on pain relief scores follow (the higher the pain relief score, the more effective the treatment).

Block i		Treatment (j, k)			
		A_1B_1	A_2B_1	A_1B_2	A_2B_2
1	(Lowest)	0.0	.5	.6	1.2
2		.3	.6	.7	1.3
...		...	.	...	...
7		1.0	1.8	1.7	2.1
8	(Highest)	1.2	1.7	1.6	2.4

a. Why do you think that pain tolerance of the subjects was used as a blocking variable?

b. Which of the assumptions involved in randomized block model (21.11) are you most concerned with here?

c. Obtain the residuals for randomized block model (21.11) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings?

d. Plot the responses Y_{ijk} by blocks in the format of Figure 21.2, ignoring the factorial structure of the treatments. What does this plot suggest about the appropriateness of the no-interaction assumption here?

e. Conduct the Tukey test for additivity of block and treatment effects, ignoring the factorial structure of the treatments; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

21.10. Refer to **Dental pain** Problem 21.9. Assume that randomized block model (21.11) is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not the two factors interact; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Prepare separate bar-interval graphs for each set of estimated factor level means using 95 percent confidence intervals. Does it appear that substantial main effects are present here?

d. Test separately whether main effects are present for each of the factors; use $\alpha = .01$ for each test. State the alternatives, decision rule, and conclusion for each test. What is the P-value of each test?

e. Estimate:

$$L_1 = \mu_{\cdot 1\cdot} - \mu_{\cdot 2\cdot} = \alpha_1 - \alpha_2$$
$$L_2 = \mu_{\cdot\cdot 1} - \mu_{\cdot\cdot 2} = \beta_1 - \beta_2$$

Use the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

f. Test whether or not blocking effects are present; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

21.11. A social scientist, after learning about generalized randomized block designs, asked: "Why would anyone use a randomized complete block design that requires the assumption that block and treatment effects do not interact when this assumption can be avoided with a generalized randomized block design?" Comment.

*21.12. Refer to the task completion example in Table 21.4.

a. Verify the analysis of variance in Figure 21.6.

b. Estimate the difference in mean effects for the two motivation levels using a 99 percent confidence interval.

*21.13. Refer to **Auditor training** Problem 21.5. The accounting firm repeated the experiment with another group of 30 auditors, but this time grouped them into five blocks of six each. In each block, each treatment was randomly assigned to two auditors. The results were:

Block	Training Method (j)			Block	Training Method (j)		
i	1	2	3	i	1	2	3
1	74	84	91	4	65	73	84
	71	78	95		70	78	87
2	73	75	93	5	64	71	81
	69	83	98		61	74	74
3	75	81	89				
	67	74	86				

Assume that generalized randomized block model (21.10) is appropriate.

a. State the generalized randomized block model for this application.

b. Obtain the analysis of variance table.

c. Test whether or not the mean proficiency scores for the three training methods differ; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Make all pairwise comparisons between the three training methods; use the Tukey procedure with a 95 percent family confidence coefficient. Summarize your findings.

e. Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot of the residuals. State your findings.

f. Test whether or not blocks interact with treatments; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

*21.14. Refer to **Auditor training** Problems 21.5 and 21.6. Assume that $\sigma = 2.5$. What is the power of the test for training method effects in Problem 21.6c if $\mu_{\cdot 1} = 70$, $\mu_{\cdot 2} = 73$, and $\mu_{\cdot 3} = 76$?

21.15. Refer to **Fat in diets** Problems 21.7 and 21.8. Assume that $\sigma = .04$. What is the power of the test for diet effects in Problem 21.8c if $\mu_{\cdot 1} = 1.1$, $\mu_{\cdot 2} = 1.0$, and $\mu_{\cdot 3} = .9$?

*21.16. Refer to **Auditor training** Problem 21.5. Another accounting firm wishes to conduct the same experiment with some of its auditors, using the same design and model. How many blocks would you recommend that this firm employ if it wishes to make all pairwise treatment comparisons with precision ± 1.5 with a 99 percent family confidence coefficient? Assume that a reasonable planning value for the error standard deviation in model (21.1) is $\sigma = 2.5$.

21.17. Refer to **Fat in diets** Problem 21.7. Suppose that the number of blocks to be used in the study, to consist of male subjects of similar ages, has not yet been determined. Assume that a reasonable planning value for the error standard deviation in model (21.1) is $\sigma = .04$.

a. What would be the required number of blocks if it is desired to make all pairwise diet comparisons with precision $\pm .03$ with a 95 percent family confidence coefficient?

b. What would be the required number of blocks if: (1) differences in lipid level reduction means for the three diets are to be detected with probability .95 or more when the range of the treatment means is .12, and (2) the α risk is to be controlled at .01?

*21.18. Refer to **Auditor training** Problems 21.5 and 21.6. According to the estimated efficiency measure (21.13), how effective was the use of the blocking variable as compared to a completely randomized design?

21.19. Refer to **Fat in diets** Problems 21.7 and 21.8. According to the estimated efficiency measure (21.14), how effective was the use of the blocking variable as compared to a completely randomized design?

21.20. Refer to **Dental pain** Problems 21.9 and 21.10. According to the estimated efficiency measure (21.13), how effective was the use of the blocking variable as compared to a completely randomized design?

Exercises

21.21. (Calculus needed.) State the likelihood function for the randomized block fixed effects model (21.1) when $n_b = 3$ and $r = 2$. Find the maximum likelihood estimators of the parameters.

21.22. For randomized block fixed effects model (21.1), derive $E\{MSTR\}$.

21.23. Show that when two treatments are studied in a randomized complete block design, the F^* test statistic (21.7b) for treatment effects is equivalent to the square of the two-sided t test statistic for paired observations based on (A.69).

21.24. Show that the two expressions for X_F^2 on page 900 are equivalent when no ties are present.

Chapter

Analysis of Covariance

Analysis of covariance (ANCOVA) is a technique that combines features of analysis of variance and regression. It can be used for either observational studies or designed experiments. The basic idea is to augment the analysis of variance model containing the factor effects with one or more additional quantitative variables that are related to the response variable. This augmentation is intended to reduce the variance of the error terms in the model, i.e., to make the analysis more precise. We considered covariance models briefly in Chapter 8 on page 329, and noted there that they are linear models containing both qualitative and quantitative predictor variables. Thus, covariance models are just a special type of regression model.

In this chapter, we shall first consider how a covariance model can be more effective than an ordinary ANOVA model. Then we shall discuss how to use a single-factor covariance model for making inferences. We conclude by taking up analysis of covariance models for two-factor studies and some additional considerations for the use of covariance analysis.

22.1 Basic Ideas

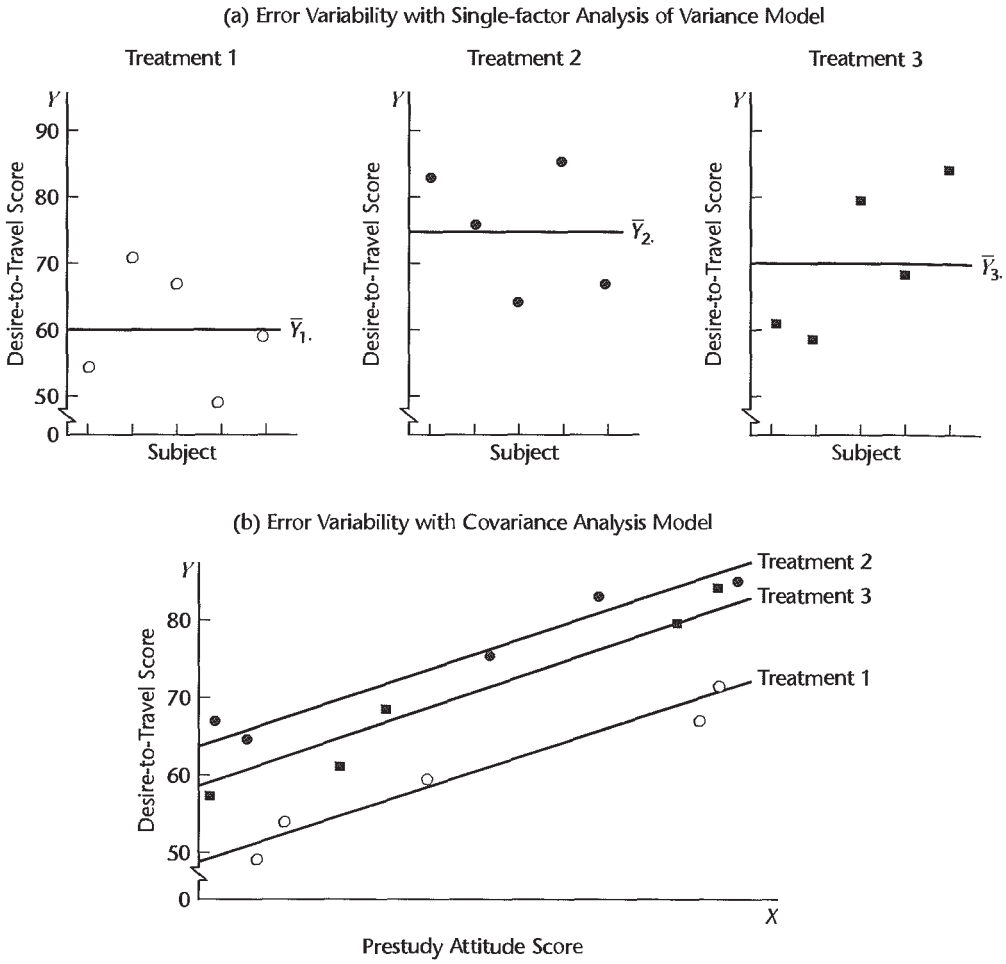
How Covariance Analysis Reduces Error Variability

Covariance analysis may be helpful in reducing large error term variances that sometimes are present in analysis of variance models. Consider a study in which the effects of three different films promoting travel in a state are studied. A subject receives an initial questionnaire to elicit information about the subject's attitudes toward the state. The subject is then shown one of the three five-minute films, and immediately afterwards is questioned about the film, about desire to travel in the state, and so on.

In this type of situation, covariance analysis can be utilized. To see why it might be highly effective, consider Figure 22.1a. Here are plotted the desire-to-travel scores, obtained after each of the three promotional films was shown to a different group of five subjects. Three different symbols are used to distinguish the different treatments. It is evident from Figure 22.1a that the error terms, as shown by the scatter around the estimated treatment means $\overline{Y}_{i\cdot}$, are fairly large, indicating a large error term variance.

Suppose now that we were to utilize also the subjects' initial attitude scores. We plot in Figure 22.1b the desire-to-travel score (obtained after exposure to the film) against the initial attitude score for each of the 15 subjects. Note that the three treatment regression relations

FIGURE 22.1 Illustration of Error Variability Reduction by Covariance Analysis.



happen to be linear (this need not be so). Also note that the scatter around the treatment regression lines is much less than the scatter in Figure 22.1a around the treatment means $\bar{Y}_{i\cdot}$, as a result of the desire-to-travel scores being highly linearly related to the initial attitude scores. The relatively large scatter in Figure 22.1a reflects the large error term variability that would be encountered with an analysis of variance model for this single-factor study. The smaller scatter in Figure 22.1b reflects the smaller error term variability that would be involved in an analysis of covariance model.

Covariance analysis, it is thus seen, utilizes the relationship between the response variable (desire-to-travel score, in our example) and one or more quantitative variables for which observations are available (prestudy attitude score, in our example) in order to reduce the error term variability and make the study a more powerful one for comparing treatment effects.

comitant Variables

In covariance analysis terminology, each quantitative variable added to the ANOVA model is called a *concomitant variable*. We already encountered concomitant variables in Chapter 9, though not by that name. We mentioned in Chapter 9 that *supplemental* or *uncontrolled* variables are sometimes used in regression models for controlled experiments to reduce the variance of the experimental error terms. We also noted in that chapter that *control* variables may be added to the regression model in confirmatory observational studies to reflect the effects of previously identified explanatory variables as the effects of the new, primary explanatory variables on the response variable are being tested. Both the supplemental or uncontrolled variables in a controlled experiment and the control variables in a confirmatory observational study are concomitant variables that are added to the model primarily to reduce the variance of the error terms. Concomitant variables are sometimes also called *covariates*.

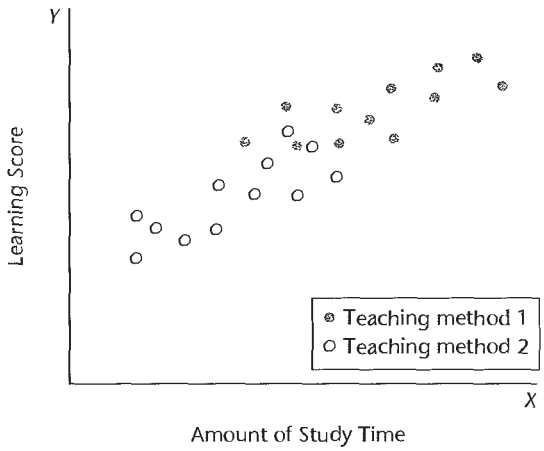
Choice of Concomitant Variables. The choice of concomitant variables is an important one. If such variables have no relation to the response variable, nothing is to be gained by covariance analysis, and one might as well use a simpler analysis of variance model. Concomitant variables frequently used with human subjects include prestudy attitudes, age, socioeconomic status, and aptitude. When retail stores are used as study units, concomitant variables might be last period's sales or number of employees.

Concomitant Variables Unaffected by Treatments. For a clear interpretation of the results, a concomitant variable should be observed before the study; or if observed during the study, it should not be influenced by the treatments in any way. A prestudy attitude score meets this requirement. Also, if a subject's age is ascertained during the study, it would be reasonable in many instances to expect that the information about age provided by the subject will not be affected by the treatment. The reason for this requirement can be seen readily from the following example. A company was conducting a training school for engineers to teach them accounting and budgeting principles. Two teaching methods were used, and engineers were assigned at random to one of the two. At the end of the program, a score was obtained for each engineer reflecting the amount of learning. The analyst decided to use as a concomitant variable in covariance analysis the amount of time devoted to study (which the engineers were required to record). After conducting the analysis of covariance, the analyst found that training method had virtually no effect. The analyst was baffled by this finding until it was pointed out that the amount of study time probably was also affected by the treatments, and analysis indeed confirmed this. One of the training methods involved computer-assisted learning which appealed to the engineers so that they spent more time studying and also learned more. In other words, both the learning score and the amount of study time were influenced by the treatment in this case. As a result of the high correlation between the amount of study time and the learning score, the marginal treatment effect of the teaching methods on amount of learning was small and the test for treatment effects showed no significant difference between the two teaching methods.

Whenever a concomitant variable is affected by the treatments, covariance analysis will fail to show some (or much) of the effects that the treatments had on the response variable, so that an uncritical analysis may be badly misleading.

A symbolic scatter plot can provide evidence as to whether the concomitant variable is affected by the treatments. Figure 22.2 shows a scatter plot of learning score and amount of

FIGURE 22.2 Illustration of Treatments Affecting the Concomitant Variable—Engineer Training Example.



study time for the engineer training example. Treatment 1 is the one using computer-assisted learning. Note that most persons with this treatment devoted large amounts of time to study. On the other hand, persons receiving treatment 2 tended to devote smaller amounts of time to study. As a result, the observations for the two treatments tend to be concentrated over different intervals on the X scale.

Contrast this situation with the one seen in Figure 22.1b for the study on promotional films. Figure 22.1b illustrates how the concomitant variable observations should be scattered in a randomized experiment if the treatments have no effect on the concomitant variable. Here, the distribution of subjects along the X scale by prestudy attitude scores is roughly similar for all treatments, subject only to chance variation.

Comment

Covariance analysis is concerned with quantitative concomitant variables. When qualitative concomitant variables need to be added (e.g., gender, geographic region), the model remains an analysis of variance model where some of the factors are of primary interest and the others represent concomitant variables that are included for the purpose of error variance reduction. ■

22.2 Single-Factor Covariance Model

The covariance models to be presented in this chapter are applicable to observational studies and to experimental studies based on a completely randomized design. In the earlier engineer training example, the 24 engineers participating in the study were randomly assigned to the two teaching methods, with 12 engineers assigned to each teaching method. Thus, this experimental study was based on a completely randomized design.

The covariance models to be taken up in this chapter are also applicable to observational studies, such as an investigation of the salary increases of a company's employees in the accounting department by gender, where age is utilized as a concomitant variable.

otation

We shall employ the notation for single-factor analysis of variance. The number of cases for the ith factor level is denoted by n_i, the total number of cases by $n_T = \sum n_i$, and the jth observation on the response variable for the ith factor level is denoted by Y_{ij}. We shall initially consider a single-factor covariance model with only one concomitant variable. Later we shall take up models with more than one concomitant variable. We shall denote the value of the concomitant variable associated with the jth case for the ith factor level by X_{ij}.

Development of Covariance Model

The single-factor ANOVA model in terms of fixed factor effects was given in (16.62):

$$Y_{ij} = \mu_. + \tau_i + \varepsilon_{ij} \tag{22.1}$$

The covariance model starts with this ANOVA model and adds another term (or several), reflecting the relationship between the response variable and the concomitant variable. Usually, a linear relation is utilized as a first approximation:

$$Y_{ij} = \mu_. + \tau_i + \gamma X_{ij} + \varepsilon_{ij} \tag{22.2}$$

Here γ is a regression coefficient for the relation between Y and X. The constant $\mu_.$ now is no longer an overall mean. We can, however, make this constant an overall mean, and incidentally simplify some computations, if we center the concomitant variable around the overall mean $\bar{X}_{..}$. The resulting model is the usual covariance model for a single-factor study with fixed factor levels:

$$Y_{ij} = \mu_. + \tau_i + \gamma(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij} \tag{22.3}$$

where:

$\mu_.$ is an overall mean
τ_i are the fixed treatment effects subject to the restriction $\sum \tau_i = 0$
γ is a regression coefficient for the relation between Y and X
X_{ij} are constants
ε_{ij} are independent $N(0, \sigma^2)$
$i = 1, \ldots, r;\ j = 1, \ldots, n_i$

Covariance model (22.3) corresponds to ANOVA model (22.1) except for the term $\gamma(X_{ij} - \bar{X}_{..})$, which is added to reflect the relationship between Y and X. Note that the concomitant observations X_{ij} are assumed to be constants. Since ε_{ij} is the only random variable on the right side of (22.3), it follows at once that:

$$E\{Y_{ij}\} = \mu_. + \tau_i + \gamma(X_{ij} - \bar{X}_{..}) \tag{22.4a}$$

$$\sigma^2\{Y_{ij}\} = \sigma^2 \tag{22.4b}$$

In view of the independence of the ε_{ij}, the Y_{ij} are also independent. Hence, an alternative statement of covariance model (22.3) is:

$$Y_{ij} \text{ are independent } N(\mu_{ij}, \sigma^2) \tag{22.5}$$

where:

$$\mu_{ij} = \mu_{\cdot} + \tau_i + \gamma(X_{ij} - \bar{X}_{\cdot\cdot})$$
$$\sum \tau_i = 0$$

Properties of Covariance Model

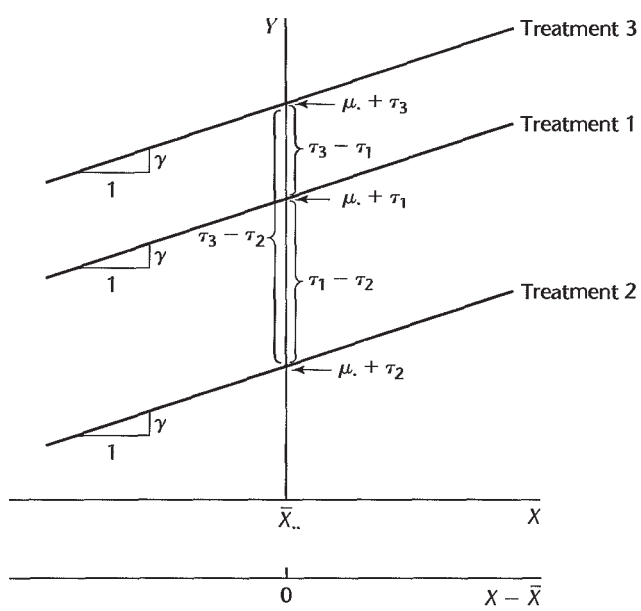
Some of the properties of covariance model (22.3) are identical to those of ANOVA model (22.1). For instance, the error terms ε_{ij} are independent and have constant variance. There are also some new properties, and we discuss these now.

Comparisons of Treatment Effects. With the analysis of variance model, all observations for the ith treatment have the same mean response; i.e., $E\{Y_{ij}\} = \mu_i$ for all j. This is not so with the covariance model, since the mean response $E\{Y_{ij}\}$ here depends not only on the treatment but also on the value of the concomitant variable X_{ij} for the study unit. Thus, the expected response for the ith treatment with covariance model (22.3) is given by a regression line:

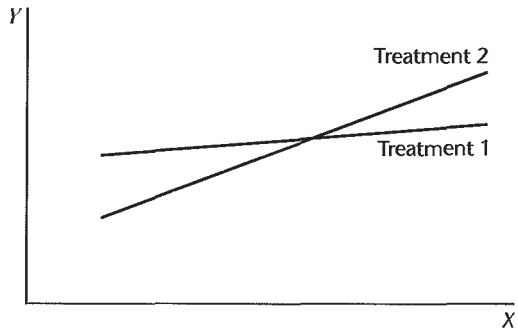
$$\mu_{ij} = \mu_{\cdot} + \tau_i + \gamma(X_{ij} - \bar{X}_{\cdot\cdot}) \tag{22.6}$$

This regression line indicates, for any value of X, the mean response with treatment i. Figure 22.3 illustrates for a study with three treatments how these treatment regression lines might appear. Note that $\mu_{\cdot} + \tau_i$ is the ordinate of the line for the ith treatment when

FIGURE 22.3 Example of Treatment Regression Lines with Covariance Model (22.3).



GURE 22.4
ple of
onparallel
tment
egression
ines.



$X - \bar{X}_{..} = 0$, that is, when $X = \bar{X}_{..}$, and that γ is the slope of each line. Since all treatment regression lines have the same slope, they are parallel.

While we no longer can speak of *the* mean response with the ith treatment since it varies with X, we can still measure the effect of any treatment compared with any other by a single number. In Figure 22.3, for instance, treatment 1 leads to a higher mean response than treatment 2 by an amount that is the same no matter what is the value of X. The difference between the two mean responses is the same for all values of X because the slopes of the regression lines are equal. Hence, we can measure the difference at any convenient X, say, at $X = \bar{X}_{..}$:

$$\mu_{.} + \tau_1 - (\mu_{.} + \tau_2) = \tau_1 - \tau_2 \tag{22.7}$$

Thus, $\tau_1 - \tau_2$ measures how much higher the mean response is with treatment 1 than with treatment 2 for any value of X. We can compare any other two treatments similarly. It follows directly from this discussion that when all treatments have the same mean responses for any X (i.e., the treatments have no differential effects), the treatment regression lines must be identical; and hence, $\tau_1 - \tau_2 = 0$, $\tau_1 - \tau_3 = 0$, etc. Indeed, all τ_i equal zero in that case.

Constancy of Slopes. The assumption in covariance model (22.3) that all treatment regression lines have the same slope is a crucial one. Without it, the difference between the effects of two treatments cannot be summarized by a single number based on the main effects, such as $\tau_2 - \tau_1$. Figure 22.4 illustrates the case of nonparallel slopes for two treatments. Here, treatment 1 leads to higher mean responses than treatment 2 for smaller values of X, and the reverse holds for larger values of X. When the treatments interact with the concomitant variable X, resulting in nonparallel slopes, covariance analysis is not appropriate. Instead, separate treatment regression lines need to be estimated and then compared.

Generalizations of Covariance Model

Covariance model (22.3) for single-factor studies can be generalized in several respects. We mention briefly three ways in which this model can be generalized.

Nonconstant *X*s. Covariance model (22.3) assumes that the observations X_{ij} on the concomitant variable are constants. At times, it might be more reasonable to consider the concomitant observations as random variables. In that case, if covariance model (22.3) can be interpreted as a conditional one, applying for any X values that might be observed, the covariance analysis to be presented is still appropriate.

Nonlinearity of Relation. The linear relation between Y and X assumed in covariance model (22.3) is not essential to covariance analysis. Any other relation could be used. For instance, the model for a quadratic relation is as follows:

$$Y_{ij} = \mu_{\cdot} + \tau_i + \gamma_1(X_{ij} - \bar{X}_{\cdot\cdot}) + \gamma_2(X_{ij} - \bar{X}_{\cdot\cdot})^2 + \varepsilon_{ij} \tag{22.8}$$

Linearity of the relation leads to simpler analysis and is often a sufficiently good approximation to provide meaningful results. If a linear relation is not a good approximation, however, a more adequate description of the relation should be utilized in the covariance model. Covariance analysis does require, however, that the treatment response functions be parallel; in other words, there must not be any interaction effects between the treatment and concomitant variables.

Several Concomitant Variables. Covariance model (22.3) uses a single concomitant variable. This is often sufficient to reduce the error variability substantially. However, the model can be extended in a straightforward fashion to include two or more concomitant variables. The single-factor covariance model for two concomitant variables, X_1 and X_2, to the first order is as follows:

$$Y_{ij} = \mu_{\cdot} + \tau_i + \gamma_1(X_{ij1} - \bar{X}_{\cdot\cdot 1}) + \gamma_2(X_{ij2} - \bar{X}_{\cdot\cdot 2}) + \varepsilon_{ij} \tag{22.9}$$

Regression Formulation of Covariance Model

An easy way to estimate the parameters of covariance model (22.3) and make inferences is through the regression approach. Computational formulas for manual calculation were developed before the advent of computers, making use of the special structure of the **X** matrix for covariance models. Today, however, covariance calculations can be carried out readily by means of standard regression packages.

As for the regression formulation of analysis of variance models, we shall employ $r-1$ indicator variables taking on the values 1, -1, or 0 to represent the r treatments in a covariance analysis model:

$$I_1 = \begin{cases} 1 & \text{if case from treatment 1} \\ -1 & \text{if case from treatment } r \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots \tag{22.10}$$

$$I_{r-1} = \begin{cases} 1 & \text{if case from treatment } r-1 \\ -1 & \text{if case from treatment } r \\ 0 & \text{otherwise} \end{cases}$$

Note that we now denote the indicator variables by the symbol I to clearly distinguish the treatment effects from the concomitant variable X.

In expressing covariance model (22.3) in regression form, we shall, as in the regression chapters, denote the centered observations $X_{ij} - \bar{X}_{\cdot\cdot}$ by x_{ij}. Covariance model (22.3) can then be expressed as follows:

$$Y_{ij} = \mu_{\cdot} + \tau_1 I_{ij1} + \cdots + \tau_{r-1} I_{ij,r-1} + \gamma x_{ij} + \varepsilon_{ij} \tag{22.11}$$

where:

$$x_{ij} = X_{ij} - \bar{X}_{\cdot\cdot}$$

Here, I_{ij1} is the value of indicator variable I_1 for the jth case from treatment i, and similarly for the other indicator variables. Note that the treatment effects $\tau_1, \ldots, \tau_{r-1}$ are the regression coefficients for the indicator variables.

Now that we have formulated covariance model (22.3) as a regression model, our discussion of regression analysis in previous chapters applies. We therefore consider only briefly how to examine the appropriateness of the covariance model and how to make relevant inferences before turning to an example to illustrate the procedures.

propriateness of Covariance Model

Some of the key issues concerning the appropriateness of covariance model (22.3) and the equivalent regression model (22.11) deal with:

1. Normality of error terms.
2. Equality of error variances for different treatments.
3. Equality of slopes of the different treatment regression lines.
4. Linearity of regression relation with concomitant variable.
5. Uncorrelatedness of error terms.

The third issue, concerning the equality of the slopes of the different treatment regression lines, is particularly important in evaluating the appropriateness of covariance model (22.3). The test in Section 8.7 to compare several regression lines is applicable for determining whether the condition of equal slopes in the covariance model is met. We shall illustrate this test in the example in Section 22.3.

nferences of Interest

The key statistical inferences of interest in covariance analysis are the same as with analysis of variance models, namely, whether the treatments have any effects, and if so what these effects are. Testing for fixed treatment effects involves the same alternatives as for analysis of variance models:

$$\begin{aligned} &H_0\colon \tau_1 = \tau_2 = \cdots = \tau_r = 0 \\ &H_a\colon \text{not all } \tau_i \text{ equal zero} \end{aligned} \tag{22.12}$$

As we can see by referring to the equivalent regression model (22.11), this test involves testing whether several regression coefficients equal zero. The appropriate test statistic therefore is (7.27).

If the treatment effects are found to differ, the next step usually is to investigate the nature of these effects. Pairwise comparisons of treatment effects $\tau_i - \tau_{i'}$ (the vertical distance between the two treatment regression lines) may be of interest, or more general contrasts of the τ_i may be relevant. In either case, linear combinations of the regression coefficients $\tau_1, \ldots, \tau_{r-1}$ are to be estimated.

Occasionally, the nature of the regression relationship between Y and X is of interest, but usually the concomitant variable X is only employed in ANCOVA models to help reduce the error variability.

Comment

In covariance analysis there is usually no concern with whether the regression coefficient γ is zero, that is, whether there is indeed a regression relation between Y and X. If there is no relation, no bias

results in the covariance analysis. The error mean square would simply be the same as for the analysis of variance model (allowing for sampling variation), and one degree of freedom would be lost for the error mean square. ■

22.3 Example of Single-Factor Covariance Analysis

A company studied the effects of three different types of promotions on sales of its crackers:

Treatment 1—Sampling of product by customers in store and regular shelf space
Treatment 2—Additional shelf space in regular location
Treatment 3—Special display shelves at ends of aisle in addition to regular shelf space

Fifteen stores were selected for the study, and a completely randomized experimental design was utilized. Each store was randomly assigned one of the promotion types, with five stores assigned to each type of promotion. Other relevant conditions under the control of the company, such as price and advertising, were kept the same for all stores in the study. Data on the number of cases of the product sold during the promotional period, denoted by Y, are presented in Table 22.1, as are also data on the sales of the product in the preceding period, denoted by X. Sales in the preceding period are to be used as the concomitant variable.

Development of Model

Figure 22.5 presents the data of Table 22.1 in the form of a symbolic scatter plot. Linear regression and parallel slopes for the treatment regression lines appear to be reasonable. Therefore, the following regression model was tentatively selected:

$$Y_{ij} = \mu_{\cdot} + \tau_1 I_{ij1} + \tau_2 I_{ij2} + \gamma x_{ij} + \varepsilon_{ij} \qquad \text{Full model} \qquad \textbf{(22.13)}$$

where:

$$I_1 = \begin{cases} 1 & \text{if store received treatment 1} \\ -1 & \text{if store received treatment 3} \\ 0 & \text{otherwise} \end{cases}$$

$$I_2 = \begin{cases} 1 & \text{if store received treatment 2} \\ -1 & \text{if store received treatment 3} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{ij} = X_{ij} - \overline{X}_{\cdot\cdot}$$

TABLE 22.1 Data—Cracker Promotion Example (number of cases sold).

Treatment	Store (j)									
	1		2		3		4		5	
i	Y_{i1}	X_{i1}	Y_{i2}	X_{i2}	Y_{i3}	X_{i3}	Y_{i4}	X_{i4}	Y_{i5}	X_{i5}
1	38	21	39	26	36	22	45	28	33	19
2	43	34	38	26	38	29	27	18	34	25
3	24	23	32	29	31	30	21	16	28	29

IGURE 22.5
bolic
catter Plot of
cker
ales—
Cracker
motion
xample.

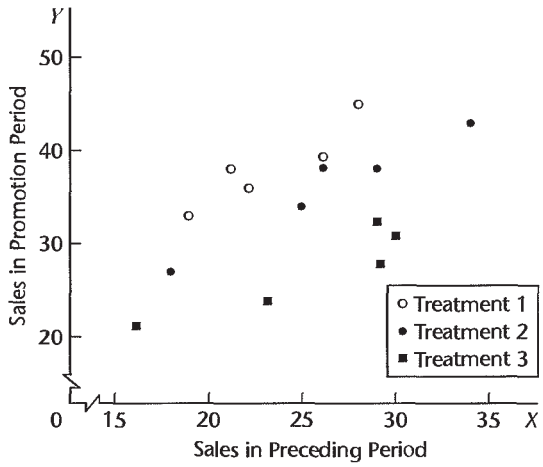


TABLE 22.2
Regression Variables for Single-Factor Covariance Analysis—Cracker Promotion Example.

i	j	(1) Y	(2) X	(3) x	(4) I_1	(5) I_2	(6) I_1x	(7) I_2x
1	1	38	21	−4	1	0	−4	0
1	2	39	26	1	1	0	1	0
...		...	...	...	...	...	...	...
2	1	43	34	9	0	1	0	9
2	2	38	26	1	0	1	0	1
...		...	...	...	...	...	...	...
3	4	21	16	−9	−1	−1	9	9
3	5	28	29	4	−1	−1	−4	−4
			$\bar{X}_{..} = 25$					

Table 22.2 repeats a portion of the data on the responses Y and the concomitant variable X in columns 1 and 2. The centered concomitant variable x is presented in column 3 and the indicator variables for the treatments in columns 4 and 5. Note that the centering of the concomitant variable is around the overall mean $\bar{X}_{..} = 25$. Regressing Y in column 1 of Table 22.2 on x, I_1, and I_2 in columns 3–5 by a computer package led to the results summarized in Table 22.3.

Various residual plots were obtained to examine the appropriateness of regression model (22.13). Figure 22.6 contains two of these. Figure 22.6a contains aligned residual dot plots for the three treatments. These do not suggest any major differences in the variances of the error terms. Figure 22.6b contains a normal probability plot of the residuals, which shows some modest departure from linearity. However, the coefficient of correlation between the ordered residuals and their expected values under normality is .958, for which Table B.6 does not suggest any significant departure from normality. The analyst also conducted a test to confirm the equality of the slopes of the three treatment regression lines. This test will be described shortly. On the basis of these analyses, the analyst concluded that regression model (22.13) is appropriate here.

TABLE 22.3 Computer Output for Covariance Model (22.13)—Cracker Promotion Example.

(a) Regression Coefficients

$\hat{\mu}_{\cdot} = 33.800 \qquad \hat{\tau}_2 = .942$

$\hat{\tau}_1 = 6.017 \qquad \hat{\gamma} = .899$

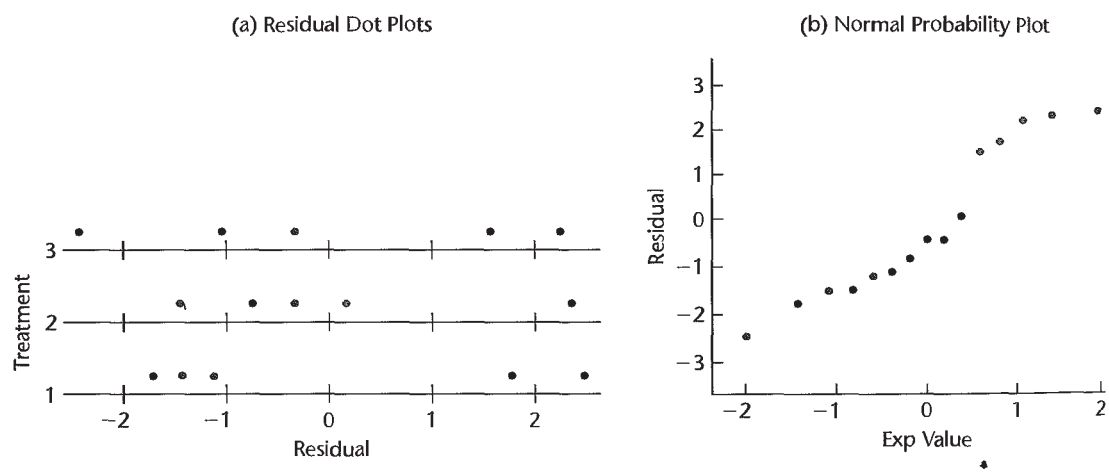
(b) Analysis of Variance

Source of Variation	*SS*	*df*	*MS*
Regression	$SSR = 607.829$	3	$MSR = 202.610$
Error	$SSE = 38.571$	11	$MSE = 3.506$
Total	$SSTO = 646.400$	14	

(c) Estimated Variance-Covariance Matrix of Regression Coefficients

$$\begin{array}{c} \\ \hat{\mu}_{\cdot} \\ \hat{\tau}_1 \\ \hat{\tau}_2 \\ \hat{\gamma} \end{array} \begin{array}{c} \begin{array}{cccc} \hat{\mu}_{\cdot} & \hat{\tau}_1 & \hat{\tau}_2 & \hat{\gamma} \end{array} \\ \begin{bmatrix} .2338 & & & \\ 0 & .5016 & & \\ 0 & -.2603 & .4882 & \\ 0 & .0189 & -.0147 & .0105 \end{bmatrix} \end{array}$$

FIGURE 22.6 Diagnostic Residual Plots—Cracker Promotion Example.



Test for Treatment Effects

To test whether or not the three cracker promotions differ in effectiveness, we can ei follow the general linear test approach of fitting full and reduced models and using statistic (2.70) or use extra sums of squares and test statistic (7.27). In either case,

TABLE 22.4 Regression ANOVA Results for Reduced Model (22.15)—Cracker Promotion Example.

Source of Variation	*SS*	*df*
Regression	$SSR = 190.678$	1
Error	$SSE = 455.722$	13
Total	$SSTO = 646.400$	14

alternatives are:

$$H_0: \tau_1 = \tau_2 = 0$$
$$H_a: \text{not both } \tau_1 \text{ and } \tau_2 \text{ equal zero} \tag{22.14}$$

Note that $\tau_3 = -\tau_1 - \tau_2$ must equal zero when $\tau_1 = \tau_2 = 0$.

We shall conduct the test by means of the general linear test approach. First, we develop the reduced model under H_0:

$$Y_{ij} = \mu. + \gamma x_{ij} + \varepsilon_{ij} \qquad \text{Reduced model} \tag{22.15}$$

Model (22.15) is just a simple linear regression model where none of the parameters vary for the different treatments. When regressing Y in column 1 of Table 22.2 on x in column 3, we obtain the analysis of variance results in Table 22.4.

We see from Table 22.4 that $SSE(R) = 455.722$ and from Table 22.3b that $SSE(F) = 38.571$. Hence, test statistic (2.70) here is:

$$F^* = \frac{SSE(R) - SSE(F)}{(n_T - 2) - [n_T - (r+1)]} \div \frac{SSE(F)}{n_T - (r+1)}$$
$$= \frac{455.722 - 38.571}{13 - 11} \div \frac{38.571}{11} = 59.5$$

The level of significance is to be controlled at $\alpha = .05$; hence, we need to obtain $F(.95; 2, 11) = 3.98$. The decision rule therefore is:

$$\text{If } F^* \leq 3.98, \text{ conclude } H_0$$
$$\text{If } F^* > 3.98, \text{ conclude } H_a$$

Since $F^* = 59.5 > 3.98$, we conclude H_a, that the three cracker promotions differ in sales effectiveness. The P-value of the test is 0+.

Comment

Occasionally, a test whether or not $\gamma = 0$ is of interest. This is simply the ordinary test whether or not a single regression coefficient equals zero. It can be conducted by means of the t^* test statistic (7.25) or by means of the F^* test statistic (7.24). ■

Estimation of Treatment Effects

Since treatment effects were found to be present in the cracker promotion study, the analyst next wished to investigate the nature of these effects. We noted earlier that a comparison of two treatments involves $\tau_i - \tau_{i'}$, the vertical distance between the two treatment regression lines. Using the fact that $\tau_3 = -\tau_1 - \tau_2$ and (A.30b) for the variance of a linear combination of two random variables, we see that the estimators of all pairwise comparisons and their variances are as follows:

Comparison	Estimator	Variance
$\tau_1 - \tau_2$	$\hat{\tau}_1 - \hat{\tau}_2$	$\sigma^2\{\hat{\tau}_1\} + \sigma^2\{\hat{\tau}_2\} - 2\sigma\{\hat{\tau}_1, \hat{\tau}_2\}$
$\tau_1 - \tau_3 = 2\tau_1 + \tau_2$	$2\hat{\tau}_1 + \hat{\tau}_2$	$4\sigma^2\{\hat{\tau}_1\} + \sigma^2\{\hat{\tau}_2\} + 4\sigma\{\hat{\tau}_1, \hat{\tau}_2\}$
$\tau_2 - \tau_3 = \tau_1 + 2\tau_2$	$\hat{\tau}_1 + 2\hat{\tau}_2$	$\sigma^2\{\hat{\tau}_1\} + 4\sigma^2\{\hat{\tau}_2\} + 4\sigma\{\hat{\tau}_1, \hat{\tau}_2\}$

(22.16)

Table 22.3a furnishes the needed estimated regression coefficients, and Table 22.3c provides their estimated variances and covariances. We obtain from there:

Comparison	Estimate	Variance
$\tau_1 - \tau_2$	$6.017 - .942$ $= 5.075$	$.5016 + .4882 - 2(-.2603)$ $= 1.5104$
$\tau_1 - \tau_3$	$2(6.017) + .942$ $= 12.976$	$4(.5016) + .4882 + 4(-.2603)$ $= 1.4534$
$\tau_2 - \tau_3$	$6.017 + 2(.942)$ $= 7.901$	$.5016 + 4(.4882) + 4(-2.2603)$ $= 1.4132$

(22.16a)

When a single interval estimate is to be constructed, the t distribution with $n_T - r - 1$ degrees of freedom is used. (The degrees of freedom are those associated with *MSE* in the full covariance model.) Usually, however, a family of interval estimates is desired. In that case, the Scheffé multiple comparison procedure may be employed with the S multiple defined by:

$$S^2 = (r-1)F(1-\alpha; r-1, n_T - r - 1) \tag{22.17}$$

or the Bonferroni method may be employed with the B multiple:

$$B = t(1 - \alpha/2g; n_T - r - 1) \tag{22.18}$$

where g is the number of statements in the family. The Tukey method is not appropriate for covariance analysis.

In the case at hand, the analyst wished to obtain all pairwise comparisons with a 95 percent family confidence coefficient. The analyst used the Scheffé procedure in anticipation that

some additional estimates of contrasts might be desired. We require therefore:

$$S^2 = (3-1)F(.95; 2, 11) = 2(3.98) = 7.96 \qquad S = 2.82$$

Using the results in (22.16a), the confidence intervals for all pairwise treatment comparisons with a 95 percent family confidence coefficient then are:

$$\begin{aligned}
1.61 &= 5.075 - 2.82\sqrt{1.5104} \le \tau_1 - \tau_2 \le 5.075 + 2.82\sqrt{1.5104} = 8.54 \\
9.58 &= 12.976 - 2.82\sqrt{1.4534} \le \tau_1 - \tau_3 \le 12.976 + 2.82\sqrt{1.4534} = 16.38 \\
4.55 &= 7.901 - 2.82\sqrt{1.4132} \le \tau_2 - \tau_3 \le 7.901 + 2.82\sqrt{1.4132} = 11.25
\end{aligned}$$

These results indicate clearly that sampling in the store (treatment 1) is significantly better for stimulating cracker sales than either of the two shelf promotions, and that increasing the regular shelf space (treatment 2) is superior to additional displays at the end of the aisle (treatment 3).

Comments

1. Occasionally, more general contrasts among treatment effects than pairwise comparisons are desired. No new problems arise either in the use of the t distribution for a single contrast or in the use of the Scheffé or Bonferroni procedures for multiple comparisons. For instance, if the analyst desired in the cracker promotion example to compare the treatment effect for sampling in the store (treatment 1) with the two treatments involving shelf displays (treatments 2 and 3), the following contrast would be of interest:

$$L = \tau_1 - \frac{\tau_2 + \tau_3}{2} \tag{22.19}$$

The appropriate estimator is:

$$\hat{L} = \hat{\tau}_1 - \frac{\hat{\tau}_2 + (-\hat{\tau}_1 - \hat{\tau}_2)}{2} = \frac{3}{2}\hat{\tau}_1 \tag{22.20}$$

The variance of this estimator is by (A.16b):

$$\sigma^2\{\hat{L}\} = \frac{9}{4}\sigma^2\{\hat{\tau}_1\} \tag{22.21}$$

2. Sometimes there is interest in estimating the mean response with the ith treatment for a "typical" value of X. Frequently $X = \bar{X}_{..}$ is considered to be a "typical" value. We know from Figure 22.3 that at $X = \bar{X}_{..}$, the mean response for the ith treatment is the intercept of the treatment regression line, $\mu_{.} + \tau_i$. An estimator of $\mu_{.} + \tau_i$ can be readily developed. For the cracker promotion example, we obtain the following estimators and their variances:

Mean Response at $X = \bar{X}_{..}$	Estimator	Variance
$\mu_{.} + \tau_1$	$\hat{\mu}_{.} + \hat{\tau}_1$	$\sigma^2\{\hat{\mu}_{.}\} + \sigma^2\{\hat{\tau}_1\} + 2\sigma\{\hat{\mu}_{.}, \hat{\tau}_1\}$
$\mu_{.} + \tau_2$	$\hat{\mu}_{.} + \hat{\tau}_2$	$\sigma^2\{\hat{\mu}_{.}\} + \sigma^2\{\hat{\tau}_2\} + 2\sigma\{\hat{\mu}_{.}, \hat{\tau}_2\}$
$\mu_{.} + \tau_3$	$\hat{\mu}_{.} - \hat{\tau}_1 - \hat{\tau}_2$	$\sigma^2\{\hat{\mu}_{.}\} + \sigma^2\{\hat{\tau}_1\} + \sigma^2\{\hat{\tau}_2\} - 2\sigma\{\hat{\mu}_{.}, \hat{\tau}_1\} - 2\sigma\{\hat{\mu}_{.}, \hat{\tau}_2\} + 2\sigma\{\hat{\tau}_1, \hat{\tau}_2\}$

(22.22)

Use of the results in Table 22.3 leads to the following estimates:

Treatment	Estimated Mean Response at $\bar{X}_{..}$	Estimated Variance
1	$33.800 + 6.017 = 39.817$	$.2338 + .5016 + 2(0) = .7354$
2	$33.800 + .942 = 34.742$	$.2338 + .4882 + 2(0) = .7220$
3	$33.800 - 6.017 - .942$ $= 26.841$	$.2338 + .5016 + .4882 - 2(0) - 2(0)$ $+ 2(-.2603) = .7030$

The estimated mean response for treatment i at $X = \bar{X}_{..}$ is often called the *adjusted estimated treatment mean*. It is said to be "adjusted" because it takes into account the effect of the concomitant variable. A comparison of the adjusted treatment means leads, of course, to the same pairwise comparisons of treatment effects as before; for instance, $39.817 - 34.742 = 5.075 = \hat{\tau}_1 - \hat{\tau}_2$. ■

Test for Parallel Slopes

An important assumption in covariance analysis is that all treatment regression lines have the same slope γ. The analyst who conducted the cracker promotion study, indeed, tested this assumption before proceeding with the analysis discussed earlier. We know from Chapter 8 that regression model (22.13) can be generalized to allow for different slopes for the treatments by introducing cross-product interaction terms. Specifically, interaction variables I_1x and I_2x will be required here. We shall denote the corresponding regression coefficients by β_1 and β_2. Thus, the generalized model is:

$$Y_{ij} = \mu_{.} + \tau_1 I_{ij1} + \tau_2 I_{ij2} + \gamma x_{ij} + \beta_1 I_{ij1} x_{ij} + \beta_2 I_{ij2} x_{ij} + \varepsilon_{ij} \qquad \text{Generalized model} \tag{22.23}$$

Table 22.2 contains in columns 6 and 7 the interaction variables for this model for the cracker promotion example. Regressing the response variable Y in column 1 of Table 22.2 on x, I_1, I_2, I_1x, I_2x in columns 3–7 by means of a computer multiple regression package yielded the ANOVA results in Table 22.5. The error sum of squares SSE obtained by fitting generalized model (22.23) is the equivalent of fitting separate regression lines for each treatment and summing these error sums of squares.

TABLE 22.5 Regression ANOVA Results for Generalized Model (22.23)—Cracker Promotion Example.

Source of Variation	*SS*	*df*
Regression	$SSR = 614.879$	5
Error	$SSE = 31.521$	9
Total	$SSTO = 646.400$	14

The test for parallel slopes is equivalent to testing for no interactions in generalized model (22.23):

$$H_0: \beta_1 = \beta_2 = 0$$
$$H_a: \text{not both } \beta_1 \text{ and } \beta_2 \text{ equal zero} \tag{22.24}$$

We need to recognize that generalized model (22.23) now is the "full" model and covariance model (22.13) is the "reduced" model. Hence, we have from Tables 22.3b and 22.5:

$$SSE(F) = 31.521 \qquad SSE(R) = 38.571$$

Thus, test statistic (2.70) becomes here:

$$F^* = \frac{38.571 - 31.521}{11 - 9} \div \frac{31.521}{9} = 1.01$$

For level of significance $\alpha = .05$, we require $F(.95; 2, 9) = 4.26$. Since $F^* = 1.01 \leq 4.26$, we conclude H_0, that the three treatment regression lines have the same slope. The P-value of the test is .40. Hence, the requirement of equal treatment slopes in analysis of covariance model (22.13) is met in the cracker promotion example.

Comments

1. An indication of the effectiveness of the analysis of covariance in reducing error variability can be obtained by comparing *MSE* for covariance analysis with *MSE* for regular analysis of variance. For the cracker promotion example, we know from Table 22.3 that *MSE* for the covariance analysis is 3.51. It can be shown that the error mean square for regular analysis of variance would have been 26.63. Hence, in this case, covariance analysis was able to reduce the residual variance by about 87 percent, a substantial reduction.

2. Covariance analysis and analysis of variance need not lead to the same conclusions about the treatment effects. For instance, analysis of variance might not indicate any treatment effects, whereas covariance analysis with a smaller error variance could show significant treatment effects. Ordinarily, of course, one should decide in advance which of the two analyses is to be used. ■

22.4 Two-Factor Covariance Analysis

We have until now considered covariance analysis for single-factor studies with r treatments. Covariance analysis can also be employed with two-factor and multifactor studies. We illustrate now the use of covariance analysis for two-factor studies with one concomitant variable. For notational simplicity, we consider the case where the treatment sample size is the same for all treatments. However, the regression approach to covariance analysis is general and applies directly when the study is unbalanced, with unequal treatment sample sizes.

Covariance Model for Two-Factor Studies

The fixed effects ANOVA model for a two-factor balanced study was given in (19.23):

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n \tag{22.25}$$

where α_i is the main effect of factor A at the ith level, β_j is the main effect of factor B at the jth level, and $(\alpha\beta)_{ij}$ is the interaction effect when factor A is at the ith level and factor B is at the jth level. The covariance model for a two-factor study with a single concomitant variable, assuming the relation between Y and the concomitant variable X is linear, is:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma(X_{ijk} - \bar{X}_{...}) + \varepsilon_{ijk}$$
$$i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n \tag{22.26}$$

Regression Approach

We illustrate the regression approach to covariance analysis for a balanced two-factor study with one concomitant variable when both factors A and B are at two levels, i.e., when $a = b = 2$. The regression model counterpart to covariance model (22.26) then is:

$$Y_{ijk} = \mu_{..} + \alpha_1 I_{ijk1} + \beta_1 I_{ijk2} + (\alpha\beta)_{11} I_{ijk1} I_{ijk2} + \gamma x_{ijk} + \varepsilon_{ijk} \tag{22.27}$$

where:

$$I_1 = \begin{cases} 1 & \text{if case from level 1 for factor } A \\ -1 & \text{if case from level 2 for factor } A \end{cases}$$

$$I_2 = \begin{cases} 1 & \text{if case from level 1 for factor } B \\ -1 & \text{if case from level 2 for factor } B \end{cases}$$

$$x_{ijk} = X_{ijk} - \bar{X}_{...}$$

Note that the regression coefficients in (22.27) are the analysis of variance factor effects α_1, β_1, and $(\alpha\beta)_{11}$ and the concomitant variable coefficient γ.

Testing for factor A main effects requires that $\alpha_1 = 0$ in the reduced model. Correspondingly, $\beta_1 = 0$ is required in the reduced model when testing for factor B main effects, and $(\alpha\beta)_{11} = 0$ is required in the reduced model when testing for AB interactions.

Estimation of factor A and factor B main effects can easily be done in terms of comparisons among the regression coefficients. The use of the Scheffé and Bonferroni multiple comparison procedures presents no new issues. For instance, the S multiple for multiple comparisons among the factor A level means is defined as follows:

$$S^2 = (a-1)F(1-\alpha; a-1, nab - ab - 1) \tag{22.28}$$

and the B multiple is the same as in (22.18), with $n_T = nab$ and $r = ab$.

Example

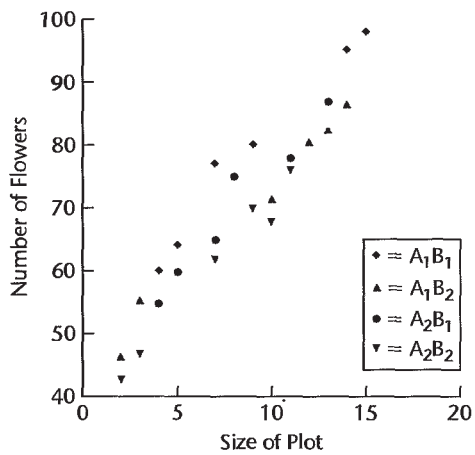
A horticulturist conducted an experiment to study the effects of flower variety (factor A: varieties LP, WB) and moisture level (factor B: low, high) on yield of salable flowers (Y). Because the plots were not of the same size, the horticulturist wished to use plot size (X) as the concomitant variable. Six replications were made for each treatment. A portion of the data are presented in Table 22.6. Figure 22.7 contains a symbolic scatter plot of the data. The model assumptions of linear relations between Y and the concomitant variable X, as well as of parallel slopes for the four treatments, appear to be reasonable here.

A fit of regression model (22.27) to the data by a computer regression package yielded the fitted regression function in Table 22.7a. The analyst plotted the data together with the fitted regression lines and made a variety of residual plots and tests (not shown). On the

E 22.6
alable
e s
le.

Factor *A* (flower variety) *i*	Factor *B* (moisture level) *j*			
	B_1 (low)		B_2 (high)	
	Y_{i1k}	X_{i1k}	Y_{i2k}	X_{i2k}
A_1 (variety LP)	98	15	71	10
	60	4	80	12
	...	...	...	...
	64	5	55	3
A_2 (variety WB)	55	4	76	11
	60	5	68	10
	...	...	...	...
	78	11	70	9

IGURE 22.7
bolic
catter Plot of
alable
owers and
ize of
lot—Salable
owers
xample.



basis of these diagnostics, the analyst was satisfied that regression model (22.27), which assumes parallel linear regression functions and constant error variances, is suitable here.

To examine the nature of the factor effects, we show in Figure 22.8 the estimated treatment means plot for the two moisture levels B_1 and B_2. These estimated means all correspond to plot size $X = \bar{X}_{...} = 8.25$ or $x = 0$. Any other plot size would yield exactly the same relationships as those in Figure 22.8. It appears from Figure 22.8 that there are no important interactions between flower variety and moisture level, and that there may be main effects for both factors, particularly for moisture level.

To study formally the factor effects, reduced models were formed by deleting from regression model (22.27) one predictor variable at a time (recall that both factors have only two levels), and the reduced models were then fitted. The extra sums of squares so obtained, as well as the error sum of squares for the full model, are presented in Table 22.7b, together with the degrees of freedom and mean squares. No total sum of squares is shown because the factor effect components are not orthogonal.

TABLE 22.7 Computer Output for Fit of Regression Model (22.27)—Salable Flowers Example.

(a) Fitted Regression Function

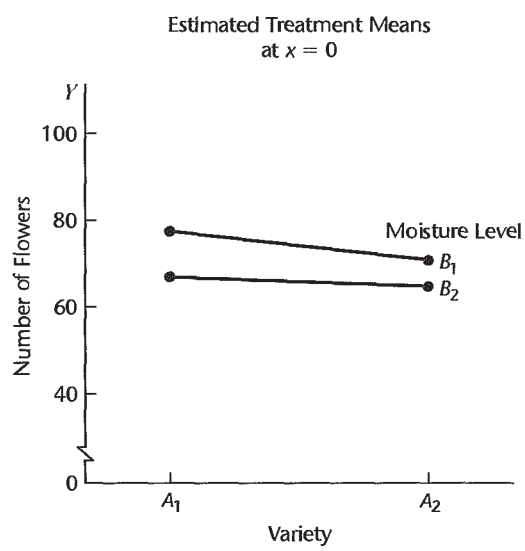
$$\hat{Y} = 70.0 + 2.04234I_1 + 3.68078I_2 + .81922I_1I_2 + 3.27688x$$

Regression Coefficient	Estimated Regression Coefficient	Estimated Standard Deviation
α_1	2.04234	.52108
β_1	3.68078	.51291
$(\alpha\beta)_{11}$	.81922	.51291
γ	3.27688	.13002

(b) Extra Sums of Squares

Effect	Source of Variation	*SS*	*df*	*MS*
Concomitant variable	$x \mid I_1, I_2, I_1I_2$	3,994.52	1	3,994.52
A	$I_1 \mid x, I_2, I_1I_2$	96.60	1	96.60
B	$I_2 \mid x, I_1, I_1I_2$	323.85	1	323.85
AB	$I_1I_2 \mid x, I_1, I_2$	16.04	1	16.04
	Error	119.48	19	6.2884

FIGURE 22.8 Estimated Treatment Means Plot—Salable Flowers Example.



We test first for the presence of interactions by means of the usual general linear statistic F^*, using the results in Table 22.7b:

$$F^* = \frac{SSR(I_1I_2|x, I_1, I_2)}{1} \div MSE = \frac{16.04}{6.2884} = 2.55$$

For $\alpha = .01$, we require $F(.99; 1, 19) = 8.18$. Since $F^* = 2.55 \leq 8.18$, we conclude no interactions are present. The P-value of the test is .13.

We now wish to compare both the factor A main effects and the factor B main effects by means of confidence intervals, with a 95 percent family confidence coefficient. Since $\alpha_2 = -\alpha_1$, we have for our example:

$$L_1 = \alpha_1 - \alpha_2 = \alpha_1 - (-\alpha_1) = 2\alpha_1$$

Similarly, we obtain for the comparison of factor B main effects:

$$L_2 = 2\beta_1$$

Point estimates are readily obtained from the results in Table 22.7a:

$$\hat{L}_1 = 2\hat{\alpha}_1 = 2(2.04234) = 4.08$$

$$\hat{L}_2 = 2\hat{\beta}_1 = 2(3.68078) = 7.36$$

The estimated standard deviations also follow easily, using (A.16b):

$$s\{\hat{L}_1\} = 2s\{\hat{\alpha}_1\} = 2(.52108) = 1.042$$

$$s\{\hat{L}_2\} = 2s\{\hat{\beta}_1\} = 2(.51291) = 1.026$$

We utilize the Bonferroni simultaneous estimation procedure for $g = 2$ comparisons. For a 95 percent family confidence coefficient, we require $t[1 - .05/2(2); 19] = t(.9875; 19) = 2.433$. The two desired confidence intervals therefore are:

$$1.5 = 4.08 - 2.433(1.042) \leq \alpha_1 - \alpha_2 \leq 4.08 + 2.433(1.042) = 6.6$$

$$4.9 = 7.36 - 2.433(1.026) \leq \beta_1 - \beta_2 \leq 7.36 + 2.433(1.026) = 9.9$$

With family confidence coefficient .95, we conclude that variety LP yields, on the average, between 1.5 and 6.6 more salable flowers for any given plot size than variety WB. Also, for any given plot size, the mean number of salable flowers is between 4.9 and 9.9 flowers greater for the low moisture level than for the high one, thus indicating a substantial effect of moisture level on yield.

If interactions had been present, we could have studied the nature of the interaction effects by, for instance, comparing the effect of the moisture level for each of the two flower varieties. It can be shown that this comparison is given by $(\alpha\beta)_{12} = -(\alpha\beta)_{11}$. Hence, we could estimate the desired interaction effect by using the estimated regression coefficient $\widehat{(\alpha\beta)}_{11}$ and its estimated standard deviation in Table 22.7a.

Covariance Analysis for Randomized Complete Block Designs

Covariance analysis can be employed to further reduce the experimental error variability in a randomized complete block design. The extension is a straightforward one from covariance analysis for a completely randomized design.

Covariance Model. The usual randomized block design model was given in (21.1). The covariance model for a randomized block design with one concomitant variable is obtained by simply adding a term (or several terms) for the relation between the response variable Y and the concomitant variable X. Assuming this relation can be described by a linear function, we obtain:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \gamma(X_{ij} - \overline{X}_{..}) + \varepsilon_{ij} \qquad i = 1, \ldots, n_b; j = 1, \ldots, r \qquad \textbf{(22.29)}$$

Regression Approach. The regression approach to covariance model (22.29) involves no new principles. We shall denote the centered variable $X_{ij} - \overline{X}_{..}$ in covariance model (22.29) by x_{ij}. Further, we shall again use $1, -1, 0$ indicator variables for the block and treatment effects. To illustrate an equivalent regression model, consider a randomized complete block design study with $n_b = 4$ blocks and $r = 3$ treatments. The regression model counterpart to covariance model (22.29) then is:

$$Y_{ij} = \mu_{..} + \rho_1 I_{ij1} + \rho_2 I_{ij2} + \rho_3 I_{ij3} + \tau_1 I_{ij4} + \tau_2 I_{ij5} + \gamma x_{ij} + \varepsilon_{ij} \qquad \text{Full model} \tag{22.30}$$

where:

$$I_1 = \begin{cases} 1 & \text{if experimental unit from block 1} \\ -1 & \text{if experimental unit from block 4} \\ 0 & \text{otherwise} \end{cases}$$

I_2, I_3 are defined similarly

$$I_4 = \begin{cases} 1 & \text{if experimental unit received treatment 1} \\ -1 & \text{if experimental unit received treatment 3} \\ 0 & \text{otherwise} \end{cases}$$

$$I_5 = \begin{cases} 1 & \text{if experimental unit received treatment 2} \\ -1 & \text{if experimental unit received treatment 3} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{ij} = X_{ij} - \overline{X}_{..}$$

To test for treatment effects:

$$\begin{aligned} H_0&: \ \tau_1 = \tau_2 = \tau_3 = 0 \\ H_a&: \ \text{not all } \tau_j \text{ equal zero} \end{aligned} \tag{22.31}$$

we would either need to fit the reduced model under H_0:

$$Y_{ij} = \mu_{..} + \rho_1 I_{ij1} + \rho_2 I_{ij2} + \rho_3 I_{ij3} + \gamma x_{ij} + \varepsilon_{ij} \qquad \text{Reduced model} \tag{22.32}$$

or else use the appropriate extra sum of squares. The test for treatment effects is then conducted in the usual way.

Comparisons of two treatment effects by the regression approach are straightforward. For estimating $\tau_1 - \tau_2$, for instance, we use the unbiased estimator $\hat{\tau}_1 - \hat{\tau}_2$ based on the estimated regression coefficients obtained when fitting the full model (22.30). The estimated variance of this estimator is:

$$s^2\{\hat{\tau}_1 - \hat{\tau}_2\} = s^2\{\hat{\tau}_1\} + s^2\{\hat{\tau}_2\} - 2s\{\hat{\tau}_1, \hat{\tau}_2\} \tag{22.33}$$

The estimated variance-covariance matrix of the regression coefficients, available in many regression package printouts, can then be used to obtain the required estimated variances and covariances.

Comment

Some computer packages for covariance analysis produce analyses that are only valid when all treatment sample sizes are equal. Computer packages should therefore be used with great care when the treatment sample sizes are unequal, to make sure that the package conducts the tests of interest. ■

5 Additional Considerations for the Use of Covariance Analysis

ovariance Analysis as Alternative to Blocking

At times, a choice exists between: (1) a completely randomized design, with covariance analysis used to reduce the experimental errors and (2) a randomized block design, with the blocks formed by means of the concomitant variable. Generally, the latter alternative is preferred. There are several reasons for this:

1. If the regression between the response variable and the concomitant (blocking) variable is linear, a randomized block design and covariance analysis are about equally efficient. If the regression is not linear but covariance analysis with a linear relationship is utilized, covariance analysis with a completely randomized design will tend to be not as effective as a randomized block design.
2. Randomized block designs are essentially free of assumptions about the nature of the relationship between the blocking variable and the response variable, while covariance analysis assumes a definite form of relationship.
3. Randomized block designs have somewhat fewer degrees of freedom available for experimental error than with covariance analysis for a completely randomized design. However, in all but small-scale experiments, this difference in degrees of freedom has little effect on the precision of the estimates.

Use of Differences

In a variety of studies, a prestudy observation X and a poststudy observation Y on the same variable are available for each unit. For instance, X may be the score for a subject's attitude toward a company prior to reading its annual report, and Y may be the score after reading the report. In this situation, an obvious alternative to covariance analysis is to do an analysis of variance on the differences $Y - X$. Sometimes, $Y - X$ is called an *index of response* because it makes one observation out of two.

If the slope of the treatment regression lines is $\gamma = 1$, analysis of covariance and analysis of variance on $Y - X$ are essentially equivalent. When $\gamma = 1$, covariance model (22.2) becomes:

$$Y_{ij} = \mu_{\cdot} + \tau_i + X_{ij} + \varepsilon_{ij} \tag{22.34}$$

which can be written as a regular analysis of variance model:

$$Y_{ij} - X_{ij} = \mu_{\cdot} + \tau_i + \varepsilon_{ij} \tag{22.34a}$$

Thus, if a unit change in X leads to about the same change in Y, it makes sense to perform an analysis of variance on $Y - X$ rather than to use covariance analysis, because

the analysis of variance model is a simpler model. If the regression slope is not near 1, however, covariance analysis may be substantially more effective than use of the differences $Y - X$.

In the earlier cracker promotion example, use of $Y - X$ would have been effective. It would have yielded the error mean square $MSE = 3.500$, which is practically the same as the error mean square for covariance analysis, $MSE = 3.506$ (see Table 22.3b). Recall that the regression slope in our example is close to 1 ($\hat{\gamma} = .899$), hence, the approximate equivalence of the two procedures.

Correction for Bias

The suggestion is sometimes made that analysis of covariance can be helpful in correcting for bias with observational data. With such data, the groups under study may differ substantially with respect to a concomitant variable, and this may bias the comparisons of the groups. Consider, for instance, a study in which attitudes toward no-fault automobile insurance were compared for persons who are risk averse and persons who are risk seeking. It was found that many persons in the risk-averse group tended to be older (50 to 70 years old), while many persons in the risk-seeking group tended to be younger (20 to 40 years old). In this type of situation, some researchers would advise that covariance analysis, with age as the concomitant variable, be employed to help remove any bias in the analysis of the observational data on attitudes toward no-fault insurance because the two age groups differ so much.

Even though there is great appeal in the idea of removing bias in observational data, covariance analysis should be used with caution for this purpose. In the first place, comparisons of means at a common value of X may require substantial extrapolation of the regression lines to a region where there are no or only few data points (in our example, to near 45 years). It may well be that the regression relationship used in the covariance analysis is not appropriate for substantial extrapolation. In the second place, the treatment variable may depend on the concomitant variable (or vice versa), which could affect the proper conclusions to be drawn.

Interest in Nature of Treatment Effects

Covariance analysis is sometimes employed for the principal purpose of shedding more light on the nature of the treatment effects, rather than merely for increasing the precision of the analysis. For instance, a market researcher in a study of the effects of three different advertisements on the maximum price consumers are willing to pay for a new type of home siding may use covariance analysis, with value of the consumer's home as the concomitant variable. The reason is because the researcher is truly interested in the relation for each advertisement between home value and maximum price. Reduction of error variance in this instance may be a secondary consideration.

As in all regression analyses, care must be used in drawing inferences about the causal nature of the relation between the concomitant variable and the response. In the advertising example, it might well be that value of a consumer's home is largely influenced by income. If this were so, the relation between value of the consumer's home and maximum price the consumer is willing to pay may actually be largely a reflection of the underlying relation between income and maximum price.

oblems

22.1. A student's reaction to the instructor's statement that covariance analysis is inappropriate when the treatment regression lines do not have the same slope was as follows: "It seems to me that this is ducking a real-world problem. If the treatment slopes are different, just use a covariance model that allows for different treatment slopes." Evaluate this reaction.

22.2. A survey analyst remarked: "When covariance analysis is used with survey data, there is a danger that the treatments may be related to the concomitant variable." What is the nature of the problem? Does this same problem exist when the treatments are randomly assigned to the experimental units?

22.3. Portray, analogously to the format of Figure 1.6 on page 11 for a regression model, the nature of covariance model (22.3) when there are three treatments and the parameter values are: $\mu_{\cdot} = 150$, $\tau_1 = 15$, $\tau_2 = -5$, $\tau_3 = -10$, $\gamma = 6$, $\bar{X}_{\cdot\cdot} = 70$, $\sigma = 5$. Show several distributions of Y for each treatment.

22.4. Refer to the cracker promotion example on page 926. A student stated, in discussing this case: "Strictly speaking, you cannot conclude anything about whether the three promotions differ in effectiveness because there was no control. The preceding period does not qualify as a control because it might have differed from the promotion period due to seasonal factors or other unique circumstances." Comment.

22.5. Refer to the cracker promotion example on pages 930 and 931, where three pairwise comparisons of treatment effects were made by the Scheffé procedure.

a. What would be the value of the Bonferroni multiple here for estimating the three comparisons?

b. Did the analyst obtain substantially less precise interval estimates using the Scheffé procedure, which permits making additional estimates without modifying the present ones?

22.6. State the analysis of covariance model for a single-factor study with four treatments when there are two concomitant variables, each with linear and quadratic terms in the model.

*22.7. Refer to **Productivity improvement** Problem 16.7. The economist also has information on annual productivity improvement in the prior year and wishes to use this information as a concomitant variable. The data on the prior year's productivity improvement (X_{ij}) follow.

	j											
i	1	2	3	4	5	6	7	8	9	10	11	12
1	8.2	7.9	7.0	5.7	7.2	7.0	6.5	7.9	6.3			
2	8.8	10.0	10.7	10.0	9.7	9.4	10.6	9.8	10.0	10.3	8.9	10.0
3	11.5	12.2	12.8	11.0	12.3	12.1						

a. Obtain the residuals for covariance model (22.3).

b. For each treatment, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

c. State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Could you conduct a formal test here as to whether the regression functions are linear? If so, how many degrees of freedom are there for the denominator mean square in the test statistic?

*22.8. Refer to **Productivity improvement** Problems 16.7 and 22.7. Assume that covariance model (22.3) is appropriate.

a. Prepare a symbolic scatter plot of the data. Does it appear that there are effects of the level of research and development expenditures on mean productivity improvement? Discuss.

b. State the regression model equivalent to covariance model (22.3) for this case; use 1, −1, 0 indicator variables. Also state the reduced regression model for testing for treatment effects.

c. Fit the full and reduced regression models and test for treatment effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Is $MSE(F)$ for the covariance model substantially smaller than MSE for the analysis of variance model in Problem 16.7d? Does this affect the conclusion reached about treatment effects? Does it affect the P-value?

e. Estimate the mean productivity improvement for firms with moderate research and development expenditures that had a prior productivity improvement of $X = 9.0$; use a 95 percent confidence interval.

f. Make all pairwise comparisons between the treatment effects; use either the Bonferroni or the Scheffé procedure with a 90 percent family confidence coefficient, whichever is more efficient. State your findings.

22.9. Refer to **Questionnaire color** Problem 16.8. It has been suggested to the investigator that size of parking lot might be a useful concomitant variable. The number of spaces (X_{ij}) in each parking lot utilized in the study follow.

	j				
i	1	2	3	4	5
1	300	381	226	350	100
2	153	334	473	264	325
3	144	359	296	243	252

a. Obtain the residuals for covariance model (22.3).

b. For each treatment, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

c. State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Could you conduct a formal test here as to whether the regression functions are linear? Explain.

22.10. Refer to **Questionnaire color** Problems 16.8 and 22.9. Assume that covariance model (22.3) is applicable.

a. Prepare a symbolic scatter plot of the data. Does it appear that there are color effects on the mean response rate? Discuss.

b. State the regression model equivalent to covariance model (22.3) for this case; use 1, −1, 0 indicator variables. Also state the reduced regression model for testing for treatment effects.

c. Fit the full and reduced regression models and test for treatment effects; use $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Is $MSE(F)$ for the covariance model substantially smaller than MSE for the analysis of variance model in Problem 16.8d? How does this affect the conclusion reached about treatment effects?

e. Estimate the mean response rate for blue questionnaires in parking lots of size $X = 280$; use a 90 percent confidence interval.

f. Make all pairwise comparisons between the treatment effects; use either the Bonferroni or the Scheffé procedure with a 90 percent family confidence coefficient, whichever is more efficient. State your findings.

22.11. Refer to **Rehabilitation therapy** Problem 16.9. The rehabilitation researcher wishes to use age of patient as a concomitant variable. The ages (X_{ij}) of patients in the study follow.

	j									
i	1	2	3	4	5	6	7	8	9	10
1	18.3	30.0	26.5	28.1	29.7	27.8	19.8	29.3		
2	20.8	25.2	29.2	20.0	21.5	22.1	19.7	24.7	20.2	22.9
3	22.7	28.7	18.9	18.0	21.7	20.0				

a. Obtain the residuals for covariance model (22.3).

b. For each treatment, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

c. State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Could you conduct a formal test here as to whether the regression functions are linear? Explain.

22.12. Refer to **Rehabilitation therapy** Problems 16.9 and 22.11. Assume that covariance model (22.3) is applicable.

a. Prepare a symbolic scatter plot of the data. Does it appear that there are effects of physical fitness status on the mean number of days required for therapy? Discuss.

b. State the regression model equivalent to covariance model (22.3) for this case; use 1, −1, 0 indicator variables. Also state the reduced regression model for testing for treatment effects.

c. Fit the full and reduced regression models and test for treatment effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Is $MSE(F)$ for the covariance model substantially smaller than MSE for the analysis of variance model in Problem 16.9d? Does this affect the conclusion reached about treatment effects? Does it affect the P-value?

e. Estimate the mean number of days required for therapy for patients of average physical fitness and age 24 years; use a 99 percent confidence interval.

f. Make all pairwise comparisons between the treatment effects; use either the Bonferroni or the Scheffé procedure with a 95 percent family confidence coefficient, whichever is more efficient. State your findings.

22.13. **Product display**. A manufacturer of felt-tip markers investigated by an experiment whether a proposed new display, featuring a picture of a physician, is more effective in drugstores

than the present counter display, featuring a picture of an athlete and designed to be located in the stationery area. Fifteen drugstores of similar characteristics were chosen for the study. They were assigned at random in equal numbers to one of the following three treatments: (1) present counter display in stationery area, (2) new display in stationery area, (3) new display in checkout area. Sales with the present display (X_{ij}) were recorded in all 15 stores for a three-week period. Then the new display was set up in the 10 stores receiving it, and sales for the next three-week period (Y_{ij}) were recorded in all 15 stores. The data on sales (in dollars) follow.

	j				
i	1	2	3	4	5
Treatment 1					
First 3 weeks	92	68	74	52	65
Second 3 weeks	69	44	58	38	54
Treatment 2					
First 3 weeks	77	80	70	73	79
Second 3 weeks	74	75	73	78	82
Treatment 3					
First 3 weeks	64	43	81	68	71
Second 3 weeks	66	49	84	75	77

The analyst wishes to analyze the effects of the three different display treatments by means of covariance analysis.

a. Obtain the residuals for covariance model (22.3).

b. For each treatment, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

c. State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Could you conduct a formal test here as to whether the regression functions are linear? Explain.

22.14. Refer to **Product display** Problem 22.13. Assume that covariance model (22.3) is applicable.

a. Prepare a symbolic scatter plot of the data. Does it appear that there are display effects on mean sales? Discuss.

b. State the regression model equivalent to covariance model (22.3) for this case; use $1, -1, 0$ indicator variables. Also state the reduced regression model for testing for treatment effects.

c. Fit the full and reduced regression models and test for treatment effects: use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Is $MSE(F)$ for the covariance model substantially smaller than the mean square error if analysis of variance model (16.2) had been employed?

e. Estimate the mean sales with display treatment 2 for stores whose sales in the preceding three-week period were \$75: use a 95 percent confidence interval.

f. Make all pairwise comparisons between the treatment effects: use either the Bonferroni or the Scheffé procedure with a 90 percent family confidence coefficient, whichever is more efficient. State your findings.

*22.15. Refer to **Cash offers** Problem 19.10. An analyst wishes to use each dealer's sales volume as a concomitant variable. The sales data (X_{ijk}, in hundred thousand dollars) follow.

$i = 1$		$i = 2$		$i = 3$	
$j = 1$	$j = 2$	$j = 1$	$j = 2$	$j = 1$	$j = 2$
3.0	3.5	6.5	2.2	5.0	4.0
5.1	4.2	4.1	5.4	3.1	.8
...	...	...	...	...	...
4.9	6.6	3.0	5.0	2.9	1.9

a. Obtain the residuals for covariance model (22.26).

b. For each treatment, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

c. State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

*22.16. Refer to **Cash offers** Problems 19.10 and 22.15. Assume that covariance model (22.26) is applicable.

a. State the regression model equivalent to covariance model (22.26) for this case; use 1, −1, 0 indicator variables. Fit this full model.

b. State the reduced regression models for testing for interaction and factor A and factor B main effects, respectively. Fit these reduced regression models.

c. Test for interaction effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test for factor A main effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Test for factor B main effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. For each factor, make all pairwise comparisons between the factor level main effects. Use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

22.17. Refer to **Eye contact effect** Problem 19.12. Age of personnel officer is to be used as a concomitant variable. The ages (X_{ijk}) of the personnel officers follow.

$i = 1$		$i = 2$	
$j = 1$	$j = 2$	$j = 1$	$j = 2$
42	51	43	42
30	35	53	47
...	...	...	...
35	49	49	56

a. Obtain the residuals for covariance model (22.26).

b. For each treatment, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered

residuals and their expected values under normality. What do you conclude from your analysis?

c. State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

22.18. Refer to **Eye contact effect** Problems 19.12 and 22.17. Assume that covariance model (22.26) is applicable.

a. State the regression model equivalent to covariance model (22.26) for this case; use 1, −1, 0 indicator variables. Fit this full model.

b. State the reduced regression models for testing for interaction and factor A and factor B main effects, respectively. Fit these reduced regression models.

c. Test for interaction effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test for factor A main effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Test for factor B main effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. Compare the gender main effects by means of a 99 percent confidence interval. Interpret your interval estimate.

g. Estimate the mean success rating by female personnel officers aged 40 when eye contact is present; use a 99 percent confidence interval.

*22.19. Refer to **Auditor training** Problem 21.5. The analyst wishes to examine whether use of pretraining statistical proficiency scores as a concomitant variable would help to reduce the experimental error variability significantly. The pretraining statistical proficiency scores for the auditors are as follows:

Block	Training Method (j)			Block	Training Method (j)		
i	1	2	3	i	1	2	3
1	93	98	91	6	75	74	78
2	94	93	94	7	79	76	72
3	89	91	92	8	71	69	64
4	86	84	90	9	74	71	70
5	78	76	84	10	63	68	64

a. Would you expect the auditor's age to have been a better concomitant variable here than the pretraining statistical proficiency score? Discuss.

b. State the regression model equivalent to covariance model (22.29); use 1, −1, 0 indicator variables.

c. Fit the full regression model.

d. State the reduced regression model for testing treatment effects. Fit the reduced model.

e. Test whether or not the training methods differ in mean effectiveness; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. Obtain a 95 percent confidence interval for $L = \tau_1 - \tau_2$. Interpret your interval estimate.

g. Has the error variance been reduced substantially by adding the concomitant variable? Explain.

22.20. Refer to **Fat in diets** Problem 21.7. The researcher wishes to examine whether each subject's body weight expressed as a percent of the ideal weight for that person would be a useful concomitant variable. The body weights as percents of the ideal weights for the 15 subjects are as follows:

Block	Fat Content of Diet		
i	$j = 1$	$j = 2$	$j = 3$
1	94	96	101
2	97	102	99
3	105	100	106
4	108	107	112
5	118	115	107

a. State the regression model equivalent to covariance model (22.29); use 1, −1, 0 indicator variables.

b. Fit the full regression model.

c. State the reduced regression model for testing treatment effects. Fit the reduced model.

d. Test whether or not the mean reductions in lipid level differ for the three diets; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Obtain confidence intervals for $L_1 = \tau_1 - \tau_2$ and $L_2 = \tau_2 - \tau_3$, using the Bonferroni procedure with a 95 percent family confidence coefficient. Interpret your interval estimates.

f. Has the error variance been reduced substantially by adding the concomitant variable? Explain.

*22.21. Refer to **Productivity improvement** Problems 22.7 and 22.8. The analyst is considering the use of the difference between the productivity improvements in the two years $(Y_{ij} - X_{ij})$ as the response variable with the regular analysis of variance model (22.29a).

a. Obtain the analysis of variance table.

b. How effective here is the use of differences with the regular ANOVA model compared to the use of covariance model (22.3)? Discuss.

22.22. Refer to **Product display** Problems 22.13 and 22.14. The analyst is considering the use of the difference in sales between the two periods $(Y_{ij} - X_{ij})$ as the response variable with the regular analysis of variance model (22.29a).

a. Obtain the analysis of variance table.

b. How effective here is the use of differences with the regular ANOVA model compared to the use of covariance model (22.3)? Discuss.

Exercise

22.23. (Calculus needed.) Denote $\mu. + \tau_i$ in covariance model (22.3) by Δ_i. Derive the least squares estimators for Δ_i and γ in covariance model (22.3).

Projects

22.24. Refer to the **SENIC** data set in Appendix C.1. The following hospitals are to be considered in a study of the effects of region (variable 9) on the mean length of hospital stay of patients (variable 2), with available facilities and services (variable 12) as a concomitant variable:

1–52 54 55 57 58 63 76 83 84 94 101 103 111

a. Obtain the residuals for covariance model (22.3).

b. For each region, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

c. State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

22.25. Refer to the **SENIC** data set in Appendix C.1 and Project 22.24. Assume that covariance model (22.3) is applicable.

a. Prepare a symbolic scatter plot of the data. Does it appear that there are region effects on the mean length of hospital stay? Discuss.

b. State the regression model equivalent to covariance model (22.3) for this case; use 1, −1, 0 indicator variables. Also state the reduced regression model for testing for treatment effects.

c. Fit the full and reduced regression models and test for treatment effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Make all pairwise comparisons between the region effects; use either the Bonferroni or the Scheffé procedure with a 90 percent family confidence coefficient, whichever is more efficient. State your findings.

22.26. Refer to the **Market share** data set in Appendix C.3 and Project 16.45. Use price (variable 3) as a concomitant variable.

a. Obtain the residuals for covariance model (22.3).

b. For each treatment, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

c. State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Could you conduct a formal test here as to whether the regression functions are linear? Explain.

22.27. Refer to the **Market share** data set in Appendix C.3 and Project 22.26.

a. Prepare a symbolic scatter plot of the data. Does it appear that mean monthly market share changes with the discount price and package promotion factor-level combinations? Discuss.

b. State the regression model equivalent to covariance model (22.3) for this case; use 1, −1, 0 indicator variables. Also state the reduced regression model for testing for treatment effects.

c. Fit the full and reduced regression models and test for treatment effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Is $MSE(F)$ for the covariance model substantially smaller than MSE for the analysis of variance model in Project 16.45? Does this affect the conclusion reached about treatment effects? Does it affect the P-value?

e. Estimate the average monthly market share for product with discount price present, package promotion absent, and average monthly price of product 2.5; use a 99 percent confidence interval.

f. Make all pairwise comparisons between the treatment effects; use either the Bonferroni or the Scheffé procedure with a 95 percent family confidence coefficient, whichever is more efficient. State your findings.

22.28. Refer to the **CDI** data set in Appendix C.2 and Project 19.53. The metropolitan areas identified in Project 19.53 are to be considered in a study of the effects of region (factor A: variable 17) and percent below poverty level (factor B: variable 13) on crime rate (variable 10 ÷ variable 5), with percent of population 65 or older (variable 7) as a concomitant variable. For purposes of this analysis of covariance study, percent below poverty level is to be classified into two categories: less than 8.0 percent, and 8.0 percent or more.

a. Obtain the residuals for covariance model (22.26).

b. For each treatment, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

c. State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = .001$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

22.29. Refer to the **CDI** data set in Appendix C.2 and Project 22.28. Assume that covariance model (22.26) is applicable.

a. State the regression model equivalent to covariance model (22.26) for this case; use 1, −1, 0 indicator variables. Fit this full model.

b. State the reduced regression models for testing for interaction and factor A and factor B main effects, respectively. Fit these reduced regression models.

c. Test for interaction effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test for factor A main effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Test for factor B main effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

22.30. Refer to the **Market share** data set in Appendix C.3 and Project 19.55. Use price (variable 3) as a concomitant variable.

a. Obtain the residuals for covariance model (22.26).

b. For each treatment, plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. What do you conclude from your analysis?

c. State the generalized regression model to be employed for testing whether or not the treatment regression lines have the same slope. Conduct this test using $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

22.31. Refer to the **Market share** data set in Appendix C.3 and Project 22.30.

a. State the regression model equivalent to covariance model (22.26) for this case; use 1, −1, 0 indicator variables. Fit this full model.

b. State the reduced regression models for testing for interaction and factor A and factor B main effects, respectively. Fit these reduced regression models.

c. Test for interaction effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test for factor A main effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Test for factor B main effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Case Studies

22.32. Refer to the **Prostate cancer** data set in Appendix C.5 and Case Study 16.49. Carry out a one-way analysis of covariance of this data set, where the response of interest is PSA level (variable 2), the single factor is Gleason score (variable 9), and the possible covariates are cancer volume (variable 3) and weight (variable 4). The analysis should consider transformations of the response variable and the covariates. Document steps taken in your analysis, and justify your conclusions.

22.33. Refer to the **Real estate sales** data set in Appendix C.7 and Case Study 16.50. Carry out a one-way analysis of covariance of this data set, where the response of interest is sales price (variable 2), the single factor is number of bedrooms (variable 4), and the possible covariates are finished square feet (variable 3) and lot size (variable 12). Recode the number of bedrooms into four categories: 0–2, 3, 4, and greater than or equal to 5. The analysis should consider transformations of the response variable and the covariates. Document steps taken in your analysis, and justify your conclusions.

22.34. Refer to the **Ischemic heart disease** data set in Appendix C.9 and Case Study 16.51. Carry out a one-way analysis of covariance of this data set, where the response of interest is total cost (variable 2), the single factor is total number of interventions (variable 5), and the possible covariates are duration (variable 10) and age (variable 3). Recode the number of interventions into six categories: 0, 1, 2, 3-4, 5–7, and greater than or equal to 8. The analysis should consider transformations of the response variable and the covariates. Document steps taken in your analysis, and justify your conclusions.

22.35. Refer to the **Real estate sales** data set in Appendix C.7 and Case Study 19.59. Carry out a balanced two-way analysis of covariance of this data set where the response of interest is sales price (variable 2), the two crossed factors are quality (variable 10) and style (variable 11), and the possible covariates are finished square feet (variable 3) and lot size (variable 12). Style is recoded as either 1 or not 1. Order the observations in the six factor-level-combination cells from smallest to largest observation number and retain the first 25 observations in each cell for a total of 150 observations. The analysis should consider transformations of the response variable and the covariates. Document the steps taken in your analysis and justify your conclusions.

22.36. Refer to the **Ischemic heart disease** data set in Appendix C.9 and Case Study 16.60. Carry out a balanced two-way analysis of covariance of this data set where the response of interest is total cost (variable 2), the two crossed factors are number of interventions (variable 5) and number of comorbidities (variable 9), and the possible covariates are duration (variable 10) and age (variable 3). Recode the number of interventions into six categories: 0, 1, 2, 3–4, 5–7, and greater than or equal to 8. Recode the number of comorbidities into two categories: 0–1, and greater than or equal to 2. Order the observations in the twelve factor-level-combination cells from smallest to largest observation number and retain the first 43 observations in each cell for a total of 516 observations. The analysis should consider transformations of the response variable and the covariates. Document the steps taken in your analysis and justify your conclusions.

apter 23

Two-Factor Studies with Unequal Sample Sizes

Up to this point in our discussion of two-factor studies we have restricted ourselves to equal treatment sample sizes for the two-factor ANOVA model (19.23). Often, however, two-factor studies involve unequal treatment sample sizes. The resulting imbalance destroys the orthogonality of the ANOVA decomposition. Consequently, the general linear test approach is utilized for ANOVA tests. In Sections 23.1 through 23.4 we shall take up procedures for handling two-factor studies with unequal treatment sample sizes. We continue to assume that all treatment means are of equal importance in these sections.

In occasional ANOVA studies, the treatment means are not of equal importance. This also makes the standard ANOVA decomposition inappropriate, and the general linear test approach consequently is employed. We consider in Section 23.5 procedures for conducting the analysis of variance when the treatment means have unequal importance. We conclude this chapter by discussing briefly in Section 23.6 the use of statistical computing packages in the presence of unequal treatment sample sizes.

23.1 Unequal Sample Sizes

Two-factor studies frequently involve unequal treatment or cell sample sizes for a variety of reasons. In observational studies, the investigator often has little or no control over the cell sample sizes. For example, in a comparative study of U.S. manufacturing practices, researchers examined the performance of manufacturing plants as a function of size of plant (factor A: small, medium, large) and ownership (factor B: Japan, United States). In this two-factor study, cell sample sizes for the six treatments were not under the complete control of the researchers. First, the number of plants available for study in each size-ownership category varied. Second, many plants were unable or unwilling to participate in the study.

Unequal treatment sample sizes are also encountered in experimental studies. For instance, an experimenter may seek to have the same number of cases for each treatment, but for a variety of reasons (e.g., illness of subject, incomplete records, technical problems) ends up with unequal cell sample sizes.

Another reason for unequal treatment sample sizes is that investigators in both observational and experimental studies may use larger sample sizes for treatments for which the cost

is lower. In still other instances, unequal treatment sample sizes may be desired to enable certain treatment means or certain linear combinations of treatment means to be estimated with greater precision. For example, a packaged foods manufacturer wished to measure the impact on consumer product ratings of a change from corn syrup to a low-calorie sweetener (factor A) in one of its breakfast cereals. Three categories of consumers, (factor B: children, female adults, and male adults) were considered to be important. It was known that about 60 percent of the consumers are children, 20 percent are adult males, and 20 percent are adult females. It was therefore considered to be reasonable to require that 60 percent of the subjects be children, 20 percent be adult males, and 20 percent be adult females to provide greater precision for the most important consumer group.

The fact that treatment sample sizes are unequal often does not affect the importance of the treatment means. As we just noted, sample sizes frequently are unequal for reasons that have nothing to do with the importance of treatment means. In our discussion of unequal treatment sample sizes in Sections 23.2–23.4, we shall continue to assume that all treatment means have the same importance. Procedures for handling ANOVA inferences when treatments have unequal importance are considered in Section 23.5.

Throughout Sections 23.1–23.3, we assume that there is at least one case for each treatment. Techniques for the analysis of studies with one or more cells empty are discussed in Section 23.4.

Notation

Our notation remains the same as before, except that the sample size for the treatment consisting of the ith level of factor A and the jth level of factor B will now be denoted by n_{ij}. The total number of cases for the ith level of factor A will be denoted by:

$$n_{i\cdot} = \sum_j n_{ij} \tag{23.1a}$$

the total number of cases for the jth level of factor B by:

$$n_{\cdot j} = \sum_i n_{ij} \tag{23.1b}$$

and the total number of cases for the entire study by:

$$n_T = \sum_i \sum_j n_{ij} \tag{23.1c}$$

The estimated treatment mean when factor A is at the ith level and factor B is at the jth level is defined as usual:

$$\bar{Y}_{ij\cdot} = \frac{Y_{ij\cdot}}{n_{ij}} \tag{23.2}$$

where:

$$Y_{ij\cdot} = \sum_{k=1}^{n_{ij}} Y_{ijk} \tag{23.2a}$$

23.2 Use of Regression Approach for Testing Factor Effects when Sample Sizes Are Unequal

When the treatment sample sizes are unequal, the analysis of variance for two-factor studies becomes more complex. The least squares equations are no longer of a simple structure that yields direct and easy solutions, and the regular analysis of variance formulas in (19.37) and (19.39) are now inappropriate. Furthermore, the factor effect component sums of squares are no longer orthogonal; that is, they do not sum to *SSTR*.

Hence, we will utilize the general linear test approach described in Section 2.8 when the treatment sample sizes are unequal. An easy way to obtain the proper error sums of squares for testing factor interactions and main effects by the general linear test approach is through the regression formulation of the ANOVA model described below. The only difference when cell sample sizes are unequal is that a reduced regression model needs to be fitted explicitly for each test of factor interactions and main effects because of the lack of orthogonality. Since no new principles are involved, we turn directly to an example to illustrate how ANOVA tests are conducted by means of the regression approach when the treatment sample sizes are unequal.

Regression Approach to Two-Factor Analysis of Variance

We shall explain the regression approach to two-factor analysis of variance in terms of the factor effects model (19.23):

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \tag{23.3}$$

As we know from (19.24), the mean responses for this model are given by:

$$E\{Y_{ijk}\} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} \tag{23.4}$$

To represent this model in matrix terms, we proceed in the same fashion as in the regression approach to single-factor ANOVA. Since $\sum \alpha_i = 0$, we need only $a - 1$ parameters α_i in the regression model, and we represent the parameter α_a as follows:

$$\alpha_a = -\alpha_1 - \alpha_2 - \cdots - \alpha_{a-1} \tag{23.5}$$

Hence, we utilize $a - 1$ indicator variables that can take on values 1, -1, or 0 for the α_i parameters, as in the single-factor ANOVA representation. Similarly, we need only $b - 1$ parameters β_j in the regression model, and we represent the parameter β_b as follows:

$$\beta_b = -\beta_1 - \beta_2 - \cdots - \beta_{b-1} \tag{23.6}$$

Hence, we utilize $b - 1$ indicator variables that can take on values 1, -1, or 0 for the β_j parameters.

For the interaction parameters, we need to recognize that:

$$\begin{aligned} \sum_i (\alpha\beta)_{ij} &= 0 \qquad j = 1, \ldots, b \\ \sum_j (\alpha\beta)_{ij} &= 0 \qquad i = 1, \ldots, a \end{aligned} \tag{23.7}$$

Therefore, we represent the parameters $(\alpha\beta)_{ib}$ and $(\alpha\beta)_{aj}$ as follows:

$$(\alpha\beta)_{ib} = -(\alpha\beta)_{i1} - (\alpha\beta)_{i2} - \cdots - (\alpha\beta)_{i,b-1} \tag{23.8}$$

$$(\alpha\beta)_{aj} = -(\alpha\beta)_{1j} - (\alpha\beta)_{2j} - \cdots - (\alpha\beta)_{a-1,j} \tag{23.9}$$

Indeed, because of the interrelations in the constraints in (23.7), only $(a-1)(b-1)$ terms $(\alpha\beta)_{ij}$ are needed in the regression model. As we shall demonstrate below, these are precisely the terms associated with the cross products between the indicator variables for the factor A and factor B main effects. We turn now to an example to illustrate how ANOVA tests are conducted by means of the regression approach when the treatment sample sizes are unequal.

Example

Synthetic growth hormone was administered at a clinical research center to growth hormone deficient, short children who had not yet reached puberty. The investigator was interested in the effects of a child's gender (factor A) and bone development (factor B) on the rate of growth induced by hormone administration. A child's bone development was classified into one of three categories: severely depressed, moderately depressed, mildly depressed. Three children were randomly selected for each gender-bone development group. The response variable (Y) of interest was the difference between the growth rate during growth hormone treatment and the normal growth rate prior to the treatment, expressed in centimeters per month. Four of the 18 children were unable to complete the year-long study, thus creating unequal treatment sample sizes. Note that this is an observational study. All children received the same hormone therapy, and, subsequently, changes in growth rates were observed for children in each bone development-by-gender category. No randomization of treatments to subjects was employed.

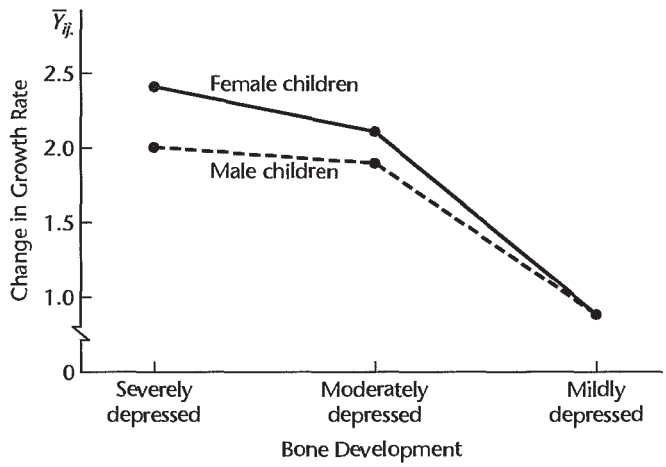
Table 23.1 presents the study data. A plot of the estimated treatment means is shown in Figure 23.1. It is clearly suggested there that a child's bone development has a major impact on the change in growth rate. The plot also raises the questions as to whether some interaction effects are present and whether the gender of a child affects the growth rate.

To test formally whether or not these factor effects are present, we utilize the general linear test approach and the equivalent regression model formulation because of the unequal sample sizes.

TABLE 23.1 Sample Data and Notation—Growth Hormone Example (growth rate difference in centimeters per month).

	Bone Development (factor B) j		
Gender (factor A) i	Severely Depressed (B_1)	Moderately Depressed (B_2)	Mildly Depressed (B_3)
Male (A_1)	1.4 (Y_{111}) 2.4 (Y_{112}) 2.2 (Y_{113})	2.1 (Y_{121}) 1.7 (Y_{122})	.7 (Y_{131}) 1.1 (Y_{132})
Mean	2.0 ($\bar{Y}_{11\cdot}$)	1.9 ($\bar{Y}_{12\cdot}$)	.9 ($\bar{Y}_{13\cdot}$)
Female (A_2)	2.4 (Y_{211})	2.5 (Y_{221}) 1.8 (Y_{222}) 2.0 (Y_{223})	.5 (Y_{231}) .9 (Y_{232}) 1.3 (Y_{233})
Mean	2.4 ($\bar{Y}_{21\cdot}$)	2.1 ($\bar{Y}_{22\cdot}$)	.9 ($\bar{Y}_{23\cdot}$)

FIGURE 23.1 Estimated Treatment Means Plot—Growth Hormone Example.



Development of Regression Model. The two-factor ANOVA model (19.23) here is:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad i = 1, 2;\ j = 1, 2, 3 \tag{23.10}$$

To express this model in regression terms, we utilize indicator variables that take on the values 1, −1, or 0, as explained below. Specifically, we need $a - 1 = 2 - 1 = 1$ indicator variable for the factor A main effects and $b - 1 = 3 - 1 = 2$ indicator variables for the factor B main effects. The interaction terms correspond to the cross products of the indicator variables for factor A and factor B main effects. Specifically, the regression model equivalent to ANOVA model (23.10) is:

$$\begin{aligned} Y_{ijk} = \mu_{..} + & \underbrace{\alpha_1 X_{ijk1}}_{\text{A main effect}} + \underbrace{\beta_1 X_{ijk2} + \beta_2 X_{ijk3}}_{\text{B main effect}} \\ & + \underbrace{(\alpha\beta)_{11} X_{ijk1} X_{ijk2} + (\alpha\beta)_{12} X_{ijk1} X_{ijk3}}_{\text{AB interaction effect}} + \varepsilon_{ijk} \qquad \text{Full model} \end{aligned} \tag{23.11}$$

where:

$$X_1 = \begin{cases} 1 & \text{if case from level 1 for factor A} \\ -1 & \text{if case from level 2 for factor A} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if case from level 1 for factor B} \\ -1 & \text{if case from level 3 for factor B} \\ 0 & \text{otherwise} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{if case from level 2 for factor B} \\ -1 & \text{if case from level 3 for factor B} \\ 0 & \text{otherwise} \end{cases}$$

The regression coefficients in (23.11) are the ANOVA model parameters:

$$\begin{aligned}
\mu_{..} \\
\alpha_1 &= \mu_{1.} - \mu_{..} \\
\beta_1 &= \mu_{.1} - \mu_{..} \\
\beta_2 &= \mu_{.2} - \mu_{..} \\
(\alpha\beta)_{11} &= \mu_{11} - \mu_{1.} - \mu_{.1} + \mu_{..} \\
(\alpha\beta)_{12} &= \mu_{12} - \mu_{1.} - \mu_{.2} + \mu_{..}
\end{aligned} \tag{23.12}$$

The remaining ANOVA model parameters are not required in the regression model because of the constraints in (19.23). Thus, for instance:

$$\begin{aligned}
\alpha_2 &= -\alpha_1 \\
\beta_3 &= -\beta_1 - \beta_2 \\
(\alpha\beta)_{13} &= -(\alpha\beta)_{11} - (\alpha\beta)_{12} \\
(\alpha\beta)_{21} &= -(\alpha\beta)_{11}
\end{aligned} \tag{23.13}$$

Table 23.2 repeats in column 1 a portion of the response data from Table 23.1. The codings of the indicator variables and the interaction terms are shown in columns 2–6. Note, for instance, that the codings for the first male child whose bone development is severely depressed ($i = 1, j = 1, k = 1$) are $X_1 = 1$, $X_2 = 1$, and $X_3 = 0$, so that $X_1X_2 = 1$ and $X_1X_3 = 0$. Table 23.3a presents the fitted regression function and regression ANOVA table when the full regression model (23.11) is fitted to the data, i.e., when Y in column 1 of Table 23.2 is regressed on the X variables in columns 2–6. Note that the fitted values for the full model are the estimated treatment means $\bar{Y}_{ij.}$, just as when all treatment sample sizes are equal. For instance, we have for the first case ($k = 1$) from treatment $i = 1, j = 1$:

$$\hat{Y}_{111} = 1.7 - .1(1) + .5(1) + .3(0) - .1(1) - 0(0) = 2.0 = \bar{Y}_{11.}$$

and for the last case ($k = 3$) from treatment $i = 2, j = 3$:

$$\hat{Y}_{233} = 1.7 - .1(-1) + .5(-1) + .3(-1) - .1(1) - 0(1) = .9 = \bar{Y}_{23.}$$

TABLE 23.2 Data for Regression Fits—Growth Hormone Example.

i	j	k	(1) Y	(2) X_1	(3) X_2	(4) X_3	(5) X_1X_2	(6) X_1X_3
1	1	1	1.4	1	1	0	1	0
1	1	2	2.4	1	1	0	1	0
	...		...	...	...	...	...	...
1	2	2	1.7	1	0	1	0	1
1	3	1	.7	1	−1	−1	−1	−1
	...		...	...	...	...	...	...
2	3	2	.9	−1	−1	−1	1	1
2	3	3	1.3	−1	−1	−1	1	1

BLE 23.3 Fits of Full and Reduced Regression Models—Growth Hormone Example.

(a) Full Model (23.11)

ce of tion	*SS*	*df*	$\hat{Y} = 1.7 - .1X_1 + .5X_2 + .3X_3 - .1X_1X_2 - 0.0X_1X_3$
ssion	4.4743	5	
or	1.3000	8	
	5.7743	13	

(b) Reduced Model (23.15)

ce of ation	*SS*	*df*	$\hat{Y} = 1.68 - .0857X_1 + .467X_2 + .327X_3$
ression	4.3989	3	
or	1.3754	10	
otal	5.7743	13	$SSE(R) - SSE(F) = 1.3754 - 1.3000 = .0754$

(c) Reduced Model (23.17)

urce of ariation	*SS*	*df*	$\hat{Y} = 1.69 + .444X_2 + .328X_3 - .0667X_1X_2 - .0167X_1X_3$
egression	4.3543	4	
rror	1.4200	9	
otal	5.7743	13	$SSE(R) - SSE(F) = 1.4200 - 1.3000 = .1200$

(d) Reduced Model (23.18)

ource of ariation	*SS*	*df*	$\hat{Y} = 1.63 + .0190X_1 + .0667X_1X_2 - .193X_1X_3$
egression	0.2846	3	
rror	5.4897	10	
otal	5.7743	13	$SSE(R) - SSE(F) = 5.4897 - 1.3000 = 4.1897$

Test for Interaction Effects. To test whether or not interaction effects are present, the ANOVA model alternatives:

$$H_0\text{: all } (\alpha\beta)_{ij} = 0$$
$$H_a\text{: not all } (\alpha\beta)_{ij} \text{ equal zero} \tag{23.14}$$

become for regression model (23.11):

$$H_0\text{: } (\alpha\beta)_{11} = (\alpha\beta)_{12} = 0$$
$$H_a\text{: not both } (\alpha\beta)_{11} \text{ and } (\alpha\beta)_{12} \text{ equal zero} \tag{23.14a}$$

Thus, we are simply testing whether or not two regression coefficients equal zero. The reduced regression model therefore is:

$$Y_{ijk} = \mu_{..} + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + \varepsilon_{ijk} \qquad \text{Reduced model} \tag{23.15}$$

When this reduced model is fitted by regressing Y in column 1 of Table 23.2 on X_1, X_2, and X_3 in columns 2–4, the results presented in Table 23.3b are obtained. The general linear test statistic (2.70) therefore is:

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F}$$

$$= \frac{1.3754 - 1.3000}{10 - 8} \div \frac{1.3000}{8} = \frac{.0377}{.1625} = .23$$

To control the risk of making a Type I error at $\alpha = .05$, we require $F(.95; 2, 8) = 4.46$. Since $F^* = .23 \leq 4.46$, we conclude H_0, that no interaction effects are present. The P-value for this test statistic is .80.

Tests for Factor Main Effects. We now proceed to test whether or not factor A and factor B main effects are present. The ANOVA model alternatives:

$$\begin{array}{ll} H_0\colon \alpha_1 = \alpha_2 = 0 & H_0\colon \beta_1 = \beta_2 = \beta_3 = 0 \\ H_a\colon \text{not both } \alpha_i \text{ equal zero} & H_a\colon \text{not all } \beta_j \text{ equal zero} \end{array} \tag{23.16}$$

become for regression model (23.11):

$$\begin{array}{ll} H_0\colon \alpha_1 = 0 & H_0\colon \beta_1 = \beta_2 = 0 \\ H_a\colon \alpha_1 \neq 0 & H_a\colon \text{not both } \beta_j \text{ equal zero} \end{array} \tag{23.16a}$$

The reduced regression models for testing for factor A main effects and factor B main effects therefore are:

$$\begin{aligned} Y_{ijk} = \mu_{..} &+ \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + (\alpha\beta)_{11} X_{ijk1} X_{ijk2} \\ &+ (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon_{ijk} \qquad \text{Reduced model} \end{aligned} \tag{23.17}$$

$$\begin{aligned} Y_{ijk} = \mu_{..} &+ \alpha_1 X_{ijk1} + (\alpha\beta)_{11} X_{ijk1} X_{ijk2} \\ &+ (\alpha\beta)_{12} X_{ijk1} X_{ijk3} + \varepsilon_{ijk} \qquad \text{Reduced model} \end{aligned} \tag{23.18}$$

Table 23.3c presents the results of fitting reduced model (23.17), where Y in column 1 of Table 23.2 is regressed on X_2, X_3, X_1X_2, and X_1X_3 in columns 3–6. Finally, Table 23.3d contains the results of fitting reduced model (23.18), where Y in column 1 of Table 23.2 is regressed on X_1, X_1X_2, and X_1X_3 in columns 2, 5, and 6. The two test statistics therefore are:

$$F_1^* = \frac{1.4200 - 1.3000}{9 - 8} \div \frac{1.3000}{8} = \frac{.1200}{.1625} = .74$$

$$F_2^* = \frac{5.4897 - 1.3000}{10 - 8} \div \frac{1.3000}{8} = \frac{2.0949}{.1625} = 12.89$$

LE 23.4
olidated
OVA
e—Growth
one
ample.

Source of Variation	*SS*	*df*	*MS*	F^*
Gender (A)	.1200	1	.1200	.74
Bone development (B)	4.1897	2	2.0949	12.89
AB interactions	.0754	2	.0377	.23
Error	1.3000	8	.1625	

For $\alpha = .05$, we require $F(.95; 1, 8) = 5.32$ and $F(.95; 2, 8) = 4.46$ for the two tests. Since $F_1^* = .74 \leq 5.32$ and $F_2^* = 12.89 > 4.46$, we conclude that there are no factor A main effects but that factor B main effects are present. The respective P-values for these two test statistics are .41 and .003.

Thus, these tests support the indications obtained previously from the estimated treatment means plot in Figure 23.1, that a child's bone development affects the change in growth rate during growth hormone treatment and that there are no gender and interaction effects. The family level of significance for the set of three tests just conducted, according to the Bonferroni inequality (4.4), is .15.

At this point, the next step in the analysis of the study results is to examine the nature of the bone development effects. We shall discuss this analysis in the next section.

Table 23.4 contains a consolidated ANOVA table presenting the results from fitting the four regression models in Table 23.3. The sum of squares for a factor effect in each instance is the difference between the error sums of squares for the reduced and full models shown in Table 23.3, and the associated degrees of freedom are the difference between the respective degrees of freedom for these error sums of squares. Note that a total sum of squares is not shown in Table 23.4 because the sums of squares for the three factor effects and for error do not add to *SSTO* when the treatment sample sizes are unequal.

Comment

In the event that pooling of sums of squares is desired for testing factor main effects when the test for interactions leads to the conclusion that there are none, as discussed in Section 19.10, the full regression model for testing factor A and factor B main effects needs to be revised. Specifically with reference to the growth hormone example, the full regression model in (23.11) would need to be revised by excluding the interaction effects and would be as follows:

$$Y_{ijk} = \mu_{..} + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + \varepsilon_{ijk} \qquad \text{Revised full model} \qquad \textbf{(23.19)}$$

■

23.3 Inferences about Factor Effects when Sample Sizes Are Unequal

The estimation of factor effects when the treatment sample sizes are unequal is completely analogous to when the sample sizes are equal. The nature of the analysis depends on whether or not important interactions are present. When no important interactions are present, the analysis generally is concerned with the factor level means $\mu_{i\cdot}$ and $\mu_{\cdot j}$. On the other

hand, when important interactions are present, the analysis usually focuses on the treatment means μ_{ij}.

The estimators and estimated variances presented in Chapter 19 for equal sample sizes must, of course, be modified to recognize the unequal treatment sample sizes. For instance, if interest is in estimating the factor level means $\mu_{i\cdot}$ as defined in (19.2) when all treatment means are of equal importance:

$$\mu_{i\cdot} = \frac{\sum_j \mu_{ij}}{b}$$

the appropriate estimator is simply the unweighted average of the estimated treatment means $\bar{Y}_{ij\cdot}$:

$$\hat{\mu}_{i\cdot} = \frac{\sum_j \bar{Y}_{ij\cdot}}{b}$$

These estimated factor level means are referred to as *least squares means*. Since the $\bar{Y}_{ij\cdot}$ are independent, the variance of this estimator is:

$$\sigma^2\{\hat{\mu}_{i\cdot}\} = \frac{1}{b^2}\sum_j \sigma^2\{\bar{Y}_{ij\cdot}\} = \frac{1}{b^2}\sum_j \frac{\sigma^2}{n_{ij}} = \frac{\sigma^2}{b^2}\sum_j \frac{1}{n_{ij}}$$

and the estimated variance is:

$$s^2\{\hat{\mu}_{i\cdot}\} = \frac{MSE}{b^2}\sum_j \frac{1}{n_{ij}}$$

Table 23.5 presents the formulas for the point estimator and estimated variance when estimating factor level means, pairwise comparisons of factor level means, and contrasts or linear combinations of factor level means, when the sample sizes are unequal. The corresponding formulas for treatment means, pairwise comparisons of treatment means, and contrasts or linear combinations of treatment means are also presented in this table.

All multiple comparison procedures applicable for equal sample sizes are appropriate when the treatment sample sizes are unequal. The Tukey pairwise comparison procedure then is conservative. The degrees of freedom associated with MSE are $n_T - ab$, as before. [Recall for equal sample sizes that $n_T = nab$; hence, $n_T - ab = (n-1)ab$.] Table 23.5 also presents the appropriate simultaneous comparison multiples for making inferences about factor level means or treatment means.

Test statistics and decision rules for simultaneous tests based on the Bonferroni, Tukey, and Scheffé procedures are easily adapted from the formulas in Chapter 19. The form of a test statistic does not change, but the degrees of freedom associated with MSE in each decision rule must now be expressed as $n_T - ab$.

Since no new issues are involved in estimating factor effects when the sample sizes are unequal, we proceed directly to two examples.

BLE 23.5 Point Estimators and Estimated Variances for Two-Factor Analyses when Sample Sizes Unequal.

(a) Factor Level Mean

$$\mu_{i\cdot} = \frac{\sum_j \mu_{ij}}{b} \qquad \mu_{\cdot j} = \frac{\sum_i \mu_{ij}}{a}$$

$$\hat{\mu}_{i\cdot} = \frac{\sum_j \bar{Y}_{ij\cdot}}{b} \qquad \hat{\mu}_{\cdot j} = \frac{\sum_i \bar{Y}_{ij\cdot}}{a} \tag{23.20}$$

$$s^2\{\hat{\mu}_{i\cdot}\} = \frac{MSE}{b^2}\sum_j \frac{1}{n_{ij}} \qquad s^2\{\hat{\mu}_{\cdot j}\} = \frac{MSE}{a^2}\sum_i \frac{1}{n_{ij}}$$

(b) Pairwise Comparison of Factor Level Means

$$D = \mu_{i\cdot} - \mu_{i'\cdot} \qquad D = \mu_{\cdot j} - \mu_{\cdot j'}$$

$$\hat{D} = \hat{\mu}_{i\cdot} - \hat{\mu}_{i'\cdot} \qquad \hat{D} = \hat{\mu}_{\cdot j} - \hat{\mu}_{\cdot j'} \tag{23.21}$$

$$s^2\{\hat{D}\} = \frac{MSE}{b^2}\sum_j \left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j}}\right) \qquad s^2\{\hat{D}\} = \frac{MSE}{a^2}\sum_i \left(\frac{1}{n_{ij}} + \frac{1}{n_{ij'}}\right)$$

(c) Contrast or Linear Combination of Factor Level Means

$$L = \sum_i c_i \mu_{i\cdot} \qquad L = \sum_j c_j \mu_{\cdot j}$$

$$\hat{L} = \sum_i c_i \hat{\mu}_{i\cdot} \qquad \hat{L} = \sum_j c_j \hat{\mu}_{\cdot j} \tag{23.22}$$

$$s^2\{\hat{L}\} = \frac{MSE}{b^2}\sum_i c_i^2 \sum_j \frac{1}{n_{ij}} \qquad s^2\{\hat{L}\} = \frac{MSE}{a^2}\sum_j c_j^2 \sum_i \frac{1}{n_{ij}}$$

(d) Confidence Interval Multiple

Single Estimate

$$t(1-\alpha/2; n_T - ab) \qquad t(1-\alpha/2; n_T - ab)$$

Multiple Comparisons

$$B = t(1-\alpha/2g; n_T - ab) \qquad B = t(1-\alpha/2g; n_T - ab)$$

$$T = \frac{1}{\sqrt{2}}q(1-\alpha; a, n_T - ab) \qquad T = \frac{1}{\sqrt{2}}q(1-\alpha; b, n_T - ab) \tag{23.23}$$

$$S^2 = (a-1)F(1-\alpha; a-1, n_T - ab) \qquad S^2 = (b-1)F(1-\alpha; b-1, n_T - ab)$$

(continued)

TABLE 23.5 Point Estimators and Estimated Variances for Two-Factor Analyses when Sample Sizes Are Unequal (concluded).

(e) Treatment Mean	
μ_{ij}	
$\hat{\mu}_{ij} = \bar{Y}_{ij\cdot}$	(23.24)
$s^2\{\hat{\mu}_{ij}\} = \dfrac{MSE}{n_{ij}}$	
(f) Pairwise Comparison of Treatment Means	
$D = \mu_{ij} - \mu_{i'j'}$	
$\hat{D} = \bar{Y}_{ij\cdot} - \bar{Y}_{i'j'\cdot}$	(23.25)
$s^2\{\hat{D}\} = MSE\left(\dfrac{1}{n_{ij}} + \dfrac{1}{n_{i'j'}}\right)$	
(g) Contrast or Linear Combination of Treatment Means	
$L = \sum\sum c_{ij}\mu_{ij}$	
$\hat{L} = \sum\sum c_{ij}\bar{Y}_{ij\cdot}$	(23.26)
$s^2\{\hat{L}\} = MSE\sum\sum \dfrac{c_{ij}^2}{n_{ij}}$	
(h) Confidence Interval Multiple	
Single Estimate	
$t(1-\alpha/2;\, n_T - ab)$	
Multiple Comparisons	
$B = t(1-\alpha/2g;\, n_T - ab)$	
$T = \dfrac{1}{\sqrt{2}}q(1-\alpha;\, ab,\, n_T - ab)$	(23.27)
$S^2 = (ab-1)F(1-\alpha;\, ab-1,\, n_T - ab)$	

Example 1—Pairwise Comparisons of Factor Level Means

We continue with the growth hormone example. We found earlier that a child's gender and bone development do not interact in their effects on the change in the growth rate when growth hormone is administered. We further found no main gender (factor A) effects, but concluded that a child's bone development (factor B) does affect the change in growth rate. We shall now analyze the nature of the bone development effects by means of pairwise comparisons among the three bone development groups. The Tukey multiple comparison procedure will be used. This procedure is conservative when sample sizes are unequal, and

use of the Bonferroni procedure would lead to wider confidence intervals here. The family confidence coefficient has been specified to be .90.

We use formulas (23.21) in Table 23.5 for the point estimates and estimated variances. The estimated treatment means are given in Table 23.1, and *MSE* is found in Table 23.4. For the pairwise comparisons of the bone development factor level means ($j = 1$: severely depressed; $j = 2$: moderately depressed; $j = 3$: mildly depressed), we obtain:

$$\hat{\mu}_{\cdot 1} = \frac{\bar{Y}_{11\cdot} + \bar{Y}_{21\cdot}}{2} = \frac{2.0 + 2.4}{2} = 2.2$$

$$\hat{\mu}_{\cdot 2} = \frac{\bar{Y}_{12\cdot} + \bar{Y}_{22\cdot}}{2} = \frac{1.9 + 2.1}{2} = 2.0$$

$$\hat{\mu}_{\cdot 3} = \frac{\bar{Y}_{13\cdot} + \bar{Y}_{23\cdot}}{2} = \frac{.9 + .9}{2} = .9$$

$$\hat{D}_1 = \hat{\mu}_{\cdot 1} - \hat{\mu}_{\cdot 2} = 2.2 - 2.0 = .2$$

$$\hat{D}_2 = \hat{\mu}_{\cdot 1} - \hat{\mu}_{\cdot 3} = 2.2 - .9 = 1.3$$

$$\hat{D}_3 = \hat{\mu}_{\cdot 2} - \hat{\mu}_{\cdot 3} = 2.0 - .9 = 1.1$$

$$s^2\{\hat{D}_1\} = \frac{.1625}{(2)^2}\left(\frac{1}{3} + \frac{1}{2} + \frac{1}{1} + \frac{1}{3}\right) = .0880 \qquad s\{\hat{D}_1\} = .297$$

$$s^2\{\hat{D}_2\} = \frac{.1625}{(2)^2}\left(\frac{1}{3} + \frac{1}{2} + \frac{1}{1} + \frac{1}{3}\right) = .0880 \qquad s\{\hat{D}_2\} = .297$$

$$s^2\{\hat{D}_3\} = \frac{.1625}{(2)^2}\left(\frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3}\right) = .0677 \qquad s\{\hat{D}_3\} = .260$$

For a 90 percent family confidence coefficient, we require:

$$T = \frac{1}{\sqrt{2}}q(.90; 3, 8) = \frac{1}{\sqrt{2}}(3.37) = 2.38$$

Hence, we obtain the following confidence intervals:

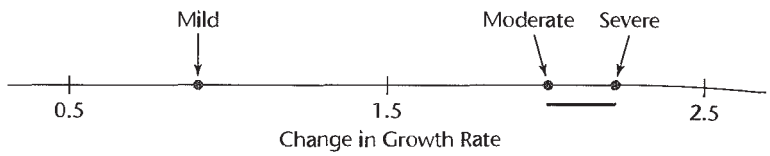
$$-.51 = .2 - 2.38(.297) \leq \mu_{\cdot 1} - \mu_{\cdot 2} \leq .2 + 2.38(.297) = .91$$

$$.59 = 1.3 - 2.38(.297) \leq \mu_{\cdot 1} - \mu_{\cdot 3} \leq 1.3 + 2.38(.297) = 2.01$$

$$.48 = 1.1 - 2.38(.260) \leq \mu_{\cdot 2} - \mu_{\cdot 3} \leq 1.1 + 2.38(.260) = 1.72$$

We conclude from these confidence intervals with 90 percent family confidence coefficient that growth hormone deficient, short children with mildly depressed bone development on the average have a substantially smaller increase in the growth rate than children with either moderately depressed or severely depressed bone development. Further, the latter two groups of children do not show significantly different mean changes in the growth rate. We summarize these findings in the following line plot of the estimated

factor level means:



Example 2—Single-Degree-of-Freedom Test

In the growth hormone example, a researcher wanted to know whether children with only mildly depressed bone development obtain, on the average, any increase in the growth rate with administration of growth hormone. Thus, the alternatives to be considered are those for a one-sided test:

$$H_0\colon \mu_{\cdot 3} \leq 0$$
$$H_a\colon \mu_{\cdot 3} > 0$$

The level of significance is to be controlled at $\alpha = .05$.

The test statistic to be employed is:

$$t^* = \frac{\hat{\mu}_{\cdot 3}}{s\{\hat{\mu}_{\cdot 3}\}}$$

We found earlier that $\hat{\mu}_{\cdot 3} = .9$ and $MSE = .1625$. Using (23.20), we obtain:

$$s^2\{\hat{\mu}_{\cdot 3}\} = \frac{.1625}{(2)^2}\left(\frac{1}{2} + \frac{1}{3}\right) = .0339 \qquad s\{\hat{\mu}_{\cdot 3}\} = .184$$

Hence, the test statistic is:

$$t^* = \frac{.9}{.184} = 4.89$$

For $\alpha = .05$, we require $t(.95; 8) = 1.860$. Therefore the one-sided decision rule is:

$$\text{If } t^* \leq 1.860, \text{ conclude } H_0$$
$$\text{If } t^* > 1.860, \text{ conclude } H_a$$

Since $t^* = 4.89 > 1.860$, we conclude H_a, that the mean change in the growth rate for children with mildly depressed bone development is greater than zero. The one-sided P-value for this test statistic is .0006.

23.4 Empty Cells in Two-Factor Studies

Occasionally after a two-factor study has been completed, it turns out that there are no cases in one or several treatment cells. Not only are the treatment sample sizes unequal then, but there is no sample information about the treatment means for the empty cells. Consider again Table 23.1 for the growth hormone study. Note that two female children with severely depressed bone condition dropped out of the study before its completion so that only one

case ($n_{21} = 1$) is present for that treatment. We can imagine easily that all three of these children could have dropped out of the study. Then we would have had $n_{21} = 0$, and no sample information would have been available about the treatment mean μ_{21}.

Partial Analysis of Factor Effects

When one or several treatment cells are empty, the analysis of variance for unequal sample sizes by means of the equivalent regression model, as explained earlier, cannot be conducted. This does not mean, however, that the entire two-factor study has become useless. A variety of partial analyses usually can be conducted that will provide at least some information about the nature of the factor effects. The analyses that can be undertaken depend on the particular cells for which no sample information is available. We illustrate by means of an example how partial information can be obtained from two-factor studies with empty cells.

Example

In the growth hormone example, suppose that there are no observations for female children with severely depressed bone development; i.e., $n_{21} = 0$. In that case no sample information is available about the treatment mean μ_{21}. We represent this situation in Figure 23.2a.

Partial information about interactions can still be obtained by restricting attention to children with moderately depressed and mildly depressed bone development, as represented in Figure 23.2b. For these children, interactions are present if the differences between the

FIGURE 23.2 Schematic Representation of Growth Hormone Study with Empty Cell—Growth Hormone Example ($n_{21} = 0$).

	Bone Development		
Gender	Severely Depressed B_1	Moderately Depressed B_2	Mildly Depressed B_3
(a) Empty Cell			
Male (A_1)	μ_{11}	μ_{12}	μ_{13}
Female (A_2)	Empty cell	μ_{22}	μ_{23}
(b) Partial Study of Interactions			
Male (A_1)		μ_{12}	μ_{13}
Female (A_2)		μ_{22}	μ_{23}
(c) Partial Study of Factor A and Factor B Main Effects			
Male (A_1)		μ_{12}	μ_{13}
Female (A_2)		μ_{22}	μ_{23}
(d) Partial Study of Factor B Main Effects			
Male (A_1)	μ_{11}	μ_{12}	μ_{13}
Female (A_2)			

treatment means for the two genders are not the same for the two bone development groups. The two differences are:

$$\mu_{12} - \mu_{22} \qquad \mu_{13} - \mu_{23}$$

Thus, we consider the following contrast among the treatment means:

$$L = \mu_{12} - \mu_{22} - \mu_{13} + \mu_{23}$$

We can either estimate L by means of a confidence interval and note whether or not the interval includes zero, or we can conduct a single degree of freedom test to establish whether or not interactions are present. With either approach, we use MSE based on all sample observations so that the associated degrees of freedom for MSE would be $n_T - (ab - 1) = 13 - 5 = 8$ (remember that $n_{21} = 0$ now).

If the partial analysis of interactions were to suggest that no interactions are present, the effect of gender can be studied by comparing the factor level means excluding children with severely depressed bone development, as represented in Figure 23.2c:

$$\mu_{1\cdot} = \frac{\mu_{12} + \mu_{13}}{2} \qquad \mu_{2\cdot} = \frac{\mu_{22} + \mu_{23}}{2}$$

In addition, the effect of bone development can be studied for male children by comparing the treatment means μ_{11}, μ_{12}, and μ_{13}, as represented in Figure 23.2d, or it can be studied for children of both genders by excluding those with severely depressed bone development, as represented in Figure 23.2c:

$$\mu_{\cdot 2} = \frac{\mu_{12} + \mu_{22}}{2} \qquad \mu_{\cdot 3} = \frac{\mu_{13} + \mu_{23}}{2}$$

Analysis if Model with No Interactions Can Be Employed

Occasionally, information is available from previous studies that the two factors in a two-factor study do not interact. In that case, a model with no interaction effects can be employed. Such a model was introduced in (20.1) for the case $n = 1$. When there are n_{ij} observations for the treatment consisting of the ith level of factor A and the jth level of factor B, the no-interaction model is:

$$Y_{ijk} = \mu_{\cdot\cdot} + \alpha_i + \beta_j + \varepsilon_{ijk} \qquad \text{No-interaction model} \qquad \textbf{(23.28)}$$

When no-interaction model (23.28) is appropriate, the analysis of variance and the analysis of factor main effects can be conducted by means of the equivalent regression model even when one or several cells are empty, as long as relevant other cells are not empty. [The relevant other cells are ones that satisfy the relations in (19.7b).]

The reason why the usual analysis of variance by means of the equivalent regression model can be conducted for ANOVA model (23.28), even though one or more cells are empty, is that the assumption of no interactions permits us in effect to estimate the empty cell means. Conceptually, estimation of an empty cell mean, say μ_{21}, requires two steps. First, we need to estimate the treatment means for the nonempty cells. These estimates are more complicated than simply using the estimated treatment means because the model assumption of no interactions needs to be utilized. We encountered such estimates for a no-interaction model in Chapter 20 when we considered studies where $n = 1$ for each cell. Once we have estimates of the treatment means μ_{ij} for the nonempty cells, the second step in

estimating the empty cell mean μ_{21} is to utilize the relation in (19.7b) for the no-interaction case, whereby we can express μ_{21} in terms of three other treatment means. For instance, μ_{21} can be estimated from $\hat{\mu}_{21} = \hat{\mu}_{22} + \hat{\mu}_{11} - \hat{\mu}_{12}$.

Example

In the growth hormone example, suppose that the cell for female children with severely depressed bone development is empty. From past knowledge, the researcher is able to assume that there are no interactions between gender and bone development. In that case, regression model (23.3) reduces to:

$$Y_{ijk} = \mu_{..} + \alpha_1 X_{ijk1} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + \varepsilon_{ijk} \qquad \text{Full model} \qquad \textbf{(23.29)}$$

To test for, say, gender main effects, we first fit this full model and obtain $SSE(F)$. The alternatives to be tested are:

$$H_0\colon \alpha_1 = 0$$
$$H_a\colon \alpha_1 \neq 0$$

Hence, the reduced model is:

$$Y_{ijk} = \mu_{..} + \beta_1 X_{ijk2} + \beta_2 X_{ijk3} + \varepsilon_{ijk} \qquad \text{Reduced model} \qquad \textbf{(23.30)}$$

We then fit this reduced model, obtain $SSE(R)$, and calculate the general linear test statistic (2.70) in the usual fashion. A test for bone development effects is carried out similarly.

Comments

1. We need to caution that it is not appropriate in the presence of empty cells to use a no-interaction model as the full model when no prior information about the absence of interactions is available. Only partial analyses of factor effects should then be undertaken, as explained earlier.

2. We have considered one cause of empty cells, when cases are missing or lost at random in an experimental study or when the sample in an observational study fails to include any cases for a particular cell. In these situations, the cell mean for the empty cell exists even though no cases are available for that cell. In contrast, a *structural empty cell* occurs when it is known a priori that it is impossible to obtain data for that cell. In this latter situation, the factorial structure is partially destroyed since the cell mean for the empty cell does not exist, and it is therefore meaningless to estimate the mean for such a structural empty cell on the basis of the other cases. ■

Missing Observations in Randomized Complete Block Designs

There are occasions when one or several observations in a randomized complete block design are "missing"—a subject may have been sick, a record may have been mislaid, a treatment may have been applied incorrectly in one instance. Such missing responses destroy the balance (orthogonality) of the complete block design and make the usual ANOVA calculations inappropriate. However, the regression approach discussed in Section 23.2, is ordinarily still appropriate when there are missing responses.

Since no new principles are involved, we turn to an example to illustrate the use of the regression approach when observations are missing in a randomized block design experiment.

Example

Table 23.6a contains the data for a simple randomized block design experiment with $r = 3$ treatments and $n_b = 3$ blocks, where observation Y_{11} is missing. We set up the regression model equivalent to randomized block design model (21.1) as follows:

$$Y_{ij} = \mu_{..} + \underbrace{\rho_1 X_{ij1} + \rho_2 X_{ij2}}_{\text{Block effect}} + \underbrace{\tau_1 X_{ij3} + \tau_2 X_{ij4}}_{\text{Treatment effect}} + \varepsilon_{ij} \qquad \text{Full model} \qquad (23.31)$$

where:

$$X_1 = \begin{cases} 1 & \text{if experimental unit from block 1} \\ -1 & \text{if experiment unit from block 3} \\ 0 & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if experimental unit from block 2} \\ -1 & \text{if experiment unit from block 3} \\ 0 & \text{otherwise} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{if experimental unit received treatment 1} \\ -1 & \text{if experiment unit received treatment 3} \\ 0 & \text{otherwise} \end{cases}$$

$$X_4 = \begin{cases} 1 & \text{if experimental unit received treatment 2} \\ -1 & \text{if experiment unit received treatment 3} \\ 0 & \text{otherwise} \end{cases}$$

Table 23.6b repeats the Y observations in column 1 and presents the four indicator variable in columns 2–5.

TABLE 23.6 Example of Missing Observation in Randomized Block Design ($r = 3, n_b = 3$).

(a) Response Data

Block i	Treatment (j) 1	2	3
1	Missing	10	9
2	11	10	7
3	6	4	3

(b) Regression Variables

i	j	(1) Y	(2) X_1	(3) X_2	(4) X_3	(5) X_4
1	2	10	1	0	0	1
1	3	9	1	0	−1	−1
2	1	11	0	1	1	0
2	2	10	0	1	0	1
2	3	7	0	1	−1	−1
3	1	6	−1	−1	1	0
3	2	4	−1	−1	0	1
3	3	3	−1	−1	−1	−1

The analysis of variance for testing treatment effects and block effects is carried out in the usual manner by first fitting the full model (23.31) and then fitting each of the following reduced models:

Test for Block Effects

$$Y_{ij} = \mu_{..} + \tau_1 X_{ij3} + \tau_2 X_{ij4} + \varepsilon_{ij} \qquad \text{Reduced model} \qquad \textbf{(23.32)}$$

Test for Treatment Effects

$$Y_{ij} = \mu_{..} + \rho_1 X_{ij1} + \rho_2 X_{ij2} + \varepsilon_{ij} \qquad \text{Reduced model} \qquad \textbf{(23.33)}$$

The extra sums of squares $SSR(X_1, X_2|X_3, X_4)$ for blocks and $SSR(X_3, X_4|X_1, X_2)$ for treatments are then calculated in the usual manner. Table 23.7a presents these extra sums of squares for our example obtained from fitting the full and reduced models, as well as the error sum of squares for the full model. No total sum of squares is shown because of lack of orthogonality as a result of the missing observation.

TABLE 23.7 ANOVA Table and Other Regression Output—Missing Data Example of Table 23.6.

(a) ANOVA Table

Source of Variation	*SS*	*df*	*MS*
Blocks	53.83	2	26.92
Treatments	12.50	2	6.25
Error	1.33	3	.44

(b) Estimated Regression Coefficients for Full Model (23.31)

Regression Coefficient	Estimated Regression Coefficient
$\mu_{..}$	$\hat{\mu}_{..} = 8.000$
ρ_1	$\hat{\rho}_1 = 2.333$
ρ_2	$\hat{\rho}_2 = 1.333$
τ_1	$\hat{\tau}_1 = 1.667$
τ_2	$\hat{\tau}_2 = 0.0$

(c) Estimated Variance-Covariance Matrix of Regression Coefficients

$$\begin{array}{c} \\ \hat{\mu}_{..} \\ \hat{\rho}_1 \\ \hat{\rho}_2 \\ \hat{\tau}_1 \\ \hat{\tau}_2 \end{array} \begin{array}{c} \begin{array}{ccccc} \hat{\mu}_{..} & \hat{\rho}_1 & \hat{\rho}_2 & \hat{\tau}_1 & \hat{\tau}_2 \end{array} \\ \begin{bmatrix} .06173 & & & & \\ .02469 & .14815 & & & \\ -.01235 & -.07407 & .11111 & & \\ .02469 & .04938 & -.02469 & .14815 & \\ -.01235 & -.02469 & .01235 & -.07407 & .11111 \end{bmatrix} \end{array}$$

The test for treatment effects is conducted as usual. From Table 23.7a we find:

$$F^* = \frac{MSR(X_3, X_4|X_1, X_2)}{MSE} = \frac{6.25}{.44} = 14.2$$

For $\alpha = .05$, we need $F(.95; 2, 3) = 9.55$. Since $F^* = 14.2 > 9.55$, we conclude that differential treatment effects are present. The P-value of this test is .03. The test for block effects can be carried out along similar lines when it is of interest.

No new problems are encountered with the regression approach in analyzing fixed treatment effects when there are missing observations. For instance, to estimate the pairwise comparison $L = \mu_{\cdot 1} - \mu_{\cdot 3} = \tau_1 - \tau_3$, we utilize the fact that $\tau_3 = -\tau_1 - \tau_2$ so that we have:

$$L = \mu_{\cdot 1} - \mu_{\cdot 3} = \tau_1 - \tau_3 = \tau_1 - (-\tau_1 - \tau_2) = 2\tau_1 + \tau_2 \tag{23.34}$$

An unbiased estimator of (23.34) is:

$$\hat{L} = 2\hat{\tau}_1 + \hat{\tau}_2 \tag{23.35}$$

whose estimated variance is, using (A.30b):

$$s^2\{\hat{L}\} = 4s^2\{\hat{\tau}_1\} + s^2\{\hat{\tau}_2\} + 4s\{\hat{\tau}_1, \hat{\tau}_2\} \tag{23.36}$$

Table 23.7b contains the estimated regression coefficients for the full model, and Table 23.7c contains the estimated variance-covariance matrix of the regression coefficients. We therefore obtain the following estimates:

$$\hat{L} = 2(1.667) + 0.0 = 3.334$$
$$s^2\{\hat{L}\} = 4(.14815) + .11111 + 4(-.07407) = .4074$$

so that the estimated standard deviation is $s\{\hat{L}\} = .638$. A 95 percent confidence interval for L requires $t(.975; 3) = 3.182$, yielding the confidence limits $3.334 \pm 3.182(.638)$ and the confidence interval:

$$1.3 \leq \mu_{\cdot 1} - \mu_{\cdot 3} \leq 5.4$$

23.5 ANOVA Inferences when Treatment Means Are of Unequal Importance

On occasion, the treatment means μ_{ij} in a two-factor study are not of equal importance, so the unweighted factor level means $\mu_{\cdot j}$ and $\mu_{i\cdot}$ defined in (19.1) and (19.2) are not relevant.

Example

In a breakfast cereal study 60 percent of the consumers of this product were children, 20 percent male adults, and 20 percent female adults. In this study, factor A was type of sweetener ($i = 1$: corn syrup, $i = 2$: low-calorie sweetener) and factor B was consumer category ($j = 1$: child, $j = 2$: male adult, $j = 3$: female adult). The company wishes to determine if a change to a low-calorie sweetener will change the mean rating of its product in the *population of consumers*. Here, the treatment means μ_{ij} have unequal importance

and the company therefore wishes to compare the two weighted means:

$$\text{Corn syrup: } .6\mu_{11} + .2\mu_{12} + .2\mu_{13}$$
$$\text{Low-calorie sweetener: } .6\mu_{21} + .2\mu_{22} + .2\mu_{23}$$

This can be done by estimating the contrast:

$$L = (.6\mu_{11} + .2\mu_{12} + .2\mu_{13}) - (.6\mu_{21} + .2\mu_{22} + .2\mu_{23})$$

or by testing the alternatives:

$$H_0\text{: } L = 0$$
$$H_a\text{: } L \neq 0$$

Note the use of the weights .6, .2, and .2 to reflect the unequal importance of the treatment means μ_{ij}.

Estimation of Treatment Means and Factor Effects

Estimation of treatment means and factor effects when the treatment means have unequal importance does not lead to any additional complexities. The general formulas in Section 23.3 for estimating treatment means μ_{ij} and for contrasts of treatment means still apply. We illustrate the analysis of factor effects when the treatment means are of unequal importance by returning to the mathematics learning example in Table 19.11.

Example

A school administrator in the mathematics learning example had requested information about which teaching method leads to better learning of college mathematics when 20 percent of the students in the class have excellent quantitative ability, 50 percent have good ability, and 30 percent have moderate ability. The mean learning scores for such a class mix with the two teaching methods are the following linear combinations of the treatment means:

$$\text{Abstract method: } L_1 = .2\mu_{11} + .5\mu_{12} + .3\mu_{13}$$
$$\text{Standard method: } L_2 = .2\mu_{21} + .5\mu_{22} + .3\mu_{23}$$

We assume here that the mean learning scores for students with different quantitative abilities will not be affected by a class mix that is somewhat different from the one in the experimental study.

Point estimates of the mean scores are (data in Table 19.11a):

$$\hat{L}_1 = .2(92) + .5(81) + .3(73) = 80.8$$
$$\hat{L}_2 = .2(90) + .5(86) + .3(82) = 85.6$$

The difference between the two mean scores is a contrast:

$$L = L_1 - L_2$$

This contrast is estimated to be:

$$\hat{L} = \hat{L}_1 - \hat{L}_2 = 80.8 - 85.6 = -4.8$$

We can obtain the estimated variance of $\hat{L}$ by (19.93b) since there are equal sample sizes

here:

$$s^2\{\hat{L}\} = \frac{28}{21}[(.2)^2 + (.5)^2 + (.3)^2 + (-.2)^2 + (-.5)^2 + (-.3)^2] = 1.013$$

so that the estimated standard deviation is $s\{\hat{L}\} = 1.006$. For a 95 percent confidence coefficient, we require $t(.975; 120) = 1.980$. Hence, the confidence limits are $-4.8 \pm 1.980(1.006)$ and the desired confidence interval is:

$$-6.79 \leq L \leq -2.81$$

With 95 percent confidence we conclude that the standard teaching method is better for the specified class mix, leading to a mean learning score that is at least 2.81 points greater than that for the abstract teaching method and may be as much as 6.79 points greater.

Test for Interactions

The test for interactions also is not affected by unequal importance of treatment means since this test is concerned with the parallelism, or lack of it, of the treatment mean curves. This was illustrated in Figures 19.3, 19.4, and 19.5. The treatment mean curves are based solely on the individual treatment means μ_{ij} and hence do not involve averages of the treatment means. Thus, the test for interactions is conducted as explained in Section 19.6 when the sample sizes are equal and as explained in Section 23.2 when the sample sizes are unequal, whether the treatment means are of equal or unequal importance.

Tests for Factor Main Effects by Use of Equivalent Regression Models

Tests for factor main effects when the treatment means are of unequal importance are carried out by the general linear test approach of Chapter 2. First, we shall explain how to implement factor tests with the general linear test approach by use of equivalent regression models; we then shall explain implementation by means of a matrix formulation.

When the treatment means are of unequal importance, the use of equivalent regression models to carry out the general linear test approach is easiest when cell means model (19.15) is employed. Since no new principles with the regression approach are involved, we turn to an example to illustrate the tests for main effects.

Example

In the growth hormone example in Table 23.1, it is known that twice as many male as female children undergo growth hormone treatment therapy, and that this ratio is the same for children who have severe, moderate, and mild depression in bone development. Inferences are desired about the target population of children undergoing therapy. Specifically, we wish to test whether or not the state of bone development affects the change in growth rate in the target population. The alternatives therefore are:

$$H_0: \frac{2\mu_{11} + \mu_{21}}{3} = \frac{2\mu_{12} + \mu_{22}}{3} = \frac{2\mu_{13} + \mu_{23}}{3} \tag{23.37}$$

H_a: not all equalities hold

We restate the alternative H_0 in the following equivalent fashion:

$$H_0: \begin{cases} \dfrac{2\mu_{11} + \mu_{21}}{3} - \dfrac{2\mu_{12} + \mu_{22}}{3} = 0 \\ \dfrac{2\mu_{11} + \mu_{21}}{3} - \dfrac{2\mu_{13} + \mu_{23}}{3} = 0 \end{cases} \tag{23.37a}$$

Implementation of the general linear test (2.70) requires that we fit the full model and then fit the reduced model under H_0. The full ANOVA model is cell means model (19.15):

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

Following the example in (16.85), we obtain the equivalent full regression model:

$$\begin{aligned} Y_{ijk} = {} & \mu_{11}X_{ijk1} + \mu_{12}X_{ijk2} + \mu_{13}X_{ijk3} + \mu_{21}X_{ijk4} \\ & + \mu_{22}X_{ijk5} + \mu_{23}X_{ijk6} + \varepsilon_{ijk} \qquad \text{Full model} \end{aligned} \tag{23.38}$$

where:

$$X_1 = \begin{cases} 1 & \text{if case from level 1 of factor } A \text{ and level 1 of factor } B \\ 0 & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if case from level 1 of factor } A \text{ and level 2 of factor } B \\ 0 & \text{otherwise} \end{cases}$$

$$\vdots$$

$$X_6 = \begin{cases} 1 & \text{if case from level 2 of factor } A \text{ and level 3 of factor } B \\ 0 & \text{otherwise} \end{cases}$$

Table 23.8 repeats in column 1 a portion of the data on the Y observations from Table 23.1 and presents in columns 2–7 the codings of the X variables for the full model. Note, for instance, that the codings of the X variables for observation Y_{111} are $X_1 = 1$, $X_2 = X_3 = X_4 = X_5 = X_6 = 0$.

When Y in column 1 of Table 23.8 is regressed on the X variables in columns 2–7 for a no-intercept regression model, we obtain $SSE(F) = 1.3000$, associated with $df_F = 14 - 6 = 8$ degrees of freedom. These results, of course, are the same as in Table 23.3a when the equivalent regression model in the factor effects form was used.

To obtain the reduced regression model under H_0, we need to incorporate the conditions in (23.37a) into the full model. We shall do this by solving the system of two equations in

TABLE 23.8 Data for Regression Fits when Treatment Means of Unequal Importance—Growth Hormone Example.

			(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
				Full Model						Reduced Model			
i	j	k	Y	X_1	X_2	X_3	X_4	X_5	X_6	Z_1	Z_2	Z_3	Z_4
1	1	1	1.4	1	0	0	0	0	0	1	0	0	0
1	1	2	2.4	1	0	0	0	0	0	1	0	0	0
	...		...	...	...	...	...	...	...	...	...	...	...
1	2	2	1.7	0	1	0	0	0	0	0	1	0	0
1	3	1	.7	0	0	1	0	0	0	0	0	1	0
	...		...	...	...	...	...	...	...	...	...	...	...
2	3	2	.9	0	0	0	0	0	1	0	2	−2	1
2	3	3	1.3	0	0	0	0	0	1	0	2	−2	1

(23.37a) for any two of the parameters and replacing these two parameters in the full model by the resulting expressions. Arbitrarily choosing μ_{21} and μ_{23}, we find in solving the two equations in (23.37a):

$$\begin{aligned} \mu_{21} &= 2\mu_{12} + \mu_{22} - 2\mu_{11} \\ \mu_{23} &= 2\mu_{12} - 2\mu_{13} + \mu_{22} \end{aligned} \tag{23.39}$$

Replacing μ_{21} and μ_{23} in full model (23.38) by the expressions in (23.39), we obtain the reduced model:

$$\begin{aligned} Y_{ijk} = {} & \mu_{11}X_{ijk1} + \mu_{12}X_{ijk2} + \mu_{13}X_{ijk3} + (2\mu_{12} + \mu_{22} - 2\mu_{11})X_{ijk4} \\ & + \mu_{22}X_{ijk5} + (2\mu_{12} - 2\mu_{13} + \mu_{22})X_{ijk6} + \varepsilon_{ijk} \end{aligned}$$

This model can be simplified algebraically, as follows:

$$Y_{ijk} = \mu_{11}Z_{ijk1} + \mu_{12}Z_{ijk2} + \mu_{13}Z_{ijk3} + \mu_{22}Z_{ijk4} + \varepsilon_{ijk} \qquad \text{Reduced model} \tag{23.40}$$

where:

$$\begin{aligned} Z_{ijk1} &= X_{ijk1} - 2X_{ijk4} \\ Z_{ijk2} &= X_{ijk2} + 2X_{ijk4} + 2X_{ijk6} \\ Z_{ijk3} &= X_{ijk3} - 2X_{ijk6} \\ Z_{ijk4} &= X_{ijk4} + X_{ijk5} + X_{ijk6} \end{aligned}$$

Table 23.8 shows the codings of the new Z variables in columns 8–11. For instance, the codings for the new Z variables associated with Y_{111} are obtained as follows:

$$X_1 = 1 \qquad X_2 = 0 \qquad X_3 = 0 \qquad X_4 = 0 \qquad X_5 = 0 \qquad X_6 = 0$$

$$\begin{aligned} Z_1 &= 1 - 2(0) = 1 \\ Z_2 &= 0 + 2(0) + 2(0) = 0 \\ Z_3 &= 0 - 2(0) = 0 \\ Z_4 &= 0 + 0 + 0 = 0 \end{aligned}$$

When Y in column 1 of Table 23.8 is regressed on the Z variables in columns 8–11 with a no-intercept regression model, we obtain $SSE(R) = 4.754$ and $df_R = 14 - 4 = 10$. Hence, the general linear test statistic (2.70) is:

$$\begin{aligned} F^* &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} \\ &= \frac{4.754 - 1.3000}{10 - 8} \div \frac{1.3000}{8} = 10.63 \end{aligned}$$

If H_0 holds, F^* follows the F distribution with 2 and 8 degrees of freedom. To control the level of significance at $\alpha = .05$, we require $F(.95; 2, 8) = 4.46$. Since $F^* = 10.63 > 4.46$, we conclude H_a, that the weighted mean change in the growth rate is not the same for the three bone development groups. The P-value of this test is .006.

ests for Factor Main Effects by Use of Matrix Formulation

We saw in the growth hormone example when using the equivalent regression models to implement the general linear test approach that it was necessary to solve a system of two equations in six unknown parameters in terms of any two of the parameters. As the number of equations in H_0 increases, the algebra can become quite tedious. Under these circumstances, it may be easier to carry out the F test when the treatment means are of unequal importance by means of formulating the general linear test in matrix terms.

The full model, as before in (23.38), is represented by:

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}}\ \underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \tag{23.41}$$

For the growth hormone example, the $\mathbf{X}$ matrix is a 14×6 matrix consisting of the columns for X_1–X_6 in Table 23.8, and the $\boldsymbol{\beta}$ vector is:

$$\underset{6\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}$$

The least squares and maximum likelihood estimators of the parameters in the full normal error model (23.41) will now be denoted by $\mathbf{b}_F$ and are, as before, given by (6.25):

$$\mathbf{b}_F = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{23.42}$$

Also, the error sum of squares is given by (6.35):

$$SSE(F) = (\mathbf{Y} - \mathbf{X}\mathbf{b}_F)'(\mathbf{Y} - \mathbf{X}\mathbf{b}_F) = \mathbf{Y}'\mathbf{Y} - \mathbf{b}_F'\mathbf{X}'\mathbf{Y} \tag{23.43}$$

A linear test hypothesis H_0 is represented in matrix form as follows:

$$H_0\colon\ \underset{s\times p}{\mathbf{C}}\ \underset{p\times 1}{\boldsymbol{\beta}} = \underset{s\times 1}{\mathbf{h}} \tag{23.44}$$

where $\mathbf{C}$ is a specified $s \times p$ matrix of rank s and $\mathbf{h}$ is a specified $s \times 1$ vector. For the growth hormone example, the hypothesis H_0 in (23.37a) can be stated in the form (23.44) with the following matrices:

$$\underset{2\times 6}{\mathbf{C}} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{3} & 0 & \frac{1}{3} & -\frac{1}{3} & 0 \\ \frac{2}{3} & 0 & -\frac{2}{3} & \frac{1}{3} & 0 & -\frac{1}{3} \end{bmatrix}$$

$$\underset{6\times 1}{\boldsymbol{\beta}} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{13} \\ \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix} \qquad \underset{2\times 1}{\mathbf{h}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Note that this formulation yields (23.37a):

$$\underset{2\times 1}{\mathbf{C}\boldsymbol{\beta}} = \begin{bmatrix} \frac{2}{3}\mu_{11} - \frac{2}{3}\mu_{12} + \frac{1}{3}\mu_{21} - \frac{1}{3}\mu_{22} \\ \frac{2}{3}\mu_{11} - \frac{2}{3}\mu_{13} + \frac{1}{3}\mu_{21} - \frac{1}{3}\mu_{23} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{h}$$

The reduced model then is:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \text{where} \qquad \mathbf{C}\boldsymbol{\beta} = \mathbf{h} \tag{23.45}$$

It can be shown that the least squares and maximum likelihood estimators under the reduced model, to be denoted by $\mathbf{b}_R$, are:

$$\mathbf{b}_R = \mathbf{b}_F - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\mathbf{b}_F - \mathbf{h}) \tag{23.46}$$

and the error sum of squares is:

$$SSE(R) = (\mathbf{Y} - \mathbf{X}\mathbf{b}_R)'(\mathbf{Y} - \mathbf{X}\mathbf{b}_R) \tag{23.47}$$

which has associated with it $df_R = n - (p - s)$ degrees of freedom. It can be shown also that the difference $SSE(R) - SSE(F)$ can be expressed as follows:

$$SSE(R) - SSE(F) = (\mathbf{C}\mathbf{b}_F - \mathbf{h})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\mathbf{b}_F - \mathbf{h}) \tag{23.48}$$

which has associated with it $df_R - df_F = (n - p + s) - (n - p) = s$ degrees of freedom.

Hence, the general linear test statistic (2.70) here is:

$$F^* = \frac{SSE(R) - SSE(F)}{s} \div \frac{SSE(F)}{n - p} \tag{23.49}$$

where $SSE(R) - SSE(F)$ is given by (23.48) and $SSE(F)$ is given by (23.43). Note for the growth hormone example that the numerator degrees of freedom are $s = 2$ and the denominator degrees of freedom are $n - p = 14 - 6 = 8$, which agree with the degrees of freedom obtained when using the equivalent regression models.

Comments

1. Many of the major statistical packages require only that the user furnish H_0 in the matrix form (23.44) and will then conduct the general linear test.

2. The least squares estimators $\mathbf{b}_R$ in (23.46) under the reduced model can be derived by minimizing the least squares criterion $Q = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$ subject to the constraint $\mathbf{C}\boldsymbol{\beta} - \mathbf{h} = \mathbf{0}$, using Lagrange multipliers.

3. The test for the alternatives (23.37a) in the growth hormone example can also be conducted by estimating the two contrasts:

$$L_1 = \frac{2\mu_{11} + \mu_{21}}{3} - \frac{2\mu_{12} + \mu_{22}}{3} \qquad L_2 = \frac{2\mu_{11} + \mu_{21}}{3} - \frac{2\mu_{13} + \mu_{23}}{3}$$

with a multiple comparison procedure (e.g., the Bonferroni procedure) and noting whether or not both confidence intervals include zero. ■

Tests for Factor Effects when Weights Are Proportional to Sample Sizes

Simplifications in determining the term $SSE(R) - SSE(F)$ in the general linear test statistic for testing weighted factor A and factor B effects occur when the weights n_{ij} for the means μ_{ij} are proportional to the total sample sizes $n_{i\cdot}$ and $n_{\cdot j}$ for factor A and factor B levels, respectively. Such weights are appropriate in some circumstances but not in many others.

Consider a study of retail stores. The effects on shoplifting losses of size of store (factor A) and location of store within the city (factor B) are to be studied. Inferences about all retail stores in the population of interest are to be made. A random sample of n_T retail stores is selected from the population of all stores, and the selected stores are then classified by size and location. We denote the resulting cell sample sizes as usual by n_{ij}. If the proportions of stores in the different size-location groups in the population were known, these known proportions would serve as the appropriate weights in making inferences about size and location main effects, and the general linear test procedures just discussed would be employed. Often, however, these proportions are not known. Under these conditions, the cell sample sizes n_{ij} may be used to estimate the unknown proportions and therefore may serve as reasonable weights.

To illustrate this, suppose that $a = 2$ store sizes and $b = 3$ locations are employed in the study of retail stores, and that a random sample of $n_T = 60$ stores resulted in the following cell sample sizes n_{ij}:

Store Size	Location (j)			
i	$j = 1$	$j = 2$	$j = 3$	Total
$i = 1$	20	5	4	29
$i = 2$	10	15	6	31
Total	30	20	10	60

Thus $n_{11} = 20$, $n_{21} = 10$, and so on. Further, denoting by $n_{i\cdot}$ and $n_{\cdot j}$ the total factor A and factor B level sample sizes as defined in (23.1a) and (23.1b), respectively, we have $n_{1\cdot} = 29$, $n_{\cdot 1} = 30$, and so on.

The test for comparing factor A effects, when the weights $n_{ij}/n_{i\cdot}$ reflect the importance of the factor A means, would then involve a comparison of the weighted mean for factor A level $i = 1$:

$$\frac{20\mu_{11} + 5\mu_{12} + 4\mu_{13}}{29}$$

and the weighted mean for factor A level $i = 2$:

$$\frac{10\mu_{21} + 15\mu_{22} + 6\mu_{23}}{31}$$

Expressed in symbolic notation, the alternatives would be:

$$H_0\text{: } \left(\frac{n_{11}}{n_{1\cdot}}\right)\mu_{11} + \left(\frac{n_{12}}{n_{1\cdot}}\right)\mu_{12} + \left(\frac{n_{13}}{n_{1\cdot}}\right)\mu_{13} = \left(\frac{n_{21}}{n_{2\cdot}}\right)\mu_{21} + \left(\frac{n_{22}}{n_{2\cdot}}\right)\mu_{22} + \left(\frac{n_{23}}{n_{2\cdot}}\right)\mu_{23}$$

H_a: equality does not hold

Similarly, the alternatives for testing weighted factor B effects would be as follows when weights $n_{ij}/n_{\cdot j}$ reflect the importance of the factor B means:

$$H_0: \left(\frac{n_{11}}{n_{\cdot 1}}\right)\mu_{11} + \left(\frac{n_{21}}{n_{\cdot 1}}\right)\mu_{21} = \left(\frac{n_{12}}{n_{\cdot 2}}\right)\mu_{12} + \left(\frac{n_{22}}{n_{\cdot 2}}\right)\mu_{22} = \left(\frac{n_{13}}{n_{\cdot 3}}\right)\mu_{13} + \left(\frac{n_{23}}{n_{\cdot 3}}\right)\mu_{23}$$

H_a: not all equalities hold

We must caution that sample sizes often do not reflect appropriate importance. Sample sizes may have been chosen arbitrarily or they may reflect unequal attrition losses in a study. Sample sizes may also reflect cost considerations; for instance, larger sample sizes may be used by a market researcher for children than for adults because selection costs are lower. In all of these instances, use of weights based on sample sizes may lead to misleading inferences.

When sample sizes do constitute appropriate weights, the alternatives for testing for weighted factor A effects can be stated in general as follows:

$$H_0: \sum_j \left(\frac{n_{1j}}{n_{1\cdot}}\right)\mu_{1j} = \cdots = \sum_j \left(\frac{n_{aj}}{n_{a\cdot}}\right)\mu_{aj} \tag{23.50}$$

H_a: not all equalities hold

and the alternatives for testing for weighted factor B effects are:

$$H_0: \sum_i \left(\frac{n_{i1}}{n_{\cdot 1}}\right)\mu_{i1} = \cdots = \sum_i \left(\frac{n_{ib}}{n_{\cdot b}}\right)\mu_{ib} \tag{23.51}$$

H_a: not all equalities hold

It can be shown that the term $SSE(R) - SSE(F)$ for testing weighted factor A effects involving the alternatives in (23.50) simplifies to the ordinary single-factor treatment sum of squares in (16.28), with the factor A levels considered to be the treatments:

$$SSA = \sum_i n_{i\cdot}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 \tag{23.52}$$

where:

$$\bar{Y}_{i\cdot\cdot} = \frac{Y_{i\cdot\cdot}}{n_{i\cdot}} \tag{23.52a}$$

$$\bar{Y}_{\cdot\cdot\cdot} = \frac{Y_{\cdot\cdot\cdot}}{n_T} \tag{23.52b}$$

$$Y_{i\cdot\cdot} = \sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} Y_{ijk} \tag{23.52c}$$

$$Y_{\cdot\cdot\cdot} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n_{ij}} Y_{ijk} \tag{23.52d}$$

Similarly, the term $SSE(R) - SSE(F)$ for testing weighted factor B effects involving the alternatives in (23.44) simplifies to the single-factor treatment sum of squares in (16.28), with the factor B levels considered to be the treatments:

$$SSB = \sum_j n_{\cdot j}(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 \tag{23.53}$$

where:

$$\bar{Y}_{\cdot j\cdot} = \frac{Y_{\cdot j\cdot}}{n_{\cdot j}} \tag{23.53a}$$

$$Y_{\cdot j\cdot} = \sum_{i=1}^{a}\sum_{k=1}^{n_{ij}} Y_{ijk} \tag{23.53b}$$

Example

In the growth hormone example of Table 23.1, suppose that the treatment sample sizes n_{ij} reflect the relative importance of the factor means. We saw in Section 23.2 that gender (factor A) and bone development (factor B) do not interact. We now wish to test whether gender affects the weighted mean change in the growth rate. The alternatives (23.50) here are:

$$H_0\colon \frac{3}{7}\mu_{11} + \frac{2}{7}\mu_{12} + \frac{2}{7}\mu_{13} = \frac{1}{7}\mu_{21} + \frac{3}{7}\mu_{22} + \frac{3}{7}\mu_{23}$$

H_a: equality does not hold

To calculate SSA in (23.52), we require from Table 23.1:

$$\begin{array}{lll} Y_{1\cdot\cdot} = 11.6 & n_{1\cdot} = 7 & \bar{Y}_{1\cdot\cdot} = 1.65714 \\ Y_{2\cdot\cdot} = 11.4 & n_{2\cdot} = 7 & \bar{Y}_{2\cdot\cdot} = 1.62857 \\ Y_{\cdot\cdot\cdot} = 23.0 & n_T = 14 & \bar{Y}_{\cdot\cdot\cdot} = 1.64286 \end{array}$$

We then obtain:

$$SSA = 7(1.65714 - 1.64286)^2 + 7(1.62857 - 1.64286)^2 = .002857$$

The number of degrees of freedom associated with SSA is $a - 1 = 2 - 1 = 1$.

We found earlier in Table 23.3a that the error sum of squares for the full model is $SSE(F) = 1.3000$, with 8 degrees of freedom associated with it. Hence, the general linear test statistic here is:

$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} \div \frac{SSE(F)}{df_F} = \frac{SSA}{1} \div MSE(F)$$

$$= \frac{.002857}{1} \div \frac{1.3000}{8} = .018$$

For $\alpha = .05$, we require $F(.95; 1, 8) = 5.32$. Since $F^* = .018 \leq 5.32$, we conclude H_0, that the weighted mean change in the growth rate is the same for male and female children. The P-value of the test is .897.

The test for factor B effects would be carried out in similar fashion.

Comments

1. A special case of weights proportional to the sample sizes occurs in designed experiments when the sample sizes themselves follow a proportional pattern. Suppose that a chain of diet establishments is experimenting with two diets that are of equal importance. The establishments cater to three times as many women as men. One hundred men and 300 women are selected, and half of each group is randomly assigned to each diet. Hence, the treatment sample sizes are as follows:

Diet	Men	Women	Total
1	50	150	200
2	50	150	200
Total	100	300	400

Note that these treatment sample sizes follow the relation:

$$n_{ij} = \frac{n_{i\cdot}n_{\cdot j}}{n_T} \tag{23.54}$$

Condition (23.54) implies that the sample sizes in any two rows (or columns) are proportional. This is called a case of *proportional frequencies*. Here the test of diet effects reduces to the comparison of $(\mu_{11} + 3\mu_{12})/4$ versus $(\mu_{21} + 3\mu_{22})/4$ and the test of gender effects reduces to the comparison of $(\mu_{11}+\mu_{21})/2$ versus $(\mu_{12}+\mu_{22})/2$. It can be shown that the terms $SSE(R)-SSE(F)$ for testing these factor A effects (diet) and factor B effects (gender) are given by (23.52) and (23.53), respectively. It can also be shown that the interaction sum of squares here is given by a simple formula:

$$SSAB = \sum_i \sum_j n_{ij}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2 \tag{23.55}$$

Furthermore, the sums of squares in this special case are orthogonal so that *SSA*, *SSB*, *SSAB*, and *SSE* sum to *SSTO*.

2. When proportional sample sizes are employed but the sample sizes do not reflect the importance of the factor level means (e.g., when the sample sizes are unequal but the factor level means are of equal importance), the regression approach or the general linear test approach explained earlier must be employed.

3. The cell sample sizes in alternatives (23.50) and (23.51) are considered to be fixed, not random variables. Thus, the relevance of the alternatives depends on the reasonableness of the actual cell sample sizes as indicators of the importance of the treatment means. ■

23.6 Statistical Computing Packages

Extreme care must be exercised when using packaged analysis of variance programs with unequal sample sizes because the default option of the package may not necessarily assign proper importance to each treatment mean. The user should read the package documentation carefully and make sure that the package generates the appropriate sums of squares for the tests of interest.

For the JMP, MINITAB, SAS, SPSS, and SYSTAT statistical packages, the outputs that are the equivalents of the regression results obtained in Sections 23.1–23.3 for the case of treatment means with equal importance and no empty cells are obtained as follows at the time of this writing:

JMP—Fit Model
MINITAB—GLM
SAS PROC GLM—Type III or Type IV sums of squares
SPSS GLM—UNIANOVA/SSTYPE(3)
SYSTAT—Default option

Extreme caution should also be used with ANOVA computer packages that provide results when some treatment cells are empty. The package may make assumptions about interactions that the researcher is unwilling to make. In the absence of a clear description of how the package handles empty cells, it is preferable that appropriate analyses be conducted by the user specifying the appropriate contrasts of interest.

When weights assigned to the treatment means are proportional to the sample sizes, numerator sums of squares *SSA* and *SSB* given in (23.52) and (23.53) may be obtained using JMP Sequential (Type 1) Tests option, MINITAB Sequential SS option, SAS PROC GLM—Type I sum of squares, SPSS GLM—UNIANOVA/SSTYPE(1), and SYSTAT—Option Weighted Means Model. When a sequential Type I sum of squares is used to obtain *SSA* and *SSB* given in (23.52) and (23.53), two separate computing runs are needed, where in one run factor A is brought in first and in the second run factor B is brought in first.

A simple option in using computer packages when the cell sample sizes are unequal, cell means have unequal importance, and/or some cells are empty is to use a single-factor ANOVA package that permits specification of contrasts to be estimated. The user can then specify the various contrasts of interest.

Problems

23.1. A market research intern selected a random sample of 400 communities and classified them according to population size (four levels) and geographic region (five levels) to study the effects of these factors on sales of the company's products. When the intern found that the treatment sample sizes were unequal, the smallest cell frequency being four, the intern generated random numbers to reduce the number of communities in each cell to four and then proceeded to analyze the effects of population size and region on the basis of the 80 communities remaining.

a. Does the method of randomly discarding cases lead to any biases? Explain.

b. Was it wise for the intern to discard 320 cases randomly in order to obtain equal treatment sample sizes?

23.2. A student asked: "If two-factor studies with unequal sample sizes must be analyzed by a regression approach, why bother with the two-factor analysis of variance model at all?" Comment.

23.3. Refer to **Eye contact effect** Problems 19.12 and 19.13.

a. Modify regression model (23.11) to apply to this two-factor study with $a = 2$ and $b = 2$.

b. Set up the $\mathbf{Y}$, $\mathbf{X}$, and $\boldsymbol{\beta}$ matrices for the regression model in part (a).

c. Obtain $\mathbf{X}\boldsymbol{\beta}$. Verify the correctness of the expected values.

d. Obtain the fitted regression function. What is estimated by the intercept term?

e. Obtain the regression analysis of variance table based on appropriate extra sums of squares. Do your results agree with those obtained using the ANOVA approach in 19.13b?

f. Test separately for interaction effects, factor A main effects, and factor B main effects. Use $\alpha = .01$ for each test and state the alternatives, decision rule, and conclusion.

*23.4. Refer to **Hay fever relief** Problems 19.14 and 19.15.

a. Modify regression model (23.11) to apply to this two-factor study with $a = 3$ and $b = 3$.

b. Set up the $\mathbf{Y}$, $\mathbf{X}$, and $\boldsymbol{\beta}$ matrices for the regression model in part (a).

c. Obtain $\mathbf{X}\boldsymbol{\beta}$. Verify the correctness of the expected values.

d. Obtain the fitted regression function. What is estimated by $\hat{\alpha}_1$?

e. Obtain the regression analysis of variance table based on appropriate extra sums of squares. Do your results agree with those obtained using the ANOVA approach in Problem 19.15b?

f. Test separately for interaction effects, factor A main effects, and factor B main effects. Use $\alpha = .05$ for each test and state the alternatives, decision rule, and conclusion.

23.5. Refer to **Disk drive service** Problems 19.16 and 19.17.

a. Modify regression model (23.11) to apply to this two-factor study with $a = 3$ and $b = 3$.

b. Obtain the fitted regression function. What is estimated by $\hat{\beta}_1$?

c. Obtain the regression analysis of variance table based on appropriate extra sums of squares. Do your results agree with those obtained using the ANOVA approach in 19.17b?

d. Test separately for interaction effects, factor A main effects, and factor B main effects. Use $\alpha = .01$ for each test and state the alternatives, decision rule, and conclusion.

*23.6. Refer to **Cash offers** Problem 19.10. Suppose that observations $Y_{214} = 28$ and $Y_{323} = 20$ are missing because the offer received in each of these cases was a trade-in offer, not a cash offer.

a. State the ANOVA model for this case. Also state the equivalent regression model; use 1, −1, 0 indicator variables.

b. Present the $\mathbf{X}$ and $\boldsymbol{\beta}$ matrices for the regression model in part (a).

c. Obtain $\mathbf{X}\boldsymbol{\beta}$ and show that the proper treatment means are obtained by your model.

d. What is the reduced regression model for testing for interaction effects?

e. Test whether or not interaction effects are present by fitting the full and reduced regression models; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. State the reduced regression models for testing for age and gender main effects, respectively, and conduct each of the tests. Use $\alpha = .05$ each time and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

g. To study the nature of the age main effects, estimate the following pairwise comparisons:

$$D_1 = \mu_{1\cdot} - \mu_{2\cdot} \qquad D_2 = \mu_{1\cdot} - \mu_{3\cdot} \qquad D_3 = \mu_{2\cdot} - \mu_{3\cdot}$$

Use the most efficient multiple comparison procedure with a 90 percent family confidence coefficient.

h. In the population of female owners, 30 percent are young, 60 percent are middle-aged, and 10 percent are elderly. Estimate the mean cash offer for this population with a 95 percent confidence interval.

*23.7. Refer to **Hay fever relief** Problem 19.14 and 23.4. Suppose that observations $Y_{113} = 2.3$, $Y_{221} = 8.9$, and $Y_{224} = 9.0$ are missing because the subjects did not immediately record the time when they began to suffer again from hay fever.

a. State the ANOVA model for this case. Also state the equivalent regression model; use 1, −1, 0 indicator variables.

b. Present the $\mathbf{X}$ and $\boldsymbol{\beta}$ matrices for the regression model in part (a).

c. Obtain $\mathbf{X}\boldsymbol{\beta}$ and show that the proper treatment means are obtained by your model.

d. What is the reduced regression model for testing for interaction effects?

e. Test whether or not interaction effects are present by fitting the full and reduced regression models; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test? How do your results compare with those obtained in 23.4f, where there is no missing data?

f. The nature of the interaction effects is to be studied by means of the following contrasts:

$$L_1 = \frac{\mu_{12} + \mu_{13}}{2} - \mu_{11} \qquad L_4 = L_2 - L_1$$

$$L_2 = \frac{\mu_{22} + \mu_{23}}{2} - \mu_{21} \qquad L_5 = L_3 - L_1$$

$$L_3 = \frac{\mu_{32} + \mu_{33}}{2} - \mu_{31} \qquad L_6 = L_3 - L_2$$

Obtain confidence intervals for these contrasts; use the Scheffé multiple comparison procedure with a 90 percent family confidence coefficient. Interpret your findings.

23.8. Refer to **Kidney failure hospitalization** Problem 19.18. Suppose that observations $Y_{124} = 12$, $Y_{216} = 2$, and $Y_{238} = 9$ are missing because the hospitalization records for these patients were not complete. Continue to work with the transformed data $Y' = \log_{10}(Y + 1)$.

a. State the ANOVA model for this case. Also state the equivalent regression model; use 1, −1, 0 indicator variables.

b. Present the $\mathbf{X}$ and $\boldsymbol{\beta}$ matrices for the regression model in part (a).

c. Obtain $\mathbf{X}\boldsymbol{\beta}$ and show that the proper treatment means are obtained by your model.

d. What is the reduced regression model for testing for interaction effects?

e. Test whether or not interaction effects are present by fitting the full and reduced regression models; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. State the reduced regression models for testing for treatment duration and weight gain main effects, respectively. Conduct each of the tests. Use $\alpha = .05$ each time and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

g. Use the single degree of freedom t^* statistic for testing whether or not the mean number of days hospitalized (in transformed units) for persons with mild weight gains exceeds .5; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

h. To analyze the nature of the factor main effects, estimate the following pairwise comparisons:

$$D_1 = \mu_{1\cdot} - \mu_{2\cdot} \qquad D_3 = \mu_{\cdot 3} - \mu_{\cdot 1}$$

$$D_2 = \mu_{\cdot 2} - \mu_{\cdot 1} \qquad D_4 = \mu_{\cdot 3} - \mu_{\cdot 2}$$

Use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

23.9. **Adjunct professors.** A sociologist selected a random sample of 45 adjunct professors who teach in the evening division of a large metropolitan university for a study of special problems associated with teaching in the evening division. The data collected include the amount of

payment received by the faculty member for teaching a course during the past semester. The sociologist classified the faculty members by subject matter of course (factor A) and highest degree earned (factor B). The earnings per course (in thousand dollars) follow.

Factor A (subject matter)	Factor B (highest degree) $j = 1$ Bachelor's	$j = 2$ Master's	$j = 3$ Doctorate
$i = 1$ Humanities	1.7	1.8	2.5
	1.9	2.1	2.7
			...
			2.9
$i = 2$ Social sciences	2.5	2.7	3.5
	2.3	2.4	3.3
	.	...	...
	2.4	2.5	3.4
$i = 3$ Engineering	2.7	2.9	3.7
	2.8	3.0	3.6
		...	...
		2.7	3.9
$i = 4$ Management	2.5	2.3	3.3
	2.6	2.8	3.4
			...
			3.6

a. State the ANOVA model for this case. Also state the equivalent regression model; use 1, -1, 0 indicator variables.

b. Present the $\mathbf{X}$ and $\boldsymbol{\beta}$ matrices for the regression model in part (a).

c. Obtain $\mathbf{X}\boldsymbol{\beta}$ and show that the proper treatment means are obtained by your model.

d. Fit the equivalent regression model and obtain the residuals. Prepare aligned residual dot plots for the treatments. What are your findings?

e. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

23.10. Refer to **Adjunct professors** Problem 23.9. Assume that ANOVA model (19.23), is appropriate, except that now $k = 1, \ldots, n_{ij}$.

a. Plot the estimated treatment means $\bar{Y}_{ij}$. in the format of Figure 23.1. Does it appear that any factor effects are present? Explain.

b. What is the reduced regression model for testing for interaction effects?

c. Test whether or not interaction effects are present by fitting the full and reduced regression models; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. State the reduced regression models for testing for subject matter and highest degree main effects, respectively, and conduct each of the tests. Use $\alpha = .01$ each time and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

e. Make all pairwise comparisons between the subject matter means; use the Tukey procedure with a 95 percent family confidence coefficient. State your findings and present a graphic summary.

f. Make all pairwise comparisons between the highest degree means; use the Tukey procedure with a 95 percent family confidence coefficient. State your findings and present a graphic summary.

23.11. Refer to **Adjunct professors** Problem 23.9. Suppose that the sociologist had prior information indicating that the two factors do not interact and that no-interaction model (23.28) is therefore appropriate.

a. State the equivalent full regression model for this case. Also state the reduced regression models for testing for factor A and factor B main effects. Use 1, −1, 0 indicator variables.

b. Fit the full and reduced regression models and test for factor A and factor B main effects; use $\alpha = .05$ for each test. State the alternatives, decision rule, and conclusion for each test. What is the P-value of each test?

*23.12. Refer to **Hay fever relief** Problem 19.14. Suppose that the data for the treatment when each of the two active ingredients is at the medium level were lost and immediate analyses of the available data are required; i.e., assume that $n_T = 32$ and $n_{22} = 0$.

a. To study whether or not interaction effects are present, estimate the following comparisons:

$$D_1 = \mu_{13} - \mu_{11} \qquad L_1 = D_1 - D_2$$
$$D_2 = \mu_{23} - \mu_{21} \qquad L_2 = D_1 - D_3$$
$$D_3 = \mu_{33} - \mu_{31}$$

Use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

b. To further explore the nature of possible interaction effects, conduct separate single degree of freedom tests of whether $\mu_{12} = \mu_{13}$ and whether $\mu_{32} = \mu_{33}$. Use $\alpha = .02$ for each test and state the alternatives, decision rule, and conclusion. What is the family level of significance, using the Bonferroni inequality?

23.13. Refer to **Kidney failure hospitalization** Problem 19.18. Suppose that there were no patients who received the dialysis treatment for long duration and had mild weight gains; i.e., assume that $n_T = 50$ and $n_{21} = 0$. Continue to work with the transformed data $Y' = \log_{10}(Y + 1)$.

On the basis of related research, the analyst believes it is reasonable to assume that the two factors do not interact and that no-interaction model (23.28) is appropriate.

a. State the equivalent full regression model for this case. Also state the reduced regression models for testing for factor A and factor B main effects. Use 1, −1, 0 indicator variables in the regression model.

b. Fit the full and reduced regression models. Test for factor A and factor B main effects; use $\alpha = .05$ for each test. State the alternatives, decision rule, and conclusion for each test. What is the P-value of each test?

*23.14. Refer to **Programmer requirements** Problem 19.20. Suppose that there were no programmers with experience on both small and large systems who had less than five years' experience; i.e., assume that $n_T = 20$ and $n_{21} = 0$.

a. To study whether or not interaction effects are present, estimate the following comparisons:

$$D_1 = \mu_{12} - \mu_{13} \qquad L_1 = D_1 - D_2$$
$$D_2 = \mu_{22} - \mu_{23}$$

Use the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

b. To study further the nature of possible interaction effects, test whether or not μ_{22} exceeds μ_{23}; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

23.15. Refer to **Adjunct professors** Problem 23.9. Suppose that there were no professors teaching humanities courses who had only a bachelor's degree, so that the study consists of $n_T = 43$ adjunct professors and $n_{11} = 0$. On the basis of previous research, the sociologist believes it is reasonable to assume that the two factors do not interact and that no-interaction model (23.28) is appropriate here.

a. State the equivalent full regression model for this case. Also state the reduced regression models for testing for factor A and factor B main effects. Use 1, −1, 0 indicator variables in the regression model.

b. Fit the full and reduced regression models and test for factor A and factor B main effects. Use $\alpha = .01$ for each test and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

*23.16. Refer to **Auditor training** Problem 21.5.

a. State the regression model equivalent to randomized block model (21.1); use 1, −1, 0 indicator variables.

b. Fit the regression model to the data.

c. Obtain the regression analysis of variance table based on appropriate extra sums of squares.

d. Test for treatment main effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

23.17. Refer to **Fat in diets** Problem 21.7.

a. State the regression model equivalent to randomized block model (21.1); use 1, −1, 0 indicator variables.

b. Fit the regression model to the data.

c. Obtain the regression analysis of variance table based on appropriate extra sums of squares.

d. Test for treatment main effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

*23.18. Refer to **Auditor training** Problems 21.5 and 23.16. Assume that observation $Y_{23} = 89$ is missing because the auditor became ill and dropped out from the study.

a. State the ANOVA model for this case. Also state the equivalent regression model; use 1, −1, 0 indicator variables.

b. State the reduced regression model for testing for differences in the mean proficiency scores for the three training methods.

c. Test whether or not the mean proficiency scores for the three training methods differ by fitting the full and reduced models; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. How do your results compare with those obtained in Problem 23.16d, where there are no missing observations?

d. Compare the mean proficiency scores for training methods 2 and 3 by means of the regression approach; use a 95 percent confidence interval.

23.19. Refer to **Fat in diets** Problems 21.7 and 23.17. Assume that observations $Y_{13} = .15$ and $Y_{51} = 1.62$ are missing because the subjects did not stay on the prescribed diet.

a. State the ANOVA model for this case. Also state the equivalent regression model; use 1, −1, 0 indicator variables.

b. State the reduced regression model for testing for differences in the mean reductions in lipid level for the three diets.

c. Test whether or not the mean reductions in lipid level differ for the three diets by fitting the full and reduced models; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. How do your results compare with those obtained in Problem 23.17d, where there are no missing observations?

d. Compare the mean reductions in lipid level for diets 1 and 3 by means of the regression approach; use a 98 percent confidence interval.

*23.20. Refer to **Cash offers** Problem 19.10. It is known that in both populations of male and female owners, 30 percent are young, 60 percent are middle-aged, and 10 percent are elderly. Test by means of the single degree of freedom t^* test statistic whether or not the mean cash offers for male and female owners are equal; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

23.21. Refer to **Kidney failure hospitalization** Problem 19.18. Continue to work with the transformed data $Y' = \log_{10}(Y + 1)$. It is known that 75 percent of patients in each weight gain group receive the short duration treatment. Inferences are desired about the target population of patients at the dialysis facility.

a. Use cell means model (19.15) to express the two alternatives for testing whether or not factor B main effects are present in the form of (23.37a).

b. State the regression model equivalent to ANOVA model (19.15), using 1, 0 indicator variables.

c. State the reduced regression model for testing for factor B main effects; express μ_{11} and μ_{13} in terms of the other cell means.

d. Fit the full and reduced regression models and test for factor B main effects; use $\alpha = .05$. State the decision rule and conclusion. What is the P-value of the test?

e. Compare the mean number of days of hospitalization (in transformed units) for patients with severe and mild weight gains; use a 95 percent confidence interval.

23.22. Refer to **Adjunct professors** Problem 23.9. It is known that 10 percent of professors in each subject matter area have a bachelor's degree, 20 percent have a master's degree, and 70 percent have a doctorate. Inferences are desired about the target population of adjunct professors.

a. Use cell means model (19.15) to express the two alternatives for testing whether or not factor A main effects are present in the form of (23.37a).

b. Define the $\mathbf{X}$ matrix and $\boldsymbol{\beta}$ vector for expressing full model (19.15) in matrix form for this case.

c. Express the two alternatives in part (a) in matrix form (23.44).

d. Use (23.48) to calculate $SSE(R) - SSE(F)$.

e. Test whether or not factor A main effects are present; use $\alpha = .01$. State the decision rule and conclusion. What is the P-value of the test?

f. Compare the mean amounts of payment received by faculty members teaching humanities and engineering courses; use a 99 percent confidence interval. Interpret your interval estimate.

*23.23. Refer to **Programmer requirements** Problem 19.20. Suppose that the observations $Y_{133} = 68$, $Y_{134} = 58$, and $Y_{234} = 45$ did not exist and that the sample sizes reflect the importance of the treatment means. Test whether or not type of experience main effects are present; control the level of significance at $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

23.24. Refer to **Adjunct professors** Problem 23.9. Assume that the sample sizes reflect the importance of the treatment means. Test whether or not subject matter main effects are present; control the level of significance at $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Exercises

23.25. Derive $\sigma^2\{\hat{L}\}$ for the estimated contrast involving $\hat{\mu}_{i\cdot}$ in (23.22).

23.26. Show that $s^2\{\hat{L}\}$ in (23.26) is an unbiased estimator of $\sigma^2\{\hat{L}\}$.

23.27. Refer to regression model (23.31), the equivalent to ANOVA model (21.1) when $n_b = 3$ and $r = 3$. Suppose that the indicator variables in model (23.31) were coded as follows:

$$X_1 = \begin{cases} 1 & \text{if experimental unit from block 1} \\ 0 & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if experimental unit from block 2} \\ 0 & \text{otherwise} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{if experimental unit from treatment 1} \\ 0 & \text{otherwise} \end{cases}$$

$$X_4 = \begin{cases} 1 & \text{if experimental unit from treatment 2} \\ 0 & \text{otherwise} \end{cases}$$

and that the regression coefficients are denoted by β_0, β_1, β_2, β_3, and β_4.

a. Exhibit the **X** matrix for this regression model.

b. Find the correspondences between the regression coefficients $\beta_0, \beta_1, \ldots, \beta_4$ and the parameters in ANOVA model (21.1).

c. Discuss the advantages and disadvantages of using 1, 0 indicator variables and 1, −1, 0 indicator variables here.

23.28. Consider a two-factor study where $a = 2$, $b = 2$, $n_{11} = n_{12} = n_{21} = 2$, $n_{22} = 1$, and no-interaction model (23.28) applies. Use the matrix methods in Section 23.5 to obtain the estimator of μ_{22}. [*Hint:* Begin with interaction model (23.3) as the full model, express the assumption of no interactions in the form of (23.44), and use (23.46) to obtain the estimator of μ_{22} for the no-interaction model.]

23.29. Refer to **Kidney failure hospitalization** Problem 23.13. Suppose that you are going to use the matrix approach in Section 23.5, rather than the regression approach, to test for factor A main effects.

a. State the **X** and $\boldsymbol{\beta}$ matrices to be used in the full model.

b. State the test hypothesis in matrix form (23.44).

Projects

23.30. Refer to the **SENIC** data set in Appendix C.1. The effects of region (factor A: variable 9) and average age of patients (factor B: variable 3) on mean length of hospital stay (variable 2) are to be studied. For purposes of this ANOVA study, average age is to be classified into three categories: under 52.0 years, 52.0–under 55.0 years, 55.0 years or more.

a. State the ANOVA model for this case. Also state the equivalent regression model; use 1, −1, 0 indicator variables.

b. Fit the regression model, obtain the residuals, and prepare aligned residual dot plots for the treatments. What are your findings?

c. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

23.31. Refer to the **SENIC** data set in Appendix C.1 and Project 23.30. Assume that ANOVA model (19.23), with $k = 1, \ldots, n_{ij}$, is appropriate.

a. Plot the estimated treatment means $\bar{Y}_{ij\cdot}$ in the format of Figure 23.1. Does it appear that any factor effects are present? Explain.

b. State the reduced regression model for testing for interaction effects.

c. Fit the reduced regression model and test whether or not interaction effects are present; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. State the reduced regression model for testing for factor A main effects. Conduct this test using $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. State the reduced regression model for testing for factor B main effects. Conduct this test using $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. Make all pairwise comparisons between regions; use the Tukey procedure and a 95 percent family confidence coefficient. State your findings and present a graphic summary.

23.32. Refer to the **CDI** data set in Appendix C.2. The effects of region (factor A: variable 17) and percent below poverty level (factor B: variable 13) on the crime rate (variable 10 ÷ variable 5) are to be studied. For purposes of this ANOVA study, percent below poverty level is to be classified into three categories: under 6.0 percent, 6.0–under 10.0 percent, 10.0 percent or more.

a. State the ANOVA model for this case. Also state the equivalent regression model; use $1, -1, 0$ indicator variables.

b. Fit the regression model, obtain the residuals, and prepare aligned residual dot plots for the treatments. What are your findings?

c. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

23.33. Refer to the **CDI** data set in Appendix C.2 and Project 23.32. Assume that ANOVA model (19.23), with $k = 1, \ldots, n_{ij}$, is appropriate.

a. Plot the estimated treatment means $\bar{Y}_{ij\cdot}$ in the format of Figure 23.1. Does it appear that any factor effects are present? Explain.

b. State the reduced regression model for testing for interaction effects.

c. Fit the reduced regression model and test whether or not interaction effects are present; use $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. State the reduced regression model for testing for factor A main effects. Conduct this test using $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. State the reduced regression model for testing for factor B main effects. Conduct this test using $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. Make all pairwise comparisons between regions: use the Tukey procedure and a 95 percent family confidence coefficient. State your findings and present a graphic summary.

23.34. Refer to the **Market share** data set in Appendix C.3. The effects of discount price (factor A: variable 5) and package promotion (factor B: variable 6) on market share (variable 2) are to be studied.

a. State the ANOVA model for this case. Also state the equivalent regression model; use 1, −1, 0 indicator variables.

b. Fit the regression model, obtain the residuals, and prepare aligned residual dot plots for the treatments. What are your findings?

c. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

23.35. Refer to the **Market share** data set in Appendix C.3 and Project 23.34. Assume that ANOVA model (19.23) with $k = 1, \ldots, n_{ij}$ is appropriate.

a. Plot the estimated treatment means $\overline{Y}_{ij.}$ in the format of Figure 23.1. Does it appear that any factor effects are present? Explain.

b. State the reduced model for testing for interaction effects.

c. Fit the reduced regression model and test whether or not interaction effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. State the reduced regression model for testing for factor A main effects. Conduct this test using $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. State the reduced regression model for testing for factor B main effects. Conduct this test using $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

23.36. Refer to the **SENIC** data set in Appendix C.1 and Projects 23.30 and 23.31. Assume that the sample sizes reflect the importance of the treatment means.

a. Test for region (factor A) main effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Test for average age of patients (factor B) main effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

23.37. Refer to the **CDI** data set in Appendix C.2 and Projects 23.32 and 23.33. Assume that the sample sizes reflect the importance of the treatment means.

a. Test for region (factor A) main effects; use $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Test for percent below poverty level (factor B) main effects; use $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

Case Studies

23.38. Refer to the **Prostate cancer** data set in Appendix C.5. Assume that the sample sizes do not reflect the importance of the treatment means. Carry out an unbalanced two-way analysis of variance of this data set, where the response of interest is PSA level (variable 2), the two crossed factors are Gleason score (variable 9) and seminal vesicle invasion (variable 7). The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

23.39. Refer to the **Prostate cancer** data set in Appendix C.5 and Case Study 23.38. Assume that the sample sizes reflect the importance of the treatment means. Carry out an unbalanced two-way analysis of variance of this data set, where the response of interest is PSA level (variable 2), the two crossed factors are Gleason score (variable 9) and seminal vesicle invasion (variable 7). The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

23.40. Refer to the **Real estate sales** data set in Appendix C.7. Assume that the sample sizes do not reflect the importance of the treatment means. Carry out an unbalanced two-way analysis of variance of this data set, where the response of interest is sales price (variable 2), the two crossed factors are quality (variable 10) and style (variable 11). Recode style as 1 or not 1. The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

23.41. Refer to the **Real estate sales** data set in Appendix C.7 and Case Study 23.40. Assume that the sample sizes reflect the importance of the treatment means. Carry out an unbalanced two-way analysis of variance of this data set, where the response of interest is sales price (variable 2), the two crossed factors are quality (variable 10) and style (variable 11). Recode style as 1 or not 1. The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

23.42. Refer to the **Ischemic heart disease** data set in Appendix C.9. Assume that the sample sizes do not reflect the importance of the treatment means. Carry out an unbalanced two-way analysis of variance of this data set, where the response of interest is total cost (variable 2), the two crossed factors are number of interventions (variable 5) and number of comorbidities (variable 9). Recode the number of interventions into six categories: 0, 1, 2, 3–4, 5–7, and greater than or equal to 8. Recode the number of comorbidities into two categories: 0–1, and greater than or equal to 2. The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

23.43. Refer to the **Ischemic heart disease** data set in Appendix C.9 and Case Study 23.42. Assume that the sample sizes reflect the importance of the treatment means. Carry out an unbalanced two-way analysis of variance of this data set, where the response of interest is total cost (variable 2), the two crossed factors are number of interventions (variable 5) and number of comorbidities (variable 9). Recode the number of interventions into six categories: 0, 1, 2, 3–4, 5–7, and greater than or equal to 8. Recode the number of comorbidities into two categories: 0–1, and greater than or equal to 2. The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

Chapter 24

Multi-Factor Studies

When three or more factors are studied simultaneously, the model and analysis employed are straightforward extensions of the two-factor case. We shall illustrate the nature of the extensions with reference to three-factor studies. Ordinarily, computer ANOVA packages will be utilized for performing the needed calculations for multi-factor studies involving three or more factors. For completeness, however, we shall present the necessary definitional formulas for three-factor studies. The ANOVA model with fixed factor levels *when all treatment sample sizes are equal and all treatment means are of equal importance* is considered in Sections 24.1–24.5. Then the analysis of variance with unequal sample sizes is taken up in Section 24.6. The chapter concludes with the planning of sample sizes for multi-factor studies.

24.1 ANOVA Model for Three-Factor Studies

We now turn to the development of the ANOVA model with fixed factor levels for three-factor studies. This ANOVA model will be applicable to observational studies and to experimental studies based on a completely randomized design.

Notation

Three factors, A, B, and C, are investigated at a, b, and c levels, respectively. The mean response for the treatment when factor A is at the ith level ($i = 1, \ldots, a$), factor B is at the jth level ($j = 1, \ldots, b$), and factor C is at the kth level ($k = 1, \ldots, c$) is denoted by μ_{ijk}. The number of cases for each treatment is assumed to be constant, denoted by n. We assume $n \geq 2$. The mean response when A is at the ith level and B is at the jth level is denoted by $\mu_{ij\cdot}$, and similar notation is used for other pairs of factor levels. Since all treatment means are assumed to have equal importance, we define:

$$\mu_{ij\cdot} = \frac{\sum_k \mu_{ijk}}{c} \tag{24.1a}$$

$$\mu_{i\cdot k} = \frac{\sum_j \mu_{ijk}}{b} \tag{24.1b}$$

$$\mu_{\cdot jk} = \frac{\sum_i \mu_{ijk}}{a} \tag{24.1c}$$

The mean response when A is at the ith level is denoted by $\mu_{i\cdot\cdot}$, and similar notation is used for the other factor level means. We define:

$$\mu_{i\cdot\cdot} = \frac{\sum_j \sum_k \mu_{ijk}}{bc} \tag{24.2a}$$

$$\mu_{\cdot j\cdot} = \frac{\sum_i \sum_k \mu_{ijk}}{ac} \tag{24.2b}$$

$$\mu_{\cdot\cdot k} = \frac{\sum_i \sum_j \mu_{ijk}}{ab} \tag{24.2c}$$

Finally, the overall mean response is denoted by $\mu_{\cdot\cdot\cdot}$ and is defined:

$$\mu_{\cdot\cdot\cdot} = \frac{\sum_i \sum_j \sum_k \mu_{ijk}}{abc} \tag{24.3}$$

Illustration

To illustrate the meaning of the model terms for a three-factor analysis of variance model, we consider a study of the effects of gender, age, and intelligence level of college graduates on learning time for a complex task. Gender is factor A and has $a = 2$ levels (male, female). Age is factor B and is defined in terms of $b = 3$ levels (young, middle, old). Finally, intelligence is factor C and is defined in terms of $c = 2$ levels (high IQ, normal IQ). Table 24.1a shows the treatment means μ_{ijk} for all factor level combinations, as well as the notational representation for each. Also shown in Table 24.1a are the various means of the μ_{ijk}. Shown in Table 24.1b are various ANOVA model parameters that were computed from the treatment means in Table 24.1a. We shall refer repeatedly to this learning time example as we explain the model terms for a three-factor study.

Main Effects

The main effects in a three-factor study are defined analogously to those for a two-factor study. Thus, the *main effect* of the ith level of factor A is defined:

$$\alpha_i = \mu_{i\cdot\cdot} - \mu_{\cdot\cdot\cdot} \tag{24.4a}$$

Similarly, we define the main effect of the jth level of factor B:

$$\beta_j = \mu_{\cdot j\cdot} - \mu_{\cdot\cdot\cdot} \tag{24.4b}$$

and the main effect of the kth level of factor C:

$$\gamma_k = \mu_{\cdot\cdot k} - \mu_{\cdot\cdot\cdot} \tag{24.4c}$$

For learning time example 1 in Table 24.1, we have, for instance:

$$\alpha_1 = \mu_{1\cdot\cdot} - \mu_{\cdot\cdot\cdot} = 16.5 - 16 = .5$$
$$\beta_1 = \mu_{\cdot 1\cdot} - \mu_{\cdot\cdot\cdot} = 14 - 16 = -2$$
$$\beta_2 = \mu_{\cdot 2\cdot} - \mu_{\cdot\cdot\cdot} = 15.5 - 16 = -.5$$
$$\gamma_1 = \mu_{\cdot\cdot 1} - \mu_{\cdot\cdot\cdot} = 12 - 16 = -4$$

TABLE 24.1 Mean Learning Times and ANOVA Model Parameters—Learning Time Example 1.

(a) Mean Learning Times (in minutes)

Factor A—Gender	$k=1$ High IQ, $j=1$ Young	$j=2$ Middle	$j=3$ Old	Average	$k=2$ Normal IQ, $j=1$ Young	$j=2$ Middle	$j=3$ Old	Average	Average, $j=1$ Young	$j=2$ Middle	$j=3$ Old	Average
$i=1$ Male	9 (μ_{111})	12 (μ_{121})	18 (μ_{131})	13 ($\mu_{1\cdot1}$)	19 (μ_{112})	20 (μ_{122})	21 (μ_{132})	20 ($\mu_{1\cdot2}$)	14 ($\mu_{11\cdot}$)	16 ($\mu_{12\cdot}$)	19.5 ($\mu_{13\cdot}$)	16.5 ($\mu_{1\cdot\cdot}$)
$i=2$ Female	9 (μ_{211})	10 (μ_{221})	14 (μ_{231})	11 ($\mu_{2\cdot1}$)	19 (μ_{212})	20 (μ_{222})	21 (μ_{232})	20 ($\mu_{2\cdot2}$)	14 ($\mu_{21\cdot}$)	15 ($\mu_{22\cdot}$)	17.5 ($\mu_{23\cdot}$)	15.5 ($\mu_{2\cdot\cdot}$)
Average	9 ($\mu_{\cdot11}$)	11 ($\mu_{\cdot21}$)	16 ($\mu_{\cdot31}$)	12 ($\mu_{\cdot\cdot1}$)	19 ($\mu_{\cdot12}$)	20 ($\mu_{\cdot22}$)	21 ($\mu_{\cdot32}$)	20 ($\mu_{\cdot\cdot2}$)	14 ($\mu_{\cdot1\cdot}$)	15.5 ($\mu_{\cdot2\cdot}$)	18.5 ($\mu_{\cdot3\cdot}$)	16 ($\mu_{\cdot\cdot\cdot}$)

(b) ANOVA Model Parameters

$\mu_{\cdot\cdot\cdot} = 16.0$	$\beta_1 = -2.0$	$\gamma_1 = -4.0$	$(\alpha\beta)_{12} = 0.0$	$(\beta\gamma)_{11} = -1.0$	$(\alpha\beta\gamma)_{111} = -.5$
$\alpha_1 = .5$	$\beta_2 = -.5$	$(\alpha\beta)_{11} = -.5$	$(\alpha\gamma)_{11} = .5$	$(\beta\gamma)_{21} = -.5$	$(\alpha\beta\gamma)_{121} = 0.0$

These parameters are shown in Table 24.1b. It follows from the definitions in (24.4) that the sums of the main effects are zero:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0 \tag{24.5}$$

For example, since $\alpha_1 + \alpha_2 = 0$, it follows that $\alpha_2 = -\alpha_1 = -.5$; β_3 and γ_2 can be obtained in similar fashion. Since all main effects terms are nonzero, we know that all three main effects are present here.

Two-Factor Interactions

The two-factor interaction effects in a three-factor study are defined in the same fashion as for a two-factor study, except that all means are averaged over the third factor. Thus, following (19.8a) we define the two-factor interaction between factor A at the ith level and factor B at the jth level, denoted as before by $(\alpha\beta)_{ij}$, as follows:

$$(\alpha\beta)_{ij} = \mu_{ij\cdot} - \mu_{i\cdot\cdot} - \mu_{\cdot j\cdot} + \mu_{\cdots} \tag{24.6a}$$

In corresponding fashion, we define the AC and BC two-factor interactions:

$$(\alpha\gamma)_{ik} = \mu_{i\cdot k} - \mu_{i\cdot\cdot} - \mu_{\cdot\cdot k} + \mu_{\cdots} \tag{24.6b}$$

$$(\beta\gamma)_{jk} = \mu_{\cdot jk} - \mu_{\cdot j\cdot} - \mu_{\cdot\cdot k} + \mu_{\cdots} \tag{24.6c}$$

For learning time example 1 in Table 24.1, we have for instance:

$$\begin{aligned}
(\alpha\beta)_{11} &= 14 - 16.5 - 14 + 16 = -.5 \\
(\alpha\beta)_{12} &= 16 - 16.5 - 15.5 + 16 = 0.0 \\
(\alpha\gamma)_{11} &= 13 - 16.5 - 12 + 16 = .5 \\
(\beta\gamma)_{11} &= 9 - 14 - 12 + 16 = -1.0 \\
(\beta\gamma)_{21} &= 11 - 15.5 - 12 + 16 = -.5
\end{aligned}$$

These parameters are shown in Table 24.1b.

The two-factor interactions $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, and $(\beta\gamma)_{jk}$ are often called *first-order interactions*. It can readily be shown that the sums of the first-order interactions over each subscript are zero:

$$\sum_i (\alpha\beta)_{ij} = 0 \quad \text{for all } j \qquad \sum_j (\alpha\beta)_{ij} = 0 \quad \text{for all } i \tag{24.7a}$$

$$\sum_i (\alpha\gamma)_{ik} = 0 \quad \text{for all } k \qquad \sum_k (\alpha\gamma)_{ik} = 0 \quad \text{for all } i \tag{24.7b}$$

$$\sum_j (\beta\gamma)_{jk} = 0 \quad \text{for all } k \qquad \sum_k (\beta\gamma)_{jk} = 0 \quad \text{for all } j \tag{24.7c}$$

All two-factor interaction terms not listed in Table 24.1b can be obtained from the five terms listed and the sum-to-zero expressions in (24.7). Since nonzero $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, and $(\beta\gamma)_{jk}$ terms are present, we know that all three two-factor interactions, AB, AC, and BC, exist.

Three-Factor Interactions

Just as in a two-factor study, where the interaction between the ith level of factor A and the jth level of factor B is defined as the difference between the treatment mean μ_{ij} and the value that would be expected if the factor effects were additive, so in a three-factor study the three-factor interaction $(\alpha\beta\gamma)_{ijk}$ is defined as the difference between the treatment mean μ_{ijk} and the value that would be expected if main effects and first-order interactions were sufficient to account for all factor effects. The value that would be expected from main effects and first-order interactions when A is at the ith level, B at the jth level, and C at the kth level is:

$$\mu_{...} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \tag{24.8}$$

Hence, the *three-factor interaction* $(\alpha\beta\gamma)_{ijk}$, also called the *second-order interaction*, is defined as:

$$(\alpha\beta\gamma)_{ijk} = \mu_{ijk} - [\mu_{...} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk}] \tag{24.9a}$$

or equivalently:

$$(\alpha\beta\gamma)_{ijk} = \mu_{ijk} - \mu_{ij\cdot} - \mu_{i\cdot k} - \mu_{\cdot jk} + \mu_{i\cdot\cdot} + \mu_{\cdot j\cdot} + \mu_{\cdot\cdot k} - \mu_{...} \tag{24.9b}$$

From the definition of the three-factor interactions, it follows that they sum to zero when added over any index:

$$\sum_i (\alpha\beta\gamma)_{ijk} = 0 \quad \text{for all } j, k \qquad \sum_j (\alpha\beta\gamma)_{ijk} = 0 \quad \text{for all } i, k \qquad \sum_k (\alpha\beta\gamma)_{ijk} = 0 \quad \text{for all } i, j \tag{24.10}$$

If *all* three-factor interactions $(\alpha\beta\gamma)_{ijk}$ are zero, we say that there are no three-factor interactions among factors A, B, and C. If some $(\alpha\beta\gamma)_{ijk}$ are not zero, we say that three-factor interactions are present.

Let us find the three-factor interaction $(\alpha\beta\gamma)_{111}$ for the learning time example in Table 24.1. From (24.9a), we have for $i = j = k = 1$:

$$(\alpha\beta\gamma)_{111} = \mu_{111} - [\mu_{...} + \alpha_1 + \beta_1 + \gamma_1 + (\alpha\beta)_{11} + (\alpha\gamma)_{11} + (\beta\gamma)_{11}]$$

Using the ANOVA model parameter values from Table 24.1b, we obtain:

$$(\alpha\beta\gamma)_{111} = 9 - (16 + .5 - 2 - 4 - .5 + .5 - 1) = -.5$$

Since $(\alpha\beta\gamma)_{111}$ is not zero, we know at once that three-factor interactions are present in this example.

Cell Means Model

Let Y_{ijkm} denote the observation for the mth case or trial ($m = 1, \ldots, n$) for the treatment consisting of the ith level of A ($i = 1, \ldots, a$), the jth level of B ($j = 1, \ldots, b$), and the kth level of C ($k = 1, \ldots, c$). Thus, the total number of cases in the study is:

$$n_T = nabc \tag{24.11}$$

The ANOVA model for a three-factor study in terms of the cell (treatment) means μ_{ijk} with fixed factor levels is:

$$Y_{ijkm} = \mu_{ijk} + \varepsilon_{ijkm} \tag{24.12}$$

where:

μ_{ijk} are parameters
ε_{ijkm} are independent $N(0, \sigma^2)$
$i = 1, \ldots, a;\ j = 1, \ldots, b;\ k = 1, \ldots, c;\ m = 1, \ldots, n$

Factor Effects Model

An equivalent factor effects model can be developed that incorporates the factorial structure by expressing each treatment mean μ_{ijk} in terms of the various factor effects. From the three-factor interaction definition (24.9a), we have the identity:

$$\mu_{ijk} \equiv \mu_{\cdots} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} \tag{24.13}$$

where:

$$\begin{aligned}
\mu_{\cdots} &= \frac{\sum\sum\sum \mu_{ijk}}{abc} \\
\alpha_i &= \mu_{i\cdot\cdot} - \mu_{\cdots} \\
\beta_j &= \mu_{\cdot j\cdot} - \mu_{\cdots} \\
\gamma_k &= \mu_{\cdot\cdot k} - \mu_{\cdots} \\
(\alpha\beta)_{ij} &= \mu_{ij\cdot} - \mu_{i\cdot\cdot} - \mu_{\cdot j\cdot} + \mu_{\cdots} \\
(\alpha\gamma)_{ik} &= \mu_{i\cdot k} - \mu_{i\cdot\cdot} - \mu_{\cdot\cdot k} + \mu_{\cdots} \\
(\beta\gamma)_{jk} &= \mu_{\cdot jk} - \mu_{\cdot j\cdot} - \mu_{\cdot\cdot k} + \mu_{\cdots} \\
(\alpha\beta\gamma)_{ijk} &= \mu_{ijk} - \mu_{ij\cdot} - \mu_{i\cdot k} - \mu_{\cdot jk} + \mu_{i\cdot\cdot} + \mu_{\cdot j\cdot} + \mu_{\cdot\cdot k} - \mu_{\cdots}
\end{aligned}$$

Hence, the equivalent factor effects ANOVA model for a three-factor study is:

$$Y_{ijkm} = \mu_{\cdots} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkm} \tag{24.14}$$

where:

ε_{ijkm} are independent $N(0, \sigma^2)$
$\alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{ij}, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}, (\alpha\beta\gamma)_{ijk}$ are constants subject to the restrictions:

$$\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0$$

$$\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = \sum_i (\alpha\gamma)_{ik} = 0$$

$$\sum_k (\alpha\gamma)_{ik} = \sum_j (\beta\gamma)_{jk} = \sum_k (\beta\gamma)_{jk} = 0$$

$$\sum_i (\alpha\beta\gamma)_{ijk} = \sum_j (\alpha\beta\gamma)_{ijk} = \sum_k (\alpha\beta\gamma)_{ijk} = 0$$

Both the cell means model (24.12) and the equivalent factor effects model (24.14) are linear models, just as in the two-factor case. We shall illustrate this for an example later in the chapter.

24.2 Interpretation of Interactions in Three-Factor Studies

To shed light on the nature of interactions in three-factor studies, we shall examine three variations of the learning time example by means of tables and graphs. The first example corresponds to learning time example 1, in which—as we have already determined—a three-factor interaction is present. In learning time example 2, there is no three-factor interaction, but two two-factor interactions are present. Finally, in learning time example 3, there is again no three-factor interaction but there is just one two-factor interaction. In each example, we present the true treatment means μ_{ijk} and the true ANOVA model parameters.

Learning Time Example 1: Interpretation of Three-Factor Interactions

In a three-factor study, the presence of a three-factor interaction indicates that responses must be explained in terms of the *combined effects of all three factors*. Thus, no simplified explanation, for example in terms of main effects or first-order interactions, is possible. Any graphical presentation of cell means should display all of the individual cell means μ_{ijk}. A convenient way to do so is to create separate two-factor treatment means or interaction plots for each level of a third factor. For example, the AB treatment means plots for the two levels of factor C are displayed in Figure 24.1 for the cell means in Table 24.1. Recall that the learning time example considers the effects of gender (factor A), age (factor B), and intelligence (factor C) on learning time. Specifically, Figure 24.1 shows that for persons with normal IQ, gender has no effect on mean learning time, and age has only a small effect leading to slightly longer learning times for older persons. For persons with high IQ, on the other hand, females tend to learn more quickly than males for older persons but not for young persons, and older persons tend to require substantially longer learning times than young persons.

Notice that the slopes of the curves in the AB cell means plots are not the same for the two levels of C. For the first level of C, the curves for middle-aged and older subjects are

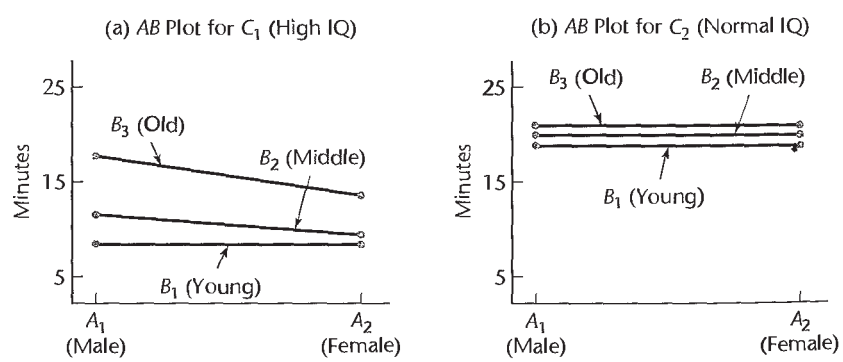


FIGURE 24.1 Cell Means Plot with *ABC* Interaction Present—Learning Time Example 1.

sloping downward, while these curves both have zero slope for the second level of C. This lack of parallelism in the two plots will always be present if a three-factor interaction exists, but this is not the only way such slope changes can arise. As we will see in the next example, if an AB interaction is present and either A or B also interacts with C, lack of parallelism will also be present when the AB interaction is displayed for each level of C.

If three-factor interactions are difficult to understand, higher-order interactions such as four-factor interactions in studies involving more than three factors are yet more abstruse. Fortunately, it is often found in practice that these higher-order interactions are quite small or nonexistent. When this is the case, they can be disregarded in the analysis of factor effects.

Learning Time Example 2: Interpretation of Multiple Two-Factor Interactions

The set up for learning time example 2 is the same as that for learning time example 1—that is, we consider the same study of the effects of gender, age, and intelligence level on learning of a complex task—but the true cell means have changed. Table 24.2 lists the cell means and the corresponding ANOVA model parameters for learning time example 2.

It is easy to see from a review of these parameters that all ABC interaction terms $(\alpha\beta\gamma)_{ijk}$ and all BC interaction terms $(\beta\gamma)_{jk}$ are zero; however, AB and AC interactions are present, since $(\alpha\beta)_{11} = -.5$ and $(\alpha\gamma)_{11} = .5$.

Figures 24.2a and 24.2b display the AB interactions for the two levels of C. The lack of parallelism of the AB curves within each panel reflects the presence of AB interactions. Notice also that the slopes of the curves in Figure 24.2a for high IQ subjects are negative, while those in Figure 24.2b for normal IQ subjects are all close to zero. The fact that the AB curves for a given level of factor B are not parallel for the two levels of factor C reflects the presence of AC interactions in this example. The AC treatment means plots are shown in Figures 24.2c–e for each of the three levels of factor B. As expected, the AC curves in each panel are not parallel. Note finally that the slopes of the AC curves change from panel to panel. This lack of parallelism reflects the presence of the AB interaction in this example.

TABLE 24.2 Mean Learning Times and ANOVA Model Parameters— Learning Time Example 2.

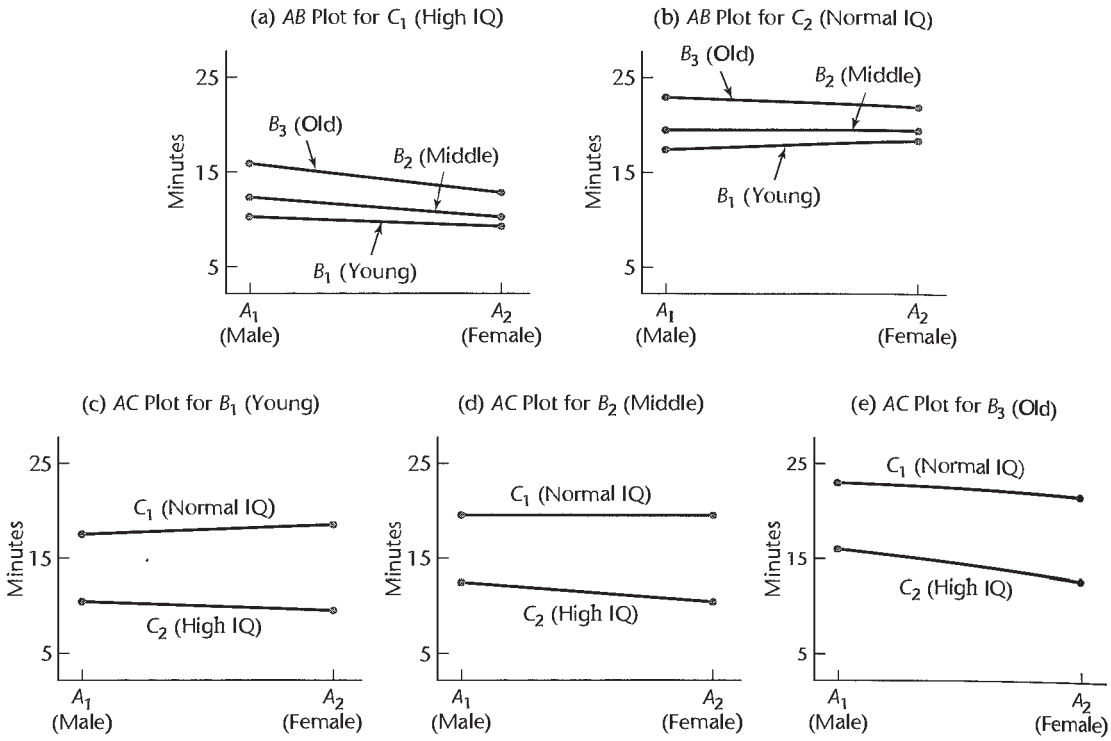
(a) Mean Learning Times (in minutes)

Factor	Intelligence (factor C) and Age (factor B)					
	$k = 1$ High IQ			$k = 2$ Normal IQ		
Gender	$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old	$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old
$i = 1$ (Males)	10.5	12.5	16	17.5	19.5	23
$i = 2$ (Females)	9.5	10.5	13	18.5	19.5	22

(b) ANOVA Model Parameters

$\mu_{...} = 16.0$	$\beta_1 = -2.0$	$\gamma_1 = -4.0$	$(\alpha\beta)_{12} = 0.0$	$(\beta\gamma)_{11} = 0.0$	$(\alpha\beta\gamma)_{111} = 0.0$
$\alpha_1 = .5$	$\beta_2 = -.5$	$(\alpha\beta)_{11} = -.5$	$(\alpha\gamma)_{11} = .5$	$(\beta\gamma)_{21} = 0.0$	$(\alpha\beta\gamma)_{121} = 0.0$

FIGURE 24.2 Cell Means Plots with AB and AC Interactions Present—Learning Time Example 2.



Learning Time Example 3: Interpretation of a Single Two-Factor Interaction

Cell means and corresponding ANOVA model parameters for learning time example 3 are given in Tables 24.3a and 24.3b, respectively. The set up is again the same as that for learning time examples 1 and 2, however the cell means have changed. Note from Table 24.3b, that all parameters corresponding to the ABC interaction are zero, as are those corresponding to AC and BC. The two-factor interaction AB is present, since $(\alpha\beta)_{11} = -.5$.

Figure 24.3a and 24.3b display the AB treatment means plots for each level of C. The slopes of the curves within each panel are not parallel, reflecting the presence of an AB interaction. Note also that the AB plots in Figure 24.3a are identical to those in Figure 24.3b, except that the cell means plotted in Figure 24.3b have been uniformly shifted up by eight minutes. This reflects the absence of the AC, BC, and ABC interactions in this example.

Since the curves in the two AB plots are identical for the different levels of factor C except for the vertical displacement (i.e., since no AC, BC, or ABC interactions are present) separate panels are not necessary for interpreting the AB interaction. The *overall* AB cell means plot displays the cell means $\mu_{ij\cdot}$ when averaged over the levels of C. This plot is shown in Figure 24.3c. Notice that the slopes in the plot are identical to those in Figures 24.3a and 24.3b. The $\mu_{ij\cdot}$ values plotted are the averages of the corresponding cell

BLE 24.3 Mean Learning Times and ANOVA Model Parameters— Learning Time Example 3.

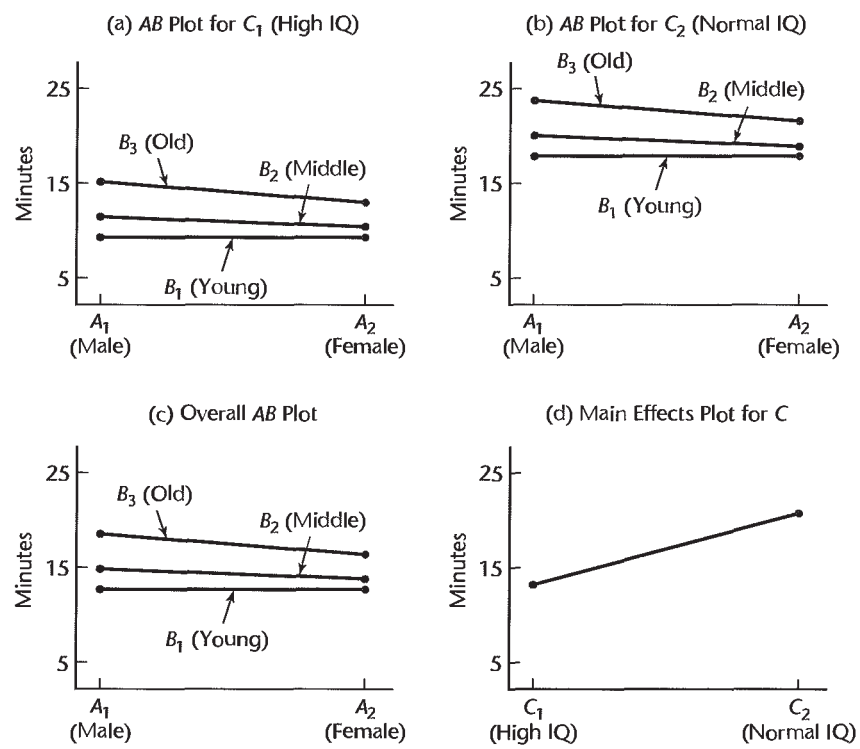
(a) Mean Learning Times (in minutes)

	Intelligence (factor C) and Age (factor B)					
	$k = 1$ High IQ			$k = 2$ Normal IQ		
or e der	$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old	$j = 1$ Young	$j = 2$ Middle	$j = 3$ Old
(Males)	10	12	15.5	18	20	23.5
2 (Females)	10	11	13.5	18	19	21.5

(b) ANOVA Model Parameters

$\mu = 16.0$	$\beta_1 = -2.0$	$\gamma_1 = -4.0$	$(\alpha\beta)_{12} = 0.0$	$(\beta\gamma)_{11} = 0.0$	$(\alpha\beta\gamma)_{111} = 0.0$
$\alpha_1 = .5$	$\beta_2 = -.5$	$(\alpha\beta)_{11} = -.5$	$(\alpha\gamma)_{11} = 0.0$	$(\beta\gamma)_{21} = 0.0$	$(\alpha\beta\gamma)_{121} = 0.0$

FIGURE 24.3 Cell Means Plots With *AB* Interaction Present— Learning Time Example 3.



means μ_{ij1} and μ_{ij2} in Figures 24.3a and 24.3b. Because factor C is present as a main effect and does not interact with either A or B, ($\gamma_1 = -4$), its effect can be shown and interpreted separately, using a bar graph, a main effects plot, or a line plot. A main effects plot for the factor C effect is shown in Figure 24.3d.

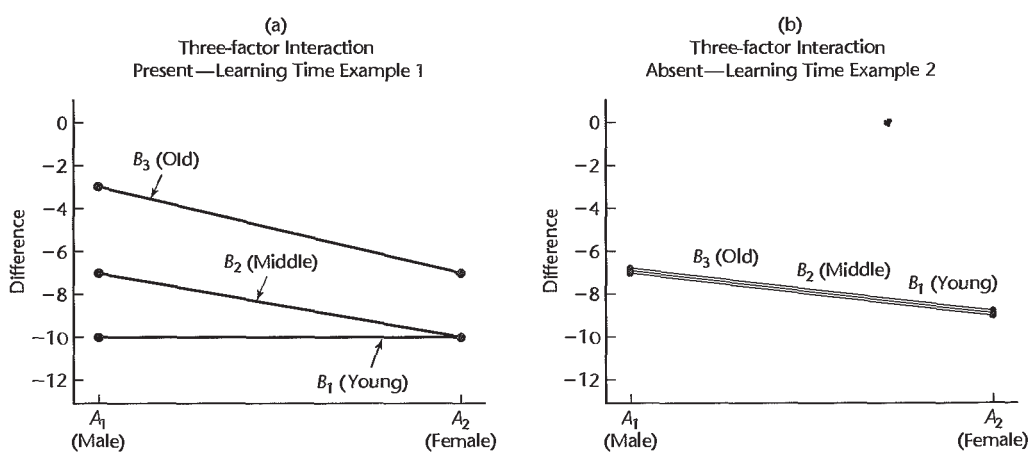
Comment

One way to determine whether or not a three-factor interaction exists is to plot *differences of treatment means* in a manner similar to two-factor interaction plots, as proposed in Reference 24.1. It can be shown that if a three-factor interaction is not present, then the differences between means with respect to any one of the factors will lead to parallel curves in the interaction plot of the differences. Conversely, if a three-factor interaction is present, the difference curves will not be parallel. For instance, in a three-factor study where the third factor is at two levels (such as in the learning time example) we would examine the differences $\mu_{ij1} - \mu_{ij2}$ for all i and j. If the AB-interaction plots for these differences show parallel curves, then no three-factor interactions are present. We refer to this plot as a *treatment means differences plot.* (If the third factor has $c > 2$ levels, $c - 1$ interaction plots of the differences $\mu_{ijk} - \mu_{ij,k+1}$ for $k = 1, \ldots, c - 1$ are constructed, and lack of parallelism in any one of the plots would indicate the presence of a three-factor interaction.)

Treatment means differences plots are shown for learning time examples 1 and 2 in Figures 24.4a and 24.4b, respectively. We see from Figure 24.4a that the difference curves are not parallel, indicating the presence of a three-factor interaction. On the other hand, there is no three-factor interaction for learning time example 2, and this is reflected by the parallelism of the three curves in Figure 24.4b. For this example, these curves happen to be identical. The curves in the plot have been jittered slightly so that all three curves can be seen.

Note that the main purpose of the treatment means differences plot is to diagnose the presence or absence of a three-factor interaction, and beyond this it does not contribute substantially to the interpretation of results. For this reason we do not advocate routine use of this plot with estimated treatment means. We shall employ analysis of variance techniques in Section 24.3 to identify which interactions are present, and then display appropriate treatment means plots or main effects plots to summarize and interpret results. ■

FIGURE 24.4 Treatment Means Differences Plot—Learning Time Examples 1 and 2.



24.3 Fitting of ANOVA Model

Notation

The notation for sample totals and means is a straightforward extension of that for two-factor studies. As usual, a dot in the subscript indicates aggregation or averaging over the index represented by the dot. We have:

$$Y_{ijk\cdot} = \sum_m Y_{ijkm} \qquad \bar{Y}_{ijk\cdot} = \frac{Y_{ijk\cdot}}{n} \tag{24.15a}$$

$$Y_{ij\cdot\cdot} = \sum_k \sum_m Y_{ijkm} \qquad \bar{Y}_{ij\cdot\cdot} = \frac{Y_{ij\cdot\cdot}}{cn} \tag{24.15b}$$

$$Y_{i\cdot k\cdot} = \sum_j \sum_m Y_{ijkm} \qquad \bar{Y}_{i\cdot k\cdot} = \frac{Y_{i\cdot k\cdot}}{bn} \tag{24.15c}$$

$$Y_{\cdot jk\cdot} = \sum_i \sum_m Y_{ijkm} \qquad \bar{Y}_{\cdot jk\cdot} = \frac{Y_{\cdot jk\cdot}}{an} \tag{24.15d}$$

$$Y_{i\cdots} = \sum_j \sum_k \sum_m Y_{ijkm} \qquad \bar{Y}_{i\cdots} = \frac{Y_{i\cdots}}{bcn} \tag{24.15e}$$

$$Y_{\cdot j\cdot\cdot} = \sum_i \sum_k \sum_m Y_{ijkm} \qquad \bar{Y}_{\cdot j\cdot\cdot} = \frac{Y_{\cdot j\cdot\cdot}}{acn} \tag{24.15f}$$

$$Y_{\cdot\cdot k\cdot} = \sum_i \sum_j \sum_m Y_{ijkm} \qquad \bar{Y}_{\cdot\cdot k\cdot} = \frac{Y_{\cdot\cdot k\cdot}}{abn} \tag{24.15g}$$

$$Y_{\cdots\cdot} = \sum_i \sum_j \sum_k \sum_m Y_{ijkm} \qquad \bar{Y}_{\cdots\cdot} = \frac{Y_{\cdots\cdot}}{abcn} \tag{24.15h}$$

Later in this section we illustrate this notation for a study of the effects of gender, body fat, and smoking history on exercise tolerance in stress testing. Each of the three factors has two levels, and there are three replications for each treatment. Tables 24.4a and b show, respectively, the data and estimated means, together with the corresponding notation.

Fitting of ANOVA Model

When the normal error cell means model (24.12) is fitted by the method of least squares or the method of maximum likelihood, the estimators as usual turn out to be the estimated treatment means:

$$\hat{\mu}_{ijk} = \bar{Y}_{ijk\cdot} \tag{24.16}$$

TABLE 24.4 Sample Data and Estimated Treatment and Factor Level Means for Three-Factor Study—Stress Test Example.

(a) Data

	Smoking History	
	$k = 1$ Light	$k = 2$ Heavy
$j = 1$ Low fat:		
$i = 1$ Male	24.1 (Y_{1111})	17.6 (Y_{1121})
	29.2 (Y_{1112})	18.8 (Y_{1122})
	24.6 (Y_{1113})	23.2 (Y_{1123})
$i = 2$ Female	20.0 (Y_{2111})	14.8 (Y_{2121})
	21.9 (Y_{2112})	10.3 (Y_{2122})
	17.6 (Y_{2113})	11.3 (Y_{2123})
$j = 2$ High fat:		
$i = 1$ Male	14.6 (Y_{1211})	14.9 (Y_{1221})
	15.3 (Y_{1212})	20.4 (Y_{1222})
	12.3 (Y_{1213})	12.8 (Y_{1223})
$i = 2$ Female	16.1 (Y_{2211})	10.1 (Y_{2221})
	9.3 (Y_{2212})	14.4 (Y_{2222})
	10.8 (Y_{2213})	6.1 (Y_{2223})

(b) Estimated Means

	$k = 1$	$k = 2$	All k
$j = 1$:			
$i = 1$	25.97 ($\bar{Y}_{111\cdot}$)	19.87 ($\bar{Y}_{112\cdot}$)	22.92 ($\bar{Y}_{11\cdot\cdot}$)
$i = 2$	19.83 ($\bar{Y}_{211\cdot}$)	12.13 ($\bar{Y}_{212\cdot}$)	15.98 ($\bar{Y}_{21\cdot\cdot}$)
All i	22.90 ($\bar{Y}_{\cdot 11\cdot}$)	16.00 ($\bar{Y}_{\cdot 12\cdot}$)	19.45 ($\bar{Y}_{\cdot 1\cdot\cdot}$)
$j = 2$:			
$i = 1$	14.07 ($\bar{Y}_{121\cdot}$)	16.03 ($\bar{Y}_{122\cdot}$)	15.05 ($\bar{Y}_{12\cdot\cdot}$)
$i = 2$	12.07 ($\bar{Y}_{221\cdot}$)	10.20 ($\bar{Y}_{222\cdot}$)	11.13 ($\bar{Y}_{22\cdot\cdot}$)
All i	13.07 ($\bar{Y}_{\cdot 21\cdot}$)	13.12 ($\bar{Y}_{\cdot 22\cdot}$)	13.09 ($\bar{Y}_{\cdot 2\cdot\cdot}$)
All j:			
$i = 1$	20.02 ($\bar{Y}_{1\cdot 1\cdot}$)	17.95 ($\bar{Y}_{1\cdot 2\cdot}$)	18.98 ($\bar{Y}_{1\cdot\cdot\cdot}$)
$i = 2$	15.95 ($\bar{Y}_{2\cdot 1\cdot}$)	11.17 ($\bar{Y}_{2\cdot 2\cdot}$)	13.56 ($\bar{Y}_{2\cdot\cdot\cdot}$)
All i	17.98 ($\bar{Y}_{\cdot\cdot 1\cdot}$)	14.56 ($\bar{Y}_{\cdot\cdot 2\cdot}$)	16.27 ($\bar{Y}_{\cdot\cdot\cdot\cdot}$)

Thus, the *fitted values* for the observations are the estimated treatment

$$\hat{Y}_{ijkm} = \bar{Y}_{ijk\cdot}$$

and the *residuals* are the deviations of the observed values from the means:

$$e_{ijkm} = Y_{ijkm} - \hat{Y}_{ijkm} = Y_{ijkm} - \bar{Y}_{ijk\cdot}$$

For the equivalent factor effects model (24.14), the least squares and maximum likelihood estimators of the parameters are as follows:

Parameter	Estimator	
$\mu_{\cdots}$	$\hat{\mu}_{\cdots} = \bar{Y}_{\cdots}$	**(24.19a)**
α_i	$\hat{\alpha}_i = \bar{Y}_{i\cdots} + \bar{Y}_{\cdots}$	**(24.19b)**
β_j	$\hat{\beta}_j = \bar{Y}_{\cdot j\cdot\cdot} + \bar{Y}_{\cdots}$	**(24.19c)**
γ_k	$\hat{\gamma}_k = \bar{Y}_{\cdot\cdot k\cdot} + \bar{Y}_{\cdots}$	**(24.19d)**
$(\alpha\beta)_{ij}$	$\widehat{(\alpha\beta)}_{ij} = \bar{Y}_{ij\cdot\cdot} + \bar{Y}_{i\cdots} + \bar{Y}_{\cdot j\cdot\cdot} + \bar{Y}_{\cdots}$	**(24.19e)**
$(\alpha\gamma)_{ik}$	$\widehat{(\alpha\gamma)}_{ik} = \bar{Y}_{i\cdot k\cdot} + \bar{Y}_{i\cdots} + \bar{Y}_{\cdot\cdot k\cdot} + \bar{Y}_{\cdots}$	**(24.19f)**
$(\beta\gamma)_{jk}$	$\widehat{(\beta\gamma)}_{jk} = \bar{Y}_{\cdot jk\cdot} + \bar{Y}_{\cdot j\cdot\cdot} + \bar{Y}_{\cdot\cdot k\cdot} + \bar{Y}_{\cdots}$	**(24.19g)**
$(\alpha\beta\gamma)_{ijk}$	$\widehat{(\alpha\beta\gamma)}_{ijk} = \bar{Y}_{ijk\cdot} - \bar{Y}_{ij\cdot\cdot} - \bar{Y}_{i\cdot k\cdot} - \bar{Y}_{\cdot jk\cdot} + \bar{Y}_{i\cdots} + \bar{Y}_{\cdot j\cdot\cdot} + \bar{Y}_{\cdot\cdot k\cdot} - \bar{Y}_{\cdots}$	**(24.19h)**

The fitted values and residuals for factor effects model (24.14) are the same as those in (24.17) and (24.18) for cell means model (24.12), as was the case for two-factor studies.

Evaluation of Appropriateness of ANOVA Model

No new problems arise in examining the appropriateness of the three-factor analysis of variance model. The residuals (24.18):

$$e_{ijkm} = Y_{ijkm} - \bar{Y}_{ijk\cdot} \tag{24.20}$$

may be examined for normality, constancy of error variance, and independence of error terms in the same fashion as for single-factor and two-factor studies.

Weighted least squares as usual is a standard remedial measure when the error variance is not constant but the distribution of the error terms is normal. A transformation of the response variable may be helpful to stabilize the error variance, to make the error distributions more normal, and/or to make important interactions unimportant. Our earlier discussions of these topics apply completely to the three-factor case.

Finally, our earlier discussion on the effects of departures from the ANOVA model applies fully to the three-factor case. In particular, the employment of equal sample sizes for all treatments minimizes the effect of unequal variances.

Example

The effects of gender of subject (factor A), body fat of subject (measured in percent, factor B), and smoking history of subject (factor C) on exercise tolerance (Y) were studied in a small-scale investigation of persons 25 to 35 years old. Exercise tolerance was measured in minutes until fatigue occurs while the subject is performing on a bicycle

TABLE 24.5 General ANOVA Table for Three-Factor Study with Fixed Factor Levels.

Source of Variation	*SS*	*df*	*MS*	$E\{MS\}$
Factor *A*	*SSA*	$a-1$	*MSA*	$\sigma^2 + bcn\frac{\sum \alpha_i^2}{a-1}$
Factor *B*	*SSB*	$b-1$	*MSB*	$\sigma^2 + acn\frac{\sum \beta_j^2}{b-1}$
Factor *C*	*SSC*	$c-1$	*MSC*	$\sigma^2 + abn\frac{\sum \gamma_k^2}{c-1}$
AB interactions	*SSAB*	$(a-1)(b-1)$	*MSAB*	$\sigma^2 + cn\frac{\sum\sum (\alpha\beta)_{ij}^2}{(a-1)(b-1)}$
AC interactions	*SSAC*	$(a-1)(c-1)$	*MSAC*	$\sigma^2 + bn\frac{\sum\sum (\alpha\gamma)_{ik}^2}{(a-1)(c-1)}$
BC interactions	*SSBC*	$(b-1)(c-1)$	*MSBC*	$\sigma^2 + an\frac{\sum\sum (\beta\gamma)_{jk}^2}{(b-1)(c-1)}$
ABC interactions	*SSABC*	$(a-1)(b-1)(c-1)$	*MSABC*	$\sigma^2 + n\frac{\sum\sum\sum (\alpha\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$
Error	*SSE*	$abc(n-1)$	*MSE*	σ^2
Total	*SSTO*	$abcn-1$		

Note: $\mu_{...}$, α_i, β_j, γ_k, $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$, and $(\alpha\beta\gamma)_{ijk}$ are defined in (24.13).

apparatus. Three subjects for each gender-body fat-smoking history group were given the exercise tolerance stress test. The results are recorded in Table 24.4a. Note that each factor has two levels ($a = b = c = 2$) and that there are three replications ($n = 3$) for each treatment.

The estimated treatment and factor level means are presented in Table 24.4b. Figure 24.5a contains the *BC* treatment means plots for each level of factor *A*, and Figure 24.5b contains the *AB* treatment means plots for each level of *C*. It appears that some factors may interact in their effect on exercise tolerance and that gender, in particular, may affect the endurance in stress testing.

Residual Analysis. The researcher first prepared aligned residual dot plots for the eight treatments. These plots (not shown), though based on only three observations for each treatment, did not suggest any gross differences in the error variances for the eight treatments. The researcher also obtained a normal probability plot of the residuals, shown in Figure 24.6. The points in this plot form a moderately linear pattern. Normality of the error terms is supported by the high coefficient of correlation between the ordered residuals and their expected values under normality, namely, .969. The researcher was therefore satisfied that three-factor ANOVA model (24.14) is applicable here, and now wishes to analyze the nature of the factor effects in detail.

FIGURE 24.5
lots of timated eatment eans—Stress est Example.

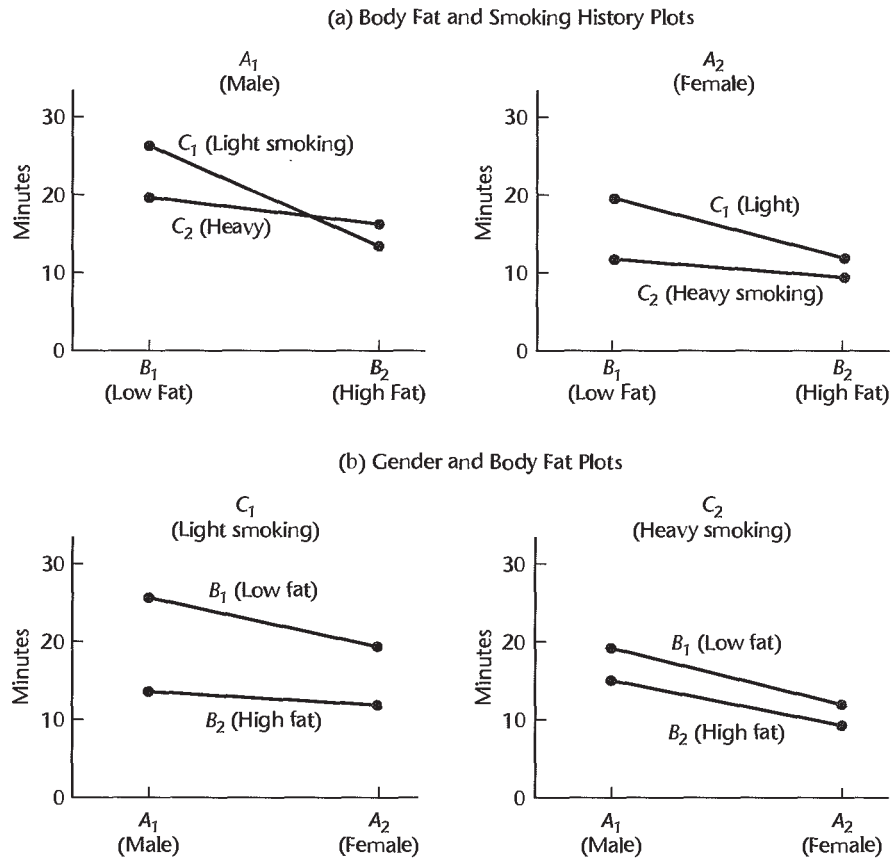
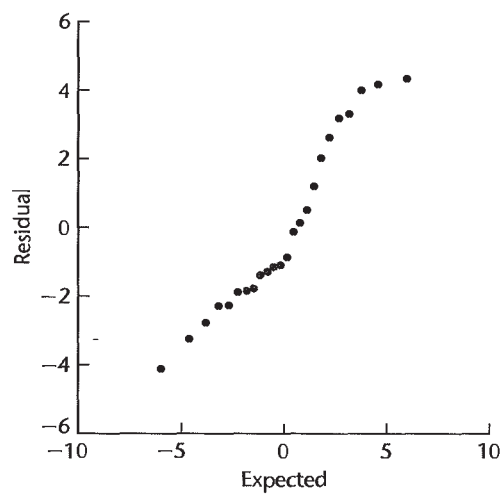


FIGURE 24.6
Normal Probability Plot of Residuals—Stress Test Example.



24.4 Analysis of Variance

Partitioning of Total Sum of Squares

Neglecting the factorial structure of the three-factor study and simply considering it to contain *abc* treatments, we obtain the usual breakdown of the total sum of squares:

$$SSTO = SSTR + SSE \tag{24.21}$$

where:

$$SSTO = \sum_i \sum_j \sum_k \sum_m (Y_{ijkm} - \bar{Y}_{....})^2 \tag{24.21a}$$

$$SSTR = n \sum_i \sum_j \sum_k (\bar{Y}_{ijk\cdot} - \bar{Y}_{....})^2 \tag{24.21b}$$

$$SSE = \sum_i \sum_j \sum_k \sum_m (Y_{ijkm} - \bar{Y}_{ijk\cdot})^2 = \sum_i \sum_j \sum_k \sum_m e_{ijkm}^2 \tag{24.21c}$$

Consider now the estimated treatment mean deviation $\bar{Y}_{ijk\cdot} - \bar{Y}_{....}$, which appears in *SSTR*. This can be decomposed in terms of the estimators in (24.19) of the main effects, two-factor interactions, and three-factor interaction:

$$\underbrace{\bar{Y}_{ijk\cdot} - \bar{Y}_{....}}_{\substack{\text{Estimated}\\ \text{treatment}\\ \text{mean deviation}}} = \underbrace{\bar{Y}_{i...} - \bar{Y}_{....}}_{A \text{ main effect}} + \underbrace{\bar{Y}_{\cdot j\cdot\cdot} - \bar{Y}_{....}}_{B \text{ main effect}} + \underbrace{\bar{Y}_{\cdot\cdot k\cdot} - \bar{Y}_{....}}_{C \text{ main effect}} + \underbrace{\bar{Y}_{ij\cdot\cdot} - \bar{Y}_{i...} - \bar{Y}_{\cdot j\cdot\cdot} + \bar{Y}_{....}}_{AB \text{ interaction effect}}$$

$$+ \underbrace{\bar{Y}_{i\cdot k\cdot} - \bar{Y}_{i...} - \bar{Y}_{\cdot\cdot k\cdot} + \bar{Y}_{....}}_{AC \text{ interaction effect}} + \underbrace{\bar{Y}_{\cdot jk\cdot} - \bar{Y}_{\cdot j\cdot\cdot} - \bar{Y}_{\cdot\cdot k\cdot} + \bar{Y}_{....}}_{BC \text{ interaction effect}}$$

$$+ \underbrace{\bar{Y}_{ijk\cdot} - \bar{Y}_{ij\cdot\cdot} - \bar{Y}_{i\cdot k\cdot} - \bar{Y}_{\cdot jk\cdot} + \bar{Y}_{i...} + \bar{Y}_{\cdot j\cdot\cdot} + \bar{Y}_{\cdot\cdot k\cdot} - \bar{Y}_{....}}_{ABC \text{ interaction effect}}$$

When we square each side and sum over i, j, k, and m, all cross-product terms drop out and we obtain:

$$SSTR = SSA + SSB + SSC + SSAB + SSAC + SSBC + SSABC \tag{24.22}$$

where:

$$SSA = nbc \sum_i (\bar{Y}_{i...} - \bar{Y}_{....})^2 \tag{24.22a}$$

$$SSB = nac \sum_j (\bar{Y}_{\cdot j\cdot\cdot} - \bar{Y}_{....})^2 \tag{24.22b}$$

$$SSC = nab \sum_k (\bar{Y}_{\cdot\cdot k\cdot} - \bar{Y}_{....})^2 \tag{24.22c}$$

$$SSAB = nc \sum_i \sum_j (\bar{Y}_{ij\cdot\cdot} - \bar{Y}_{i\cdot\cdot\cdot} - \bar{Y}_{\cdot j\cdot\cdot} + \bar{Y}_{\cdot\cdot\cdot\cdot})^2 \tag{24.22d}$$

$$SSAC = nb \sum_i \sum_k (\bar{Y}_{i\cdot k\cdot} - \bar{Y}_{i\cdot\cdot\cdot} - \bar{Y}_{\cdot\cdot k\cdot} + \bar{Y}_{\cdot\cdot\cdot\cdot})^2 \tag{24.22e}$$

$$SSBC = na \sum_j \sum_k (\bar{Y}_{\cdot jk\cdot} - \bar{Y}_{\cdot j\cdot\cdot} - \bar{Y}_{\cdot\cdot k\cdot} + \bar{Y}_{\cdot\cdot\cdot\cdot})^2 \tag{24.22f}$$

$$SSABC = n \sum_i \sum_j \sum_k (\bar{Y}_{ijk\cdot} - \bar{Y}_{ij\cdot\cdot} - \bar{Y}_{i\cdot k\cdot} - \bar{Y}_{\cdot jk\cdot} + \bar{Y}_{i\cdot\cdot\cdot} + \bar{Y}_{\cdot j\cdot\cdot} + \bar{Y}_{\cdot\cdot k\cdot} - \bar{Y}_{\cdot\cdot\cdot\cdot})^2 \tag{24.22g}$$

Combining (24.21) and (24.22), we have thus established the orthogonal decomposition:

$$SSTO = SSA + SSB + SSC + SSAB + SSAC + SSBC + SSABC + SSE \tag{24.23}$$

SSA, SSB, and *SSC* are the usual main effects sums of squares. For instance, the larger (absolutely) are the estimated main B effects $\bar{Y}_{\cdot j\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot\cdot}$, the larger will be *SSB*.

SSAB, SSAC, and *SSBC* are the usual two-factor interactions sums of squares. For instance, the larger (absolutely) are the estimated AB interactions $\bar{Y}_{ij\cdot\cdot} - \bar{Y}_{i\cdot\cdot\cdot} - \bar{Y}_{\cdot j\cdot\cdot} + \bar{Y}_{\cdot\cdot\cdot\cdot}$, the larger will be *SSAB*.

Finally, *SSABC* is the three-factor interactions sum of squares. The larger (absolutely) are these estimated three-factor interactions, the larger will be *SSABC*.

Degrees of Freedom and Mean Squares

Table 24.5 contains the general ANOVA table for three-factor ANOVA model (24.14). The degrees of freedom for main effects and two-factor interactions sums of squares correspond to those for two-factor studies. The number of degrees of freedom associated with *SSABC* is obtained by subtraction and corresponds to the number of independent linear relations among all the interaction terms $(\alpha\beta\gamma)_{ijk}$.

The expected mean squares are also given in Table 24.5. Note that *MSA, MSB, MSC, MSAB, MSAC, MSBC,* and *MSABC* all have expectations equal to σ^2 if there are no factor effects of the type reflected by the mean square. If such effects are present, each mean square has an expectation exceeding σ^2. As usual, $E\{MSE\} = \sigma^2$ always. Hence, the tests for factor effects consist of comparing the appropriate mean square against *MSE* by means of an F^* test statistic, with large values of F^* indicating the presence of factor effects.

Tests for Factor Effects

The various tests for factor effects all follow the same pattern; we illustrate them with the test for three-factor interactions. The alternatives are:

$$\begin{aligned} &H_0\text{: all } (\alpha\beta\gamma)_{ijk} = 0 \\ &H_a\text{: not all } (\alpha\beta\gamma)_{ijk} \text{ equal zero} \end{aligned} \tag{24.24a}$$

The appropriate test statistic is:

$$F^* = \frac{MSABC}{MSE} \tag{24.24b}$$

TABLE 24.6 Test Statistics for Three-Factor Study with Fixed Factor Levels.

Alternatives	Test Statistic	Percentile
H_0: all $\alpha_i = 0$ H_a: not all $\alpha_i = 0$	$F^* = \frac{MSA}{MSE}$	$F[1-\alpha; a-1, (n-1)abc]$
H_0: all $\beta_j = 0$ H_a: not all $\beta_j = 0$	$F^* = \frac{MSB}{MSE}$	$F[1-\alpha; b-1, (n-1)abc]$
H_0: all $\gamma_k = 0$ H_a: not all $\gamma_k = 0$	$F^* = \frac{MSC}{MSE}$	$F[1-\alpha; c-1, (n-1)abc]$
H_0: all $(\alpha\beta)_{ij} = 0$ H_a: not all $(\alpha\beta)_{ij} = 0$	$F^* = \frac{MSAB}{MSE}$	$F[1-\alpha; (a-1)(b-1), (n-1)abc]$
H_0: all $(\alpha\gamma)_{ik} = 0$ H_a: not all $(\alpha\gamma)_{ik} = 0$	$F^* = \frac{MSAC}{MSE}$	$F[1-\alpha; (a-1)(c-1), (n-1)abc]$
H_0: all $(\beta\gamma)_{jk} = 0$ H_a: not all $(\beta\gamma)_{jk} = 0$	$F^* = \frac{MSBC}{MSE}$	$F[1-\alpha; (b-1)(c-1), (n-1)abc]$
H_0: all $(\alpha\beta\gamma)_{ijk} = 0$ H_a: not all $(\alpha\beta\gamma)_{ijk} = 0$	$F^* = \frac{MSABC}{MSE}$	$F[1-\alpha; (a-1)(b-1)(c-1), (n-1)abc]$

If H_0 holds, F^* follows the F distribution with $(a-1)(b-1)(c-1)$ degrees of freedom for the numerator and $abc(n-1)$ degrees of freedom for the denominator. Hence, the decision rule to control the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; (a-1)(b-1)(c-1), (n-1)abc], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; (a-1)(b-1)(c-1), (n-1)abc], \text{ conclude } H_a \end{aligned} \tag{24.24c}$$

Table 24.6 contains the test statistics and percentiles of the F distribution for the various tests in a three-factor study.

Kimball Inequality. The Kimball inequality for the family level of significance α in a three-factor study when the family consists of the combined set of seven tests, including three on main effects, three on two-factor interactions, and one on three-factor interactions, is:

$$\alpha < 1 - (1-\alpha_1)(1-\alpha_2)\cdots(1-\alpha_7) \tag{24.25}$$

where α_i is the level of significance for the ith test.

Comments

1. If the three-factor interactions (and also perhaps some sets of two-factor interactions) equal zero, the question sometimes arises whether the corresponding sums of squares should be pooled with the error sum of squares. Our earlier discussion on revising the ANOVA model in Section 19.10 is applicable here also.

2. If there is only one case per treatment in a three-factor study with fixed factor levels, analysis of variance tests can only be conducted if it is possible to assume that some interactions equal zero. Usually, the interactions most likely to equal zero are the three-factor interactions. If it is possible to assume that all three-factor interactions equal zero, *MSABC* has expectation σ^2 and plays the role of the error mean square *MSE*. All mean squares are calculated in the usual manner, except that $n = 1$.

3. The F^* test statistics in Table 24.6 can be obtained by the general linear test approach explained in Chapter 2. For example, for testing whether all three-factor interactions are zero, the full model is that in (24.14), the alternatives are those in (24.24a), and the reduced model under H_0: $(\alpha\beta\gamma)_{ijk} \equiv 0$ is:

$$Y_{ijkm} = \mu_{\cdots} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \varepsilon_{ijkm} \qquad \text{Reduced model} \qquad (24.26)$$

■

Example

In the stress test example, the researcher first wished to test for the various factor effects. Figure 24.7 contains a portion of the SYSTAT ANOVA output. The researcher desired to conduct the seven potential tests with a family level of significance of $\alpha = .10$. This will ensure that if in fact no factor effects are present, there will be only one chance in 10 for one or more of the seven tests to lead to the conclusion of the presence of factor effects. Using the Kimball inequality (24.25), the researcher solved the equation:

$$\alpha = .10 = 1 - (1 - \alpha_i)^7$$

and found $\alpha_i = .015$. Thus, use of significance level $\alpha_i = .015$ for each test ensures that the family level of significance will not exceed .10.

The ANOVA table in Figure 24.7 shows the seven test statistics and their P-values. Each test statistic has in the numerator the appropriate factor effect mean square, and the denominator of each test statistic is *MSE*.

Test for Three-Factor Interactions. The first test was conducted for three-factor interactions. The alternatives are:

$$H_0\text{: all } (\alpha\beta\gamma)_{ijk} = 0$$

$$H_a\text{: not all } (\alpha\beta\gamma)_{ijk} \text{ equal zero}$$

The decision rule is:

$$\text{If } F^* \leq F(.985; 1, 16) = 7.42, \text{ conclude } H_0$$

$$\text{If } F^* > F(.985; 1, 16) = 7.42, \text{ conclude } H_a$$

FIGURE 24.7 SYSTAT ANOVA Output—Stress Test Example.

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATIO	P
GENDER	176.584	1	176.584	18.915	0.000
FAT	242.570	1	242.570	25.984	0.000
SMOKING	70.384	1	70.384	7.539	0.014
GENDER*FAT	13.650	1	13.650	1.462	0.244
GENDER *SMOKING	11.070	1	11.070	1.186	0.292
FAT*SMOKING	72.454	1	72.454	7.761	0.013
GENDER*FAT *SMOKING	1.870	1	1.870	0.200	0.660
ERROR	149.367	16	9.335		

The F^* test statistic obtained from Figure 24.7 is:

$$F^* = \frac{MSABC}{MSE} = \frac{1.870}{9.335} = .20$$

Since $F^* = .20 \leq 7.42$, the researcher concluded that no ABC interactions are present. The P-value of this test is .66.

Tests for Two-Factor Interactions. The researcher next tested for two-factor interactions. In the test for AB interactions, the decision rule is (the alternatives are given in Table 24.6):

$$\text{If } F^* \leq F(.985; 1, 16) = 7.42, \text{ conclude } H_0$$
$$\text{If } F^* > F(.985; 1, 16) = 7.42, \text{ conclude } H_a$$

and the test statistic is:

$$F^* = \frac{MSAB}{MSE} = \frac{13.650}{9.335} = 1.46$$

Since $F^* = 1.46 \leq 7.42$, the researcher concluded that no AB interactions are present. The P-value of this test is .24.

The tests for AC and BC interactions proceeded similarly. We obtain:

1. $F^* = \dfrac{MSAC}{MSE} = \dfrac{11.070}{9.335} = 1.19 \leq F(.985; 1, 16) = 7.42$ $\quad P\text{-value} = .29$

Conclusion: No AC interactions are present.

2. $F^* = \dfrac{MSBC}{MSE} = \dfrac{72.454}{9.335} = 7.76 > F(.985; 1, 16) = 7.42$ $\quad P\text{-value} = .01$

Conclusion: Some BC interactions are present.

Tests for Main Effects. Since factor A (gender) did not interact with the other two factors, attention next turned to testing for factor A main effects. In testing for factor A main effects, the decision rule is (the alternatives are given in Table 24.6):

$$\text{If } F^* \leq F(.985; 1, 16) = 7.42, \text{ conclude } H_0$$
$$\text{If } F^* > F(.985; 1, 16) = 7.42, \text{ conclude } H_a$$

The test statistic is:

$$F^* = \frac{MSA}{MSE} = \frac{176.584}{9.335} = 18.92$$

Since $F^* = 18.92 > 7.42$, the conclusion was reached that factor A main effects are present; specifically, we conclude that the mean endurance time for males is greater than that for females. The P-value of this test is 0+.

The factor B and factor C main effects were not tested at this point because BC interactions were found to be present. The researcher first wished to study the nature of the BC interaction effects before determining whether the factor B and factor C main effects are of any practical interest under the circumstances.

Family of Conclusions. The five separate F tests for factor effects led the researcher to conclude (with family level of significance $\leq .10$):

1. There are no three-factor interactions.
2. There are no two-factor interactions between gender (factor A) and either of the other two factors—body fat (factor B) and smoking history (factor C). Body fat and smoking history interactions do exist, however.
3. Main effects for gender (factor A) are present—mean endurance time for males is larger than for females.

This set of test results was most useful to the researcher. The next step in the analysis was to examine the nature of the BC interaction effects.

24.5 Analysis of Factor Effects

No new problems are encountered in the analysis of factor effects for three-factor studies with fixed factor levels. As for two-factor studies, the focus of the analysis is usually on factor level means when no important interactions are present, and on various two-factor level means ($\mu_{ij\cdot}$, $\mu_{i\cdot k}$, or $\mu_{\cdot jk}$) or individual cell means (μ_{ijk}) when there are important interactions. We first present a formal strategy for determining which level of analysis is appropriate. We then present some selected results for estimating factor effects.

Strategy for Analysis

As described in Section 19.7 for two-factor studies, the presence of interacting effects in multifactor studies complicates the explanation of the factor effects because they must then be described in terms of the combined effects of multiple factors. Of course, some phenomena are too complex to be described simply by additive main effects. The desire for a simple, parsimonious explanation, when possible, suggests the following basic strategy for analyzing factor effects in three-factor studies:

1. Examine whether or not important three-factor interactions exist.
2. If no important three-factor interactions exist, determine whether or not important two-factor interactions are present.
3. If no important two-factor or higher-order interactions are present, examine the main effects. For important A, B, or C main effects, describe the nature of these effects in terms of the factor level means $\mu_{i\cdot\cdot}$, $\mu_{\cdot j\cdot}$ and $\mu_{\cdot\cdot k}$, respectively.
4. If three-factor interactions are important, consider whether they can be made unimportant by a meaningful simple transformation of scale. If so, make the transformation and proceed as in step 2.
5. For important three-factor interactions that cannot be made unimportant by a simple transformation, which is often the case, analyze the three factors jointly in terms of the treatment means μ_{ijk}.
6. If there is just one important two-factor interaction, analyze the effects jointly in terms of the appropriate two-factor treatment means $\mu_{ij\cdot}$, $\mu_{i\cdot k}$, or $\mu_{\cdot jk}$. Analyze the effects of the third factor separately. For example, if the AB interaction is present and no AC or BC interactions exist, analyze the marginal means $\mu_{ij\cdot}$. If a C main effect is present, analyze the single-factor level means $\mu_{\cdot\cdot k}$ separately.

7. If there are two or three important two-factor interactions in a three-factor study, analyze the three factors jointly in terms of the treatment means μ_{ijk}. This principle extends to multifactor studies having more than three factors in the following way. If any two two-factor interactions are overlapping—that is they each involve a common factor—then the cell means should be analyzed in terms of the joint effects of the three factors. For example, if in a four-factor study two interactions AB and BC are found to be important (and no higher-order interactions are present), analysis of the three-factor level means $\mu_{ijk\cdot}$ is indicated.

Occasionally, exceptions to the strategy outlined above may arise. For example, on page 826 we commented on a situation in which an investigator might be interested in inferences concerning a main factor effect even though the factor was also present in an important two-factor interaction.

We have already discussed the testing for interaction effects, the possible diminution of important interactions by a simple transformation, and how to test for the presence of factor main effects. Now we turn to steps 2 through 7 of the strategy for analysis, namely, how to compare single-factor level means $\mu_{i\cdot\cdot}$, $\mu_{\cdot j\cdot}$ and $\mu_{\cdot\cdot k}$ when there are unimportant three-factor and two-factor interactions, how to compare two-factor level means $\mu_{ij\cdot}$, $\mu_{i\cdot k}$, and $\mu_{\cdot jk}$ when there is a single important two-factor interaction, and finally, how to compare treatment means μ_{ijk} when there are important overlapping two factor interactions or important three-factor interactions.

Analysis of Factor Effects when Factors Do Not Interact

Estimation of Factor Level Mean. The factor A level mean $\mu_{i\cdot\cdot}$ is estimated by:

$$\hat{\mu}_{i\cdot\cdot} = \bar{Y}_{i\cdot\cdot\cdot} \tag{24.27}$$

The estimated variance of this estimator is:

$$s^2\{\bar{Y}_{i\cdot\cdot\cdot}\} = \frac{MSE}{nbc} \tag{24.28}$$

Confidence limits for $\mu_{i\cdot\cdot}$ are obtained by means of the t distribution with $(n-1)abc$ degrees of freedom:

$$\bar{Y}_{i\cdot\cdot\cdot} \pm t[1-\alpha/2; (n-1)abc]s\{\bar{Y}_{i\cdot\cdot\cdot}\} \tag{24.29}$$

Estimation of factor level means for factors B or C is done in similar fashion.

Inferences for Contrast of Factor Level Means. Inference procedures for a contrast involving the factor A level means $\mu_{i\cdot\cdot}$:

$$L = \sum c_i \mu_{i\cdot\cdot} \qquad \text{where} \qquad \sum c_i = 0 \tag{24.30}$$

are easily developed. The $1-\alpha$ confidence limits for L are:

$$\hat{L} \pm t[1-\alpha/2; (n-1)abc]s\{\hat{L}\} \tag{24.31}$$

where L is estimated unbiasedly by:

$$\hat{L} = \sum c_i \bar{Y}_{i\cdot\cdot\cdot} \tag{24.31a}$$

and the estimated variance of $\hat{L}$ is:

$$s^2\{\hat{L}\} = \frac{MSE}{nbc}\sum c_i^2 \tag{24.31b}$$

Contrasts of factor level means for factors B or C are estimated in similar fashion.

The test statistic and decision rule for the following alternatives concerning a contrast L in (24.30):

$$\begin{aligned} H_0&: L = 0 \\ H_a&: L \neq 0 \end{aligned} \tag{24.32}$$

are:

$$t^* = \frac{\hat{L}}{s\{\hat{L}\}}; \text{If } |t^*| > t[1-\alpha/2; (n-1)abc], \text{ conclude } H_a \tag{24.33}$$

where $\hat{L}$ and $s\{\hat{L}\}$ are given by (24.31). Again for conciseness, we present only the portion of the decision rule leading to conclusion H_a.

Multiple Contrasts of Factor Level Means. When inferences are to be made concerning a number of contrasts of factor A level means $\mu_{i..}$, the Tukey, Scheffé, and Bonferroni procedures are easily adapted. As before, the Tukey procedure applies to the set of all pairwise comparisons of the form $D = \mu_{i..} - \mu_{i'..}$.

To obtain simultaneous confidence interval estimates, the t multiple in (24.31) is replaced by the T, S, or B multiple defined as follows:

Procedure	Multiple	
Tukey	$T = \frac{1}{\sqrt{2}}q[1-\alpha; a, (n-1)abc]$	(24.34a)
Scheffé	$S^2 = (a-1)F[1-\alpha; a-1, (n-1)abc]$	(24.34b)
Bonferroni	$B = t[1-\alpha/2g; (n-1)abc]$	(24.34c)

Test statistics and decision rules for simultaneous testing of a number of contrasts of the form (24.30) for the alternatives H_0: $L = 0$, H_a: $L \neq 0$ are:

Procedure	Test Statistic and Decision Rule	
Tukey	$q^* = \frac{\sqrt{2}\hat{D}}{s\{\hat{D}\}}$	(24.35a)
	If $\lvert q^*\rvert > q[1-\alpha; a, (n-1)abc]$, conclude H_a	
Scheffé	$F^* = \frac{\hat{L}^2}{(a-1)s^2\{\hat{L}\}}$	(24.35b)
	If $F^* > F[1-\alpha; a-1, (n-1)abc]$, conclude H_a	
Bonferroni	$t^* = \frac{\hat{L}}{s\{\hat{L}\}}$	(24.35c)
	If $\lvert t^*\rvert > t[1-\alpha/2g; (n-1)abc]$, conclude H_a	

Inferences concerning multiple contrasts based on the factor level means $\mu_{\cdot j\cdot}$ or $\mu_{\cdot\cdot k}$ are made in corresponding fashion.

Analysis of Factor Effects with Multiple Two-Factor Interactions or Three-Factor Interaction

As explained earlier in the strategy for analysis, when a three-factor interaction is present or overlapping two-factor interactions are present, the results of the study are typically analyzed in terms of the treatment means μ_{ijk}.

Estimation of Treatment Mean. The treatment mean μ_{ijk} is estimated by:

$$\hat{\mu}_{ijk} = \bar{Y}_{ijk\cdot} \tag{24.36}$$

The estimated variance of $\bar{Y}_{ijk\cdot}$ is:

$$s^2\{\bar{Y}_{ijk\cdot}\} = \frac{MSE}{n} \tag{24.37}$$

Confidence limits for μ_{ijk} are:

$$\bar{Y}_{ijk\cdot} \pm t[1-\alpha/2; (n-1)abc]s\{\bar{Y}_{ijk\cdot}\} \tag{24.38}$$

Inferences for Contrast of Treatment Means. When important interactions are present, contrasts among the treatment means μ_{ijk} are ordinarily desired. Let, as usual, L denote such a contrast:

$$L = \sum\sum\sum c_{ijk}\mu_{ijk} \qquad \text{where} \qquad \sum\sum\sum c_{ijk} = 0 \tag{24.39}$$

Confidence limits for L are:

$$\hat{L} \pm t[1-\alpha/2; (n-1)abc]s\{\hat{L}\} \tag{24.40}$$

where:

$$\hat{L} = \sum\sum\sum c_{ijk}\bar{Y}_{ijk\cdot} \tag{24.40a}$$

$$s^2\{\hat{L}\} = \frac{MSE}{n}\sum\sum\sum c_{ijk}^2 \tag{24.40b}$$

The test statistic and decision rule for alternatives H_0: $L = 0$, H_a: $L \neq 0$ are:

$$t^* = \frac{\hat{L}}{s\{\hat{L}\}}; \text{ If } |t^*| > t[1-\alpha/2; (n-1)abc], \text{ conclude } H_a \tag{24.41}$$

Analysis of Factor Effects with Single Two-Factor Interaction

When a single two-factor interaction is present in a three-factor study, desired contrasts may involve means of the μ_{ijk} taken over one of the factors. For example, when the only interactions present are the BC interactions, there may be interest in contrasts of the

means $\mu_{\cdot jk}$:

$$L = \sum\sum c_{jk}\mu_{\cdot jk} \qquad \text{where} \qquad \sum\sum c_{jk} = 0 \tag{24.42}$$

Such contrasts are, of course, special cases of contrasts of the treatment means μ_{ijk} in (24.39). The estimator of the contrast in (24.42) can be obtained from (24.40a) and the estimated variance from (24.40b); they are:

$$\hat{L} = \sum\sum c_{jk}\bar{Y}_{\cdot jk}. \tag{24.43}$$

$$s^2\{\hat{L}\} = \frac{MSE}{na}\sum\sum c_{jk}^2 \tag{24.44}$$

Multiple Contrasts of Treatment Means. For simultaneous interval estimates of contrasts of treatment means μ_{ijk}, the t multiple in (24.40) is replaced by the T, S, or B multiple defined as follows:

Procedure	Multiple	
Tukey	$T = \frac{1}{\sqrt{2}}q[1-\alpha;\ ABC,\ (n-1)abc]$	(24.45a)
Scheffé	$S^2 = (abc-1)F[1-\alpha;\ abc-1,\ (n-1)abc]$	(24.45b)
Bonferroni	$B = t[1-\alpha/2g;\ (n-1)abc]$	(24.45c)

Simultaneous testing of a number of alternatives of the form H_0: $L = 0$, H_a: $L \neq 0$ using the Tukey, Scheffé, and Bonferroni procedures can be accomplished with the following test statistics and decision rules:

Procedure	Test Statistic and Decision Rule	
Tukey	$q^* = \frac{\sqrt{2}\hat{D}}{s\{\hat{D}\}}$	(24.46a)
	If $\lvert q^*\rvert > q[1-\alpha;\ ABC,\ (n-1)abc]$, conclude H_a	
Scheffé	$F^* = \frac{\hat{L}^2}{(abc-1)s^2\{\hat{L}\}}$	(24.46b)
	If $F^* > F[1-\alpha;\ abc-1,\ (n-1)abc]$, conclude H_a	
Bonferroni	$t^* = \frac{\hat{L}}{s\{\hat{L}\}}$	(24.46c)
	If $\lvert t^*\rvert > t[1-\alpha/2g;\ (n-1)abc]$, conclude H_a	

As before, the Tukey procedure concerns only pairwise comparisons.

Example—Estimation of Contrasts of Treatment Means

To study the nature of the BC interaction effects in the stress test example, the researcher wished to estimate separately, for persons with high and low body fat, the difference in mean fatigue time for light smokers and heavy smokers. The desired contrasts are:

$$L_1 = \mu_{\cdot 11} - \mu_{\cdot 12}$$
$$L_2 = \mu_{\cdot 21} - \mu_{\cdot 22}$$

In addition, a single comparison between the factor level means for factor A is sufficient to analyze the factor A main effects since factor A has only two levels. The contrast of interest (here a pairwise comparison of factor level means) is:

$$L_3 = \mu_{1\cdot\cdot} - \mu_{2\cdot\cdot}$$

These three contrasts are estimated as follows, using the results in Table 24.4b:

$$\hat{L}_1 = \bar{Y}_{\cdot 11\cdot} - \bar{Y}_{\cdot 12\cdot} = 22.90 - 16.00 = 6.90$$
$$\hat{L}_2 = \bar{Y}_{\cdot 21\cdot} - \bar{Y}_{\cdot 22\cdot} = 13.07 - 13.12 = -.05$$
$$\hat{L}_3 = \bar{Y}_{1\cdot\cdot\cdot} - \bar{Y}_{2\cdot\cdot\cdot} = 18.98 - 13.56 = 5.42$$

The researcher obtained the estimated variances by using (24.44) and (24.31b) and the Bonferroni multiple for a 95 percent family confidence coefficient:

$$s^2\{\hat{L}_1\} = s^2\{\hat{L}_2\} = \frac{MSE}{na}[(1)^2 + (-1)^2] = \frac{9.335}{6}(2) = 3.112$$
$$s^2\{\hat{L}_3\} = \frac{MSE}{nbc}[(1)^2 + (-1)^2] = \frac{9.335}{12}(2) = 1.556$$
$$s\{\hat{L}_1\} = s\{\hat{L}_2\} = 1.764 \qquad s\{\hat{L}_3\} = 1.247$$
$$B = t(1 - .05/6; 16) = 2.673$$

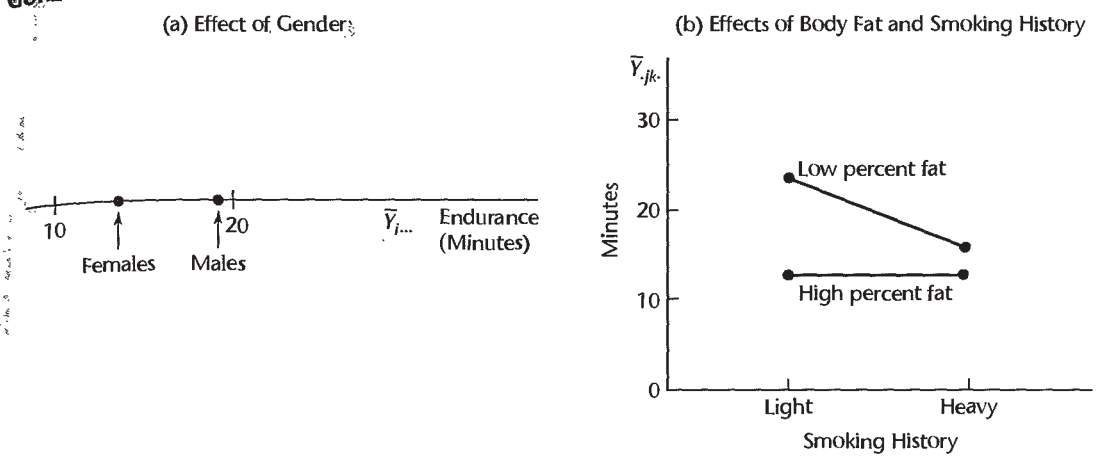
The desired confidence intervals using (24.40) therefore are:

$$2.2 = 6.90 - 2.673(1.764) \leq \mu_{\cdot 11} - \mu_{\cdot 12} \leq 6.90 + 2.673(1.764) = 11.6$$
$$-4.8 = -.05 - 2.673(1.764) \leq \mu_{\cdot 21} - \mu_{\cdot 22} \leq -.05 + 2.673(1.764) = 4.7$$
$$2.1 = 5.42 - 2.673(1.247) \leq \mu_{1\cdot\cdot} - \mu_{2\cdot\cdot} \leq 5.42 + 2.673(1.247) = 8.8$$

The researcher therefore concluded with family confidence coefficient .95: (1) Among people with low body fat, those who have a light smoking history have a mean stress test endurance that is 2.2 to 11.6 minutes longer than the mean endurance for people with a heavy smoking history. (2) People with high body fat do not differ in mean stress test endurance whether they have a light or a heavy smoking history. (3) The mean stress test endurance for men is 2.1 to 8.8 minutes longer than the mean endurance for women.

In view of the important interaction effects between body fat and smoking history on stress test endurance noted in the study findings, the researcher concluded that factor B and factor C main effects are of no interest, and therefore terminated the analysis at this

FIGURE 24.8 Key Findings from Stress Test Endurance Study.



point. The principal findings are presented graphically in Figure 24.8. Figure 24.8a shows the magnitude of the effect of gender on stress test endurance, and Figure 24.8b shows the nature of the interaction effects between body fat and smoking history on stress test endurance.

24.6 Unequal Sample Sizes in Multi-Factor Studies

When the treatment sample sizes in a multi-factor study are not equal, the procedures explained in Sections 23.1–23.3 for two-factor studies with unequal treatment sample sizes should be followed with routine modifications. We continue to assume that *all treatment means are of equal importance and that there are no empty cells.*

Tests for Factor Effects

Tests for factor effects in multifactor studies with unequal sample sizes can be conducted by means of the regression approach. Indicator variables taking on the values 1, −1, 0, are designated for each factor, the number of such variables for each factor being one less than the number of factor levels. Interaction effects are represented by cross-product terms, as usual. Since the sums of squares are no longer orthogonal when the treatment sample sizes are unequal, different reduced models need to be fitted for the tests of interest.

Example

Suppose that in the stress test example of Table 24.4, observations Y_{1113} and Y_{2212} were missing. To develop a regression model for this example, we note that each of the three factors is at two levels. Hence, one indicator variable is required for each factor. The full regression model therefore is:

$$\begin{aligned} Y_{ijkm} = \mu_{...} &+ \alpha_1 X_{ijkm1} + \beta_1 X_{ijkm2} + \gamma_1 X_{ijkm3} + (\alpha\beta)_{11} X_{ijkm1} X_{ijkm2} \\ &+ (\alpha\gamma)_{11} X_{ijkm1} X_{ijkm3} + (\beta\gamma)_{11} X_{ijkm2} X_{ijkm3} \\ &+ (\alpha\beta\gamma)_{111} X_{ijkm1} X_{ijkm2} X_{ijkm3} + \varepsilon_{ijkm} \qquad \text{Full model} \end{aligned} \tag{24.47}$$

TABLE 24.7 Data for Regression Model (24.47)—Stress Test Example with Y_{1113} and Y_{2212} Missing.

i	j	k	m	(1) Y	(2) X_1	(3) X_2	(4) X_3	(5) X_1X_2	(6) X_1X_3	(7) X_2X_3	(8) $X_1X_2X_3$
1	1	1	1	24.1	1	1	1	1	1	1	1
1	1	1	2	29.2	1	1	1	1	1	1	1
1	1	2	1	17.6	1	1	−1	1	−1	−1	−1
	...			...	...	...	...	...	...	...	...
2	2	1	1	16.1	−1	−1	1	1	−1	−1	1
2	2	1	3	10.8	−1	−1	1	1	−1	−1	1
2	2	2	1	10.1	−1	−1	−1	1	1	1	−1
2	2	2	2	14.4	−1	−1	−1	1	1	1	−1
2	2	2	3	6.1	−1	−1	−1	1	1	1	−1

where:

$$X_1 = \begin{cases} 1 & \text{if case from level 1 for factor } A \\ -1 & \text{if case from level 2 for factor } A \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if case from level 1 for factor } B \\ -1 & \text{if case from level 2 for factor } B \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{if case from level 1 for factor } C \\ -1 & \text{if case from level 2 for factor } C \end{cases}$$

The regression parameters in model (24.47) are the ANOVA model parameters as defined in (24.13).

Table 24.7 repeats in column 1 a portion of the Y observations for the stress test example in Table 24.4 with observations Y_{1113} and Y_{2212} missing. The coded indicator variables X_1, X_2, and X_3 are shown in columns 2–4 and the cross-product interaction terms are shown in columns 5–8. The full model in (24.47) is fitted by regressing Y in column 1 of Table 24.7 on the X variables in columns 2–8. To test a particular factor effect, the reduced model is obtained by dropping the appropriate X variable(s). For instance, to test for factor A main effects, X_1 would be dropped to obtain the reduced model and Y would be regressed on the X variables in columns 3–8.

Comment

The discussion in Section 23.6 on the use of statistical packages for analysis of variance with unequal sample sizes and/or empty cells is applicable in its entirety for multifactor studies. ■

Inferences for Contrasts of Factor Level Means

Estimation and testing of contrasts of factor level means in multi-factor studies with unequal sample sizes are conducted in similar fashion as for two-factor studies. The formulas in Table 23.5 for the development of interval estimates need simply be extended to three or more factors. Testing procedures may be devised from these extensions in the usual fashion.

To illustrate such an extension, consider pairwise comparisons of factor A level means in a three-factor study with unequal samples sizes. Extending formula (23.21), we obtain

for the comparison, its estimator, and the estimated variance:

$$D = \mu_{i\cdot\cdot} - \mu_{i'\cdot\cdot} \tag{24.48a}$$

$$\hat{D} = \hat{\mu}_{i\cdot\cdot} - \hat{\mu}_{i'\cdot\cdot} \qquad \text{where} \qquad \hat{\mu}_{i\cdot\cdot} = \frac{\sum_j \sum_k \bar{Y}_{ijk\cdot}}{bc} \tag{24.48b}$$

$$s^2\{\hat{D}\} = \frac{MSE}{b^2c^2} \sum_j \sum_k \left(\frac{1}{n_{ijk}} + \frac{1}{n_{i'jk}} \right) \tag{24.48c}$$

The appropriate degrees of freedom associated with MSE are $n_T - abc$.

24.7 Planning of Sample Sizes

We considered the planning of sample sizes for single-factor studies with power approach and estimation approach in Chapters 16 and 17. Then we considered the planning of sample sizes for two-factor studies in Chapter 19. Now we take up the planning of samples sizes for multi-factor studies.

Power of *F* Test for Multi-Factor Studies

Table B.11 can be used for determining the power of tests for multi-factor studies in the same fashion as for single-factor and two-factor studies. The only differences arise in the definition of the noncentrality parameter and the degrees of freedom. For three-factor fixed effects ANOVA model (24.14) with equal treatment sample sizes, the noncentrality parameter ϕ for a given test is defined as follows:

$$\phi = \frac{1}{\sigma} \left[\frac{\text{numerator of second term in } E\{MS\} \text{ in Table 24.5}}{\text{denominator of second term in } E\{MS\} \text{ plus 1}} \right]^{1/2} \tag{24.49}$$

For example, for testing for three-factor interactions, we have:

$$\phi = \frac{1}{\sigma} \left[\frac{n \sum\sum\sum (\alpha\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)+1} \right]^{1/2}$$

Use of Table B.12 for Multi-Factor Studies

When planning sample sizes for three-factor studies with the power approach, one is typically concerned with the power of detecting factor A main effects, the power of detecting factor B main effects, and the power of detecting factor C main effects. One can first specify the minimum range of factor A level means for which it is important to detect factor A main effects and obtain the needed sample sizes from Table B.12, with $r = a$. The resulting sample size is bcn, from which n can be obtained readily. The use of Table B.12 for this purpose is appropriate provided the resulting sample sizes are not small, specifically provided $a(bcn - 1) \geq 20$. If this condition is not met, the ANOVA power tables in Table B.11 should be used with an iterative approach.

In the same way, the values for the minimum range of factor level means for factors B and C can be specified for which it is important to detect the factor main effects, and the needed sample sizes found. If the sample sizes obtained from the factor A, factor B, and

factor C power specifications differ substantially, a judgment will need to be made as to the final sample sizes.

Cited Reference

24.1. Monlezun, C. J. "Two-Dimensional Plots for Interpreting Interactions in the Three-Factor Analysis of Variance Model." *The American Statistician* 33 (1989), pp. 63–69.

Problems

24.1. Refer to Table 24.1 containing the mean responses μ_{ijk} for a three-factor study.

a. Find the main effects of age.

b. Find the interaction effect of young age and normal IQ.

c. Find the interaction effect of young age, normal IQ, and female gender.

24.2. Prepare AC plots of the mean responses μ_{ijk} in Table 24.1 in the format of Figures 24.2c–e. Do your plots convey the same information as Figure 24.1? Discuss.

24.3. Prepare BC plots of the mean responses μ_{ijk} in Table 24.1. Do your plots bring out any information on main effects and interactions not readily seen from Figure 24.1? Discuss.

24.4. In a three-factor study, the mean responses μ_{ijk} are as follows:

	$k=1$		$k=2$	
	$j=1$	$j=2$	$j=1$	$j=2$
$i=1$	130	138	140	144
$i=2$	126	130	134	136
$i=3$	122	125	122	131

a. Find α_1, α_2, and α_3.

b. Find β_2 and γ_1.

c. Find $(\alpha\beta)_{12}$, $(\alpha\gamma)_{21}$, and $(\beta\gamma)_{12}$.

d. Find $(\alpha\beta\gamma)_{111}$ and $(\alpha\beta\gamma)_{322}$.

24.5. Refer to Problem 24.4. Prepare AB plots of the mean responses μ_{ijk} in the format of Figure 24.1. What do these plots show about factor main effects and interactions?

*24.6. **Case hardening.** An experiment involving the case hardening of lightweight shafts machined from bars of an alloy was run to study the effects of the amount of a chemical agent added to the alloy in a molten state (factor A), the temperature of the hardening process (factor B), and the time duration of the hardening process (factor C) on the outside hardness of the shaft. All factors were at two levels (1: low, 2: high), and the number of rods tested for each treatment was $n = 3$. The data on hardness (in Brinell units) follow.

	$k=1$		$k=2$	
	$j=1$	$j=2$	$j=1$	$j=2$
$i=1$	39.9	53.5	56.0	70.9
	32.2	50.7	56.9	73.3
	36.3	52.8	56.6	71.6
$i=2$	45.2	63.3	69.4	82.9
	48.0	65.5	66.6	85.2
	47.5	63.6	68.8	82.3

a. Obtain the residuals for ANOVA model (24.14) and prepare aligned residual dot plots for each level of factor A. Do the same for each of the other two factors. What information do these plots provide about the appropriateness of ANOVA model (24.14)?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

Refer to **Case hardening** Problem 24.6. Assume that fixed ANOVA model (24.14) is appropriate.

a. Prepare AB plots of the estimated treatment means $\bar{Y}_{ijk\cdot}$ in the format of Figure 24.5b. Does it appear that any interactions are present? Any main effects?

b. Obtain the analysis of variance table.

c. Test for three-factor interactions; use $\alpha = .025$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test for AB, AC, and BC interactions. For each test, use $\alpha = .025$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

e. Test for A, B, and C main effects. For each test, use $\alpha = .025$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

f. State the set of conclusions that can be reached from the tests in parts (c), (d), and (e). Obtain an upper bound for the family level of significance for the set of tests; use the Kimball inequality (24.25).

g. Do the results in part (f) confirm your graphic analysis in part (a)?

Refer to **Case hardening** Problems 24.6 and 24.7.

a. To study the nature of the main factor effects, estimate the following pairwise comparisons:

$$D_1 = \mu_{2\cdot\cdot} - \mu_{1\cdot\cdot} \qquad D_2 = \mu_{\cdot 2\cdot} - \mu_{\cdot 1\cdot} \qquad D_3 = \mu_{\cdot\cdot 2} - \mu_{\cdot\cdot 1}$$

Use the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

b. Estimate μ_{222} with a 95 percent confidence interval.

Marketing research contractors. A marketing research consultant evaluated the effects of fee schedule (factor A), scope of work (factor B), and type of supervisory control (factor C) on the quality of work performed under contract by independent marketing research agencies. The factor levels in the study were as follows:

Factor		Factor Levels	
A	Fee level	$i = 1$:	High
		$i = 2$:	Average
		$i = 3$:	Low
B	Scope	$j = 1$:	All contract work performed in house
		$j = 2$:	Some work subcontracted out
C	Supervision	$k = 1$:	Local supervisors
		$k = 2$:	Traveling supervisors only

The quality of work performed was measured by an index taking into account several characteristics of quality. Four agencies were chosen for each factor level combination and the quality of their work evaluated. The data on quality follow.

	$k=1$		$k=2$	
	$j=1$	$j=2$	$j=1$	$j=2$
$i=1$	124.3	115.1	112.7	88.2
	...	..	...	...
	122.6	117.3	108.6	90.1
$i=2$	119.3	117.2	113.6	92.7
	...	...	...	.
	121.4	120.0	112.3	87.9
$i=3$	90.9	89.9	78.6	58.6
	.	...	...	
	92.0	82.7	77.1	62.3

a. Obtain the residuals for ANOVA model (24.14) and plot them against the fitted values. What does your plot suggest about the appropriateness of ANOVA model (24.14)?

b. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

24.10. Refer to **Marketing research contractors** Problem 24.9. Assume that fixed ANOVA model (24.14) is appropriate.

a. Prepare AB plots of the estimated treatment means $\bar{Y}_{ijk\cdot}$ in the format of Figure 24.5b. Does it appear that any interactions are present? Any main effects?

b. Prepare AC plots of the estimated treatment means $\bar{Y}_{ijk\cdot}$ in the format of Figure 24.5b. Do your plots convey the same information as those in part (a)? Discuss.

c. Obtain the analysis of variance table.

d. Test for three-factor interactions: use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Test for AB, AC, and BC interactions. For each test, use $\alpha = .01$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

f. Test for factor A main effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

g. State the set of conclusions that can be reached from the tests in parts (d), (e), and (f). Obtain an upper bound for the family level of significance for the set of tests; use the Kimball inequality (24.25).

h. Do the results in part (g) confirm your graphic analysis in parts (a) and (b)?

24.11. Refer to **Marketing research contractors** Problems 24.9 and 24.10.

a. To study the nature of the factor A main effects and the BC interactions, it is desired to estimate the following comparisons:

$$D_1 = \mu_{1\cdot\cdot} - \mu_{2\cdot\cdot} \qquad D_4 = \mu_{\cdot 11} - \mu_{\cdot 12}$$
$$D_2 = \mu_{2\cdot\cdot} - \mu_{3\cdot\cdot} \qquad D_5 = \mu_{\cdot 21} - \mu_{\cdot 22}$$
$$D_3 = \mu_{1\cdot\cdot} - \mu_{3\cdot\cdot} \qquad L_1 = D_4 - D_5$$

Use the Bonferroni procedure with a 90 percent family confidence coefficient to make the desired comparisons. State your findings.

b. Estimate $D = \mu_{121} - \mu_{221}$ with a 95 percent confidence interval.

c. The consultant wishes to identify the type(s) of independent marketing research agencies that provide the highest quality of work. Use the Tukey testing procedure with family level of significance $\alpha = .10$ to make the desired identifications.

Electronics assembly. Assemblers in an electronics firm will attach 12 components to a newly developed "board" that will be used in automatic-control equipment in manufacturing plants. An operations analyst conducted an experiment to study the effects of three factors on the mean time to assemble a board. Factor A was the gender of the assembler ($i = 1$: male; $i = 2$: female), factor B was the sequence of assembling the components ($j = 1, 2, 3$), and factor C was the amount of experience by the assembler ($k = 1$: under 18 months; $k = 2$: 18 months or more). Randomization was used to assign 15 assemblers of each gender with a given amount of experience to each of the three assembly sequences, with each sequence assigned to five assemblers. After a learning period, the total time (in minutes) to assemble 50 boards was observed. The data follow.

	$k = 1$			$k = 2$		
	$j = 1$	$j = 2$	$j = 3$	$j = 1$	$j = 2$	$j = 3$
$i = 1$	1,250	1,319	1,217	1,021	1,119	1,033
	1,175	1,251	1,190	1,099	1,110	1,067
	...	...	...	...	...	...
	1,193	1,265	1,251	1,070	1,163	1,022
$i = 2$	1,066	1,105	1,021	864	927	841
	1,076	1,043	1,020	848	944	865
	...	...	...	...	...	...
	1,034	1,060	1,026	868	933	868

d. Obtain the residuals for ANOVA model (24.14) and plot them against the fitted values. What does your plot suggest about the appropriateness of ANOVA model (24.14)?

e. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear to be reasonable here?

Refer to **Electronics assembly** Problem 24.12. Assume that fixed ANOVA model (24.14) is appropriate.

a. Prepare AB plots of the estimated treatment means $\bar{Y}_{ijk\cdot}$ in the format of Figure 24.5b. Does it appear that any interactions are present? Any main effects?

b. Obtain the analysis of variance table.

c. Test for three-factor interactions; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test for AB, AC, and BC interactions. For each test, use $\alpha = .05$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

e. Test for A, B, and C main effects. For each test, use $\alpha = .05$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

f. State the set of conclusions that can be reached from the tests in parts (c), (d), and (e). Obtain an upper bound for the family level of significance for the set of tests; use the Kimball inequality (24.25).

g. Do the results in part (f) confirm your graphic analysis in part (a)?

24.14. Refer to **Electronics assembly** Problems 24.12 and 24.13.

a. To study the nature of the factor main effects, estimate the following pairwise comparisons:

$$D_1 = \mu_{1\cdot\cdot} - \mu_{2\cdot\cdot} \qquad D_4 = \mu_{\cdot 2\cdot} - \mu_{\cdot 3\cdot}$$
$$D_2 = \mu_{\cdot 1\cdot} - \mu_{\cdot 2\cdot} \qquad D_5 = \mu_{\cdot\cdot 1} - \mu_{\cdot\cdot 2}$$
$$D_3 = \mu_{\cdot 1\cdot} - \mu_{\cdot 3\cdot}$$

Use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

b. Estimate μ_{231} with a 95 percent confidence interval.

*24.15. Refer to **Case hardening** Problem 24.6. Suppose that observations $Y_{1211} = 53.5$ and $Y_{1212} = 50.7$ are missing.

a. State the full regression model equivalent to ANOVA model (24.14); use 1, −1, 0 indicator variables.

b. What is the reduced regression model for testing for factor A main effects?

c. Test whether or not factor A main effects are present by fitting the full and reduced regression models; use $\alpha = .025$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Estimate $D = \mu_{2\cdot\cdot} - \mu_{1\cdot\cdot}$ with a 95 percent confidence interval.

24.16. Refer to **Electronics assembly** Problem 24.12. Suppose that observations $Y_{1224} = 1{,}097$, $Y_{2213} = 1{,}051$, and $Y_{2125} = 868$ are missing.

a. State the full regression model equivalent to ANOVA model (24.14); use 1, −1, 0 indicator variables.

b. What is the reduced regression model for testing for factor C main effects?

c. Test whether or not factor C main effects are present by fitting the full and reduced regression models; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Estimate $D = \mu_{\cdot\cdot 1} - \mu_{\cdot\cdot 2}$ with a 95 percent confidence interval.

*24.17. Refer to **Case hardening** Problem 24.6. Suppose that the sample sizes have not yet been determined but it has been decided to use equal sample sizes for all treatments. The chief objective is to identify the treatment that leads to the highest mean hardness. The probability should be at least .99 that the correct treatment is identified when the mean hardness for the second best treatment differs by 2.0 or more Brinell units. Assume that a reasonable planning value for the error standard deviation is $\sigma = 1.8$. What are the required sample sizes?

24.18. Refer to **Electronics assembly** Problem 24.12. Suppose that the sample sizes have not yet been determined but it has been decided to use equal sample sizes for all treatments. The chief objective is to estimate the following pairwise comparisons:

$$L_1 = \mu_{1\cdot\cdot} - \mu_{2\cdot\cdot} \qquad L_4 = \mu_{\cdot 2\cdot} - \mu_{\cdot 3\cdot}$$
$$L_2 = \mu_{\cdot 1\cdot} - \mu_{\cdot 2\cdot} \qquad L_5 = \mu_{\cdot\cdot 1} - \mu_{\cdot\cdot 2}$$
$$L_3 = \mu_{\cdot 1\cdot} - \mu_{\cdot 3\cdot}$$

What are the required sample sizes if the precision of each of the estimates should not exceed ±20, using the Bonferroni procedure with a 90 percent family confidence coefficient for the joint set of comparisons? A reasonable planning value for the error standard deviation is $\sigma = 29$.

ercises

24.19. For fixed ANOVA model (24.14), show that $\sum_i (\alpha\beta\gamma)_{ijk} = 0$.

24.20. State the fixed ANOVA model for a three-factor study with $n = 1$ when all three-factor interactions are zero. Show the ANOVA table for this case.

24.21. For fixed ANOVA model (24.14), derive the variance of the estimated contrast $\hat{L} = \sum\sum c_{ij}\overline{Y}_{ij\cdot\cdot}$.

ojects

24.22. Refer to the **SENIC** data set in Appendix C.1. The following hospitals are to be considered in a study of the effects of average age of patients (factor A: variable 3), available facilities and services (factor B: variable 12), and region (factor C: variable 9) on the mean length of hospital stay of patients (variable 2):

1–14	16–28	31	32	34	35	37–39	41	44	46	50
52	53	57	58	63	66	76	77	83	111	

For purposes of this ANOVA study, average age is to be classified into two categories (less than 53.0 years, 53.0 years or more) and available facilities and services are to be classified into two categories (less than 40.2 percent, 40.2 percent or more).

a. Assemble the required data and obtain the residuals for ANOVA model (24.14).

b. Plot the residuals against the fitted values. What does your plot suggest about the appropriateness of ANOVA model (24.14)?

c. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear reasonable here?

24.23. Refer to the **SENIC** data set in Appendix C.1 and Project 24.22. Assume that fixed ANOVA model (24.14) is appropriate.

a. Prepare AB interaction plots of the estimated treatment means $\overline{Y}_{ijk\cdot}$ in the format of Figure 24.5b. Does it appear that any factor effects are present? Explain.

b. Obtain the analysis of variance table. Does any one source account for most of the total variability in the study? Explain.

c. Test for three-factor interactions; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test for AB, AC, and BC interactions. For each test, use $\alpha = .01$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

e. Test for A, B, and C main effects. For each test, use $\alpha = .01$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

f. To study the nature of the available facilities and region main effects, make all pairwise comparisons for each of these two factors. Use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

24.24. Refer to the **CDI** data set in Appendix C.2. The effects of region (factor A: variable 17), percent below poverty level (factor B: variable 13), and percent of population 65 or older (factor C: variable 7) on the crime rate (variable 10 ÷ variable 5) are to be studied. For purposes of this ANOVA study, percent below poverty level is to be classified into two categories (less than 8.0 percent, 8.0 percent or more) and percent of population 65 or older is to be classified into two categories (less than 12.0 percent, 12.0 percent or more).

a. Assemble the required data and obtain the residuals for ANOVA model (24.14) with $m = 1, \ldots, n_{ijk}$.

b. Plot the residuals against the fitted values. What does your plot suggest about the appropriateness of ANOVA model (24.14)?

c. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. Does the normality assumption appear reasonable here?

24.25. Refer to the **CDI** data set in Appendix C.2 and Project 24.24. Assume that fixed ANOVA model (24.14) with $m = 1, \ldots, n_{ijk}$ is appropriate.

a. Prepare AB interaction plots of the estimated treatment means $\bar{Y}_{ijk\cdot}$ in the format of Figure 24.5b. Does it appear that any factor effects are present?

b. State the equivalent regression model for this case; use $1, -1, 0$ indicator variables, and fit this full model.

c. Test for three-factor interactions and for AB, AC, and BC interactions. For each test, use $\alpha = .025$ and state the alternatives, reduced regression model, decision rule, and conclusion. What is the P-value of each test?

d. Test for A, B, and C main effects. For each test, use $\alpha = .025$ and state the alternatives, reduced regression model, decision rule, and conclusion. What is the P-value of each test?

e. To study the nature of the region main effects, make all pairwise comparisons between the region means. Use the Tukey procedure with a 95 percent family confidence coefficient. State your findings.

Case Studies

24.26. Refer to the **Real estate sales** data set in Appendix C.7. Assume that the sample sizes do not reflect the importance of the treatment means. Carry out an unbalanced three-way analysis of variance of this data set, where the response of interest is sales price (variable 2), and the three crossed factors are quality (variable 10), style (variable 11), and number of bedrooms (variable 4). Recode quality into two categories: 1–2, and 3. Recode the number of bedrooms into three categories: 0–2, 3, and 4 or more. Recode style as either 1 or not 1. The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

24.27. Refer to the **Real estate sales** data set in Appendix C.7 and Case Study 24.26. Assume that the sample sizes reflect the importance of the treatment means. Carry out an unbalanced three-way analysis of variance of this data set, where the response of interest is sales price (variable 2), and the three crossed factors are quality (variable 10), style (variable 11), and number of bedrooms (variable 4). Recode quality into two categories: 1–2, and 3. Recode the number of bedrooms into three categories: 0–2, 3, and 4 or more. Recode style as either 1 or not 1. The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

24.28. Refer to the **Ischemic heart disease** data set in Appendix C.9. Assume that the sample sizes do not reflect the importance of the treatment means. Carry out an unbalanced three-way analysis of variance of this data set, where the response of interest is total cost (variable 2), and the three crossed factors are gender (variable 4), number of interventions (variable 5), and number of comorbidities (variable 9). Recode the number of interventions into three categories: 0–1, 2–4, and greater than or equal to 5. Recode the number of comorbidities into two categories: 0–1, and greater than or equal to 2. The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

24.29. Refer to the **Ischemic heart disease** data set in Appendix C.9 and Case Study 24.28. Assume that the sample sizes reflect the importance of the treatment means. Carry out an unbalanced three-way analysis of variance of this data set, where the response of interest is total cost (variable 2), and the three crossed factors are gender (variable 4), number of interventions (variable 5) and number of comorbidities (variable 9). Recode the number of interventions into three categories: 0–1, 2–4, and greater than or equal to 5. Recode the number of comorbidities into two categories: 0–1, and greater than or equal to 2. The analysis should consider transformations of the response variable. Document the steps taken in your analysis and justify your conclusions.

Chapter 25

Random and Mixed Effects Models

Until now, we have been concerned exclusively with ANOVA model I in which the factor levels are considered fixed. This model is applicable for studies where our interest centers on the effects of the specific factor levels chosen. There are still other studies where the factor levels are a sample from a larger population of potential factor levels and inferences are desired about the populations of factor levels. For example, in Section 16.3 we described a single-factor study by a company that owns several hundred retail stores. Seven of these stores were selected at random, and a sample of employees in each store was asked to evaluate the management of the store. The seven stores chosen for the study constitute the seven levels of the random factor, retail stores. In this case, management was not just interested in the management of the seven stores chosen; it wanted to generalize the results to the entire population of stores. Because the retail stores were selected at random, the factor retail stores in this example is considered a random factor. Random factors may also be present in two-factor and multi-factor studies; either all of the factors may be random or some may be random and some fixed. For instance, suppose in the previous example that eight employees were selected at random from each of the five departments in each of the stores. Interest now is in the employee evaluations of management by department and store. Here, stores would be a random factor because the seven selected stores are a sample of all stores. On the other hand, departments would be a fixed factor because there are only five departments in each store and interest is in these five departments.

Analysis of variance models for studies in which all factors are random are called ANOVA models II and those for studies in which some factors are random and some fixed are called ANOVA models III. In Sections 25.1 to 25.4 and 25.6, we consider ANOVA model II for single-factor studies and ANOVA models II and III for two-factor and three-factor studies. Completely randomized block designs with random block effects are taken up in Section 25.5. Throughout Sections 25.1 to 25.6, we assume that all treatment sample sizes are equal. In Section 25.7 we consider studies where the treatment sample sizes are unequal. We begin our discussion with random ANOVA model II for single-factor studies.

25.1 Single-Factor Studies—ANOVA Model II

As we noted earlier, there are occasions when the factor levels or treatments in a single-factor study are not of intrinsic interest in themselves but constitute a sample from a larger population of factor levels. ANOVA model II is designed for this type of situation. Consider, for instance, Apex Enterprises, a company that builds roadside restaurants carrying one of several promoted trade names, leases franchises to individuals to operate the restaurants, and provides management services. This company employs a large number of personnel officers who interview applicants for jobs in the restaurants. At the end of an interview, the personnel officer assigns a rating between 0 and 100 to indicate the applicant's potential value on the job. Five personnel officers were selected at random, and each was assigned four candidates at random. In this case, the company did not wish to make inferences concerning the five personnel officers who happened to be selected but rather about the population of all personnel officers. Questions of interest included: How great is the variation in ratings among all personnel officers? What is the mean rating by all personnel officers?

The distinction between this situation, for which ANOVA model II is designed, and one where fixed ANOVA model I is appropriate can be seen readily by modifying our example slightly. If a smaller company had only five personnel officers who were all included in the study and interest is limited to these five officers, ANOVA model I would be relevant since the factor levels (the five personnel officers) would then not be considered a sample from a larger population. A repetition of the experiment for the smaller company would involve the same five personnel officers, but in the case of Apex Enterprises a repetition would involve a new random sample of five personnel officers which would probably consist of different officers.

Random Cell Means Model

The cell means version of ANOVA model II for single-factor studies is as follows when all factor level sample sizes are equal, i.e., when $n_i \equiv n$:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \tag{25.1}$$

where:

μ_i are independent $N(\mu., \sigma_\mu^2)$
ε_{ij} are independent $N(0, \sigma^2)$
μ_i and ε_{ij} are independent random variables
$i = 1, \ldots, r; j = 1, \ldots, n$

ANOVA model (25.1) is similar in appearance to fixed ANOVA model (16.2). The main distinction is that the factor level means μ_i are constants for ANOVA model I but are random variables for ANOVA model II. Hence, ANOVA model II is often called a *random* ANOVA model.

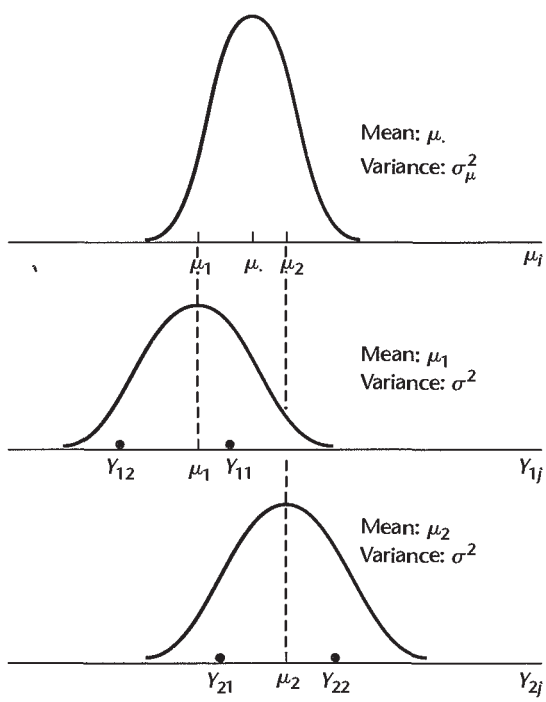
Meaning of Model Terms. We shall explain the meaning of the model terms with reference to the personnel officers in the Apex Enterprises example. The term μ_i corresponds to the mean of all ratings by the ith personnel officer if the officer interviewed all prospective

employees. The expected value of μ_i is $\mu_.$. Thus, $\mu_.$ represents here the mean rating for all prospective employees by all personnel officers. The variability of the personnel officers' mean ratings μ_i is measured by the variance σ_μ^2. The more the different personnel officers vary in their mean ratings (for instance, some may rate consistently higher than others), the greater will be σ_μ^2. If all personnel officers rate at the same mean level, all μ_i will be equal to $\mu_.$ and then $\sigma_\mu^2 = 0$.

The term ε_{ij} represents the variation associated with the different potential values as assessed by the ith personnel officer for the different prospective employees. Note that ANOVA model (25.1) assumes that all ε_{ij} have the same variance σ^2. This means that the distributions of ratings for prospective employees by the different personnel officers are assumed to have the same variability. The distributions for the different personnel officers may differ with respect to their means but not with respect to their variability according to ANOVA model (25.1).

Figure 25.1 illustrates ANOVA model II. On the top is shown the distribution of the personnel officers' mean ratings μ_i, which is normal. Several μ_i (two personnel officers' mean ratings in the illustration) are selected at random from this distribution. Each in turn leads to a distribution of the potential values of prospective employees as evaluated by the ith personnel officer, $Y_{ij} = \mu_i + \varepsilon_{ij}$, which are all normal distributions with the same variance. Several Y_{ij} responses are then selected from each of these distributions (two responses for each personnel officer in the illustration).

FIGURE 25.1 Representation of ANOVA Model II.



Important Features of Model

1. The expected value of a response Y_{ij} is:

$$E\{Y_{ij}\} = \mu. \tag{25.2a}$$

because we have by (25.1):

$$\begin{aligned} E\{Y_{ij}\} &= E\{\mu_i\} + E\{\varepsilon_{ij}\} \\ &= \mu. + 0 \\ &= \mu. \end{aligned}$$

Note that this expectation averages over the selections of both μ_i and ε_{ij}.

2. The variance of Y_{ij}, to be denoted by σ_Y^2, is:

$$\sigma^2\{Y_{ij}\} = \sigma_Y^2 = \sigma_\mu^2 + \sigma^2 \tag{25.2b}$$

Thus, all observations Y_{ij} have the same variance. The result in (25.2b) follows because ANOVA model II assumes that μ_i and ε_{ij} are independent random variables, and $\sigma^2\{\mu_i\} = \sigma_\mu^2$ and $\sigma^2\{\varepsilon_{ij}\} = \sigma^2$ according to ANOVA model (25.1). Because the variance of Y in this model is the sum of two components, σ_μ^2 and σ^2, this model is sometimes called a *components of variance model* and σ_Y^2 is referred to as the *total variance*. (Reference 25.1 provides detailed discussions of variance components models.)

3. The Y_{ij} are normally distributed because they are linear combinations of the independent normal variables μ_i and ε_{ij}.

4. Unlike for fixed ANOVA model I where all observations Y_{ij} are independent, the Y_{ij} for random ANOVA model II are only independent if they pertain to different factor levels. The covariance of any two observations with random ANOVA model (25.1) can be shown to be:

$$\sigma\{Y_{ij}, Y_{ij'}\} = \sigma_\mu^2 \qquad j \neq j' \tag{25.2c}$$

$$\sigma\{Y_{ij}, Y_{i'j'}\} = 0 \qquad i \neq i' \tag{25.2d}$$

Thus, random ANOVA model (25.1) assumes that the covariance between any two responses for the same factor level is constant for all factor levels.

We illustrate the nature of the variance-covariance matrix of the responses Y_{ij} for random ANOVA model (25.1) for a simple illustration where there are $r = 2$ factor levels and $n = 2$ cases for each level. The observations vector is:

$$\mathbf{Y} = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix}$$

and the variance-covariance matrix of the Y observations is:

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \begin{bmatrix} \sigma_Y^2 & \sigma_\mu^2 & 0 & 0 \\ \sigma_\mu^2 & \sigma_Y^2 & 0 & 0 \\ 0 & 0 & \sigma_Y^2 & \sigma_\mu^2 \\ 0 & 0 & \sigma_\mu^2 & \sigma_Y^2 \end{bmatrix}$$

Note that all observations have the same variance σ_Y^2, as indicated by (25.2b), any two observations from the same factor level have covariance σ_μ^2 as indicated by (25.2c), and any two observations from different factor levels are uncorrelated as indicated by (25.2d).

The reason why any two responses from the same factor level are correlated is that, in advance of the random trials, the responses are expected to be similar because they will both have the same random component μ_i and will differ only because of the error terms ε_{ij}.

Once the factor levels have been selected, however, random ANOVA model (25.1) assumes that any two responses from the same factor level are independent because the factor level mean μ_i is then fixed and the two observations differ only because of the error terms ε_{ij} which are assumed to be independent. Thus, in the Apex Enterprises example, once the personnel officers have been selected, random ANOVA model (25.1) assumes that the different ratings Y_{ij} by a given personnel officer are independent.

Comment

At times, the population of the μ_i may be relatively small and should be treated as a finite population. This can be done, but we do not discuss this case here. If the population of the μ_i is finite but large, little is lost in treating it as an infinite population. We did this, in fact, in our Apex Enterprises illustration. The number of personnel officers employed by Apex Enterprises is finite, but since there are many we treated the population of the μ_i as an infinite one. Thus, there are two basic situations when the population of the μ_i is treated as infinite—when the population is finite but large, and when interest centers in the underlying *process* generating the μ_i. ■

Questions of Interest

When ANOVA model II is appropriate, there is usually no interest in inferences about the particular μ_i included in the study, such as which is the largest or smallest, but rather in inferences about the entire population of the μ_i. Specifically, interest often centers on $\mu_.$, the mean of the μ_i, and on σ_μ^2, the variability of the μ_i. In the Apex Enterprises example, for instance, management would not ordinarily be as interested in the mean ratings of the five personnel officers who happened to be included in the study as in the mean rating by all personnel officers and in the variability of mean ratings among all personnel officers.

While σ_μ^2 is a direct measure of the variability of the μ_i, the effect of this variability is often measured more meaningfully relative to the total variability σ_Y^2 in (25.2b):

$$\frac{\sigma_\mu^2}{\sigma_Y^2} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2} \tag{25.3}$$

Note that this ratio measures the proportion of the total variability of the Y_{ij} that is accounted for by the variability of the μ_i. It takes on the value 0 when $\sigma_\mu^2 = 0$ and values near 1 when σ_μ^2 is large relative to σ^2.

With reference to the Apex Enterprises example, the ratio measures the proportion of the total variability of ratings for all candidates by all personnel officers that is accounted for by differences in the mean ratings among the personnel officers. If the ratio is near zero, differences in the mean ratings among personnel officers are relatively insignificant. On the other hand, if the ratio is large, say, .8 or more, then much of the total variability is accounted for by differences between personnel officers, and management may wish to study the advisability of giving the personnel officers more training to obtain improved consistency of ratings between officers.

It can be shown that the coefficient of correlation between any two responses from the same factor level with random ANOVA model (25.1) is:

$$\rho\{Y_{ij}, Y_{ij'}\} = \frac{\sigma_\mu^2}{\sigma_Y^2} = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2} \qquad j \neq j' \tag{25.4}$$

Thus, the measure in (25.3), which indicates the proportion of the total variability of the Y_{ij} that is accounted for by the variability of the μ_i, is actually the coefficient of correlation between any two observations from the same factor level. It is called the *intraclass correlation coefficient.*

Comment

The result in (25.4) follows from the definition of the coefficient of correlation in (A.25a):

$$\rho\{Y_{ij}, Y_{ij'}\} = \frac{\sigma\{Y_{ij}, Y_{ij'}\}}{\sigma\{Y_{ij}\}\sigma\{Y_{ij'}\}}$$

The covariance in the numerator is given in (25.2c), and $\sigma\{Y_{ij}\} = \sigma\{Y_{ij'}\} = \sigma_Y$ according to (25.2b). ■

Test whether $\sigma_\mu^2 = 0$

We first consider how to test whether all μ_i are equal:

$$\begin{aligned} H_0&: \sigma_\mu^2 = 0 \\ H_a&: \sigma_\mu^2 > 0 \end{aligned} \tag{25.5}$$

H_0 implies that all μ_i are equal; that is, $\mu_i \equiv \mu_.$. H_a implies that the μ_i differ. For the personnel officers example, H_0 implies that the mean ratings for all personnel officers are the same, while H_a implies that they differ.

Despite the fact that ANOVA model II differs from ANOVA model I, the analysis of variance for a single-factor study is conducted in identical fashion. (This is not always the case in more complex situations.) The difference between the two models appears in the expected mean squares. It can be shown, in a manner similar to that employed in our derivation for ANOVA model I, that the expected mean squares for ANOVA model II when all treatment sample sizes equal n are as follows:

$$E\{MSE\} = \sigma^2 \tag{25.6}$$

$$E\{MSTR\} = \sigma^2 + n\sigma_\mu^2 \tag{25.7}$$

It follows from (25.6) and (25.7) that if $\sigma_\mu^2 = 0$, MSE and $MSTR$ have the same expectation σ^2. Otherwise, $E\{MSTR\} > E\{MSE\}$ since $n > 0$ always. Hence, large values of the test

statistic:

$$F^* = \frac{MSTR}{MSE} \tag{25.8}$$

will lead to conclusion H_a in (25.5). Since F^* again follows the F distribution when H_0 holds, the decision rule for controlling the risk of making a Type I error at α is the same as the one for ANOVA model I:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; r-1, r(n-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; r-1, r(n-1)], \text{ conclude } H_a \end{aligned} \tag{25.9}$$

Note that the degrees of freedom associated with MSE here are $n_T - r = r(n-1)$ since $n_T = rn$ when all factor level sample sizes are equal.

Example

Table 25.1 contains the results of the study by Apex Enterprises on the evaluation ratings of potential employees by its personnel officers. Five personnel officers were selected at random, and four prospective employee candidates were assigned at random to each selected officer. Figure 25.2 contains dot plots of the ratings for each of the five personnel officers. It appears that the locations of the rating distributions for the personnel officers differ, that the variability within each of the five distributions is approximately the same, and that the

TABLE 25.1 Ratings by Five Personnel Officers—Apex Enterprises Example.

Officer i	Candidate (j) 1	2	3	4	Mean
A	76	65	85	74	$\bar{Y}_{1\cdot} = 75.00$
B	59	75	81	67	$\bar{Y}_{2\cdot} = 70.50$
C	49	63	61	46	$\bar{Y}_{3\cdot} = 54.75$
D	74	71	85	89	$\bar{Y}_{4\cdot} = 79.75$
E	66	84	80	79	$\bar{Y}_{5\cdot} = 77.25$
Mean					$\bar{Y}_{\cdot\cdot} = 71.45$

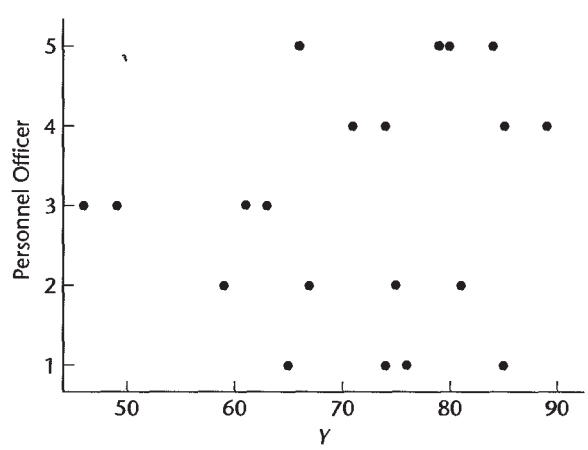


FIGURE 25.2 Dot Plots of Ratings by Five Personnel Officers—Apex Enterprises Example.

TABLE 25.2 ANOVA Table for Single-factor ANOVA Model II—Apex Enterprises Example.

Source of Variation	SS	df	MS	E{MS} General	E{MS} Example
Between personnel officers	$SSTR = 1{,}579.7$	4	$MSTR = 394.9$	$\sigma^2 + n\sigma_\mu^2$	$\sigma^2 + 4\sigma_\mu^2$
Error (within personnel officers)	$SSE = 1{,}099.3$	15	$MSE = 73.3$	σ^2	σ^2
Total	$SSTO = 2{,}678.9$	19			

variability within each of the rating distributions may be almost as large as the variability between the personnel officers.

The ANOVA calculations are routine and are shown in Table 25.2, which also shows the expected mean squares in general and for the Apex Enterprises example. Using the results from Table 25.2, the appropriate test statistic for determining whether $\sigma_\mu^2 = 0$ is:

$$F^* = \frac{394.9}{73.3} = 5.39$$

To control the risk of making a Type I error at $\alpha = .05$, we require $F(.95; 4, 15) = 3.06$. Hence, the decision rule is:

$$\text{If } F^* \leq 3.06, \text{ conclude } H_0$$
$$\text{If } F^* > 3.06, \text{ conclude } H_a$$

Since $F^* = 5.39 > 3.06$, we conclude H_a, that $\sigma_\mu^2 > 0$ or that the mean ratings of the personnel officers differ. The P-value of the test is .01.

Comments

1. We illustrate the derivation of an expected mean square for ANOVA model II by sketching the development for deriving $E\{MSTR\}$ in (25.7) when $n_i \equiv n$. The proof parallels that for ANOVA model I. According to ANOVA model (25.1), we can write:

$$\bar{Y}_{i\cdot} = \mu_i + \bar{\varepsilon}_{i\cdot}$$
$$\bar{Y}_{\cdot\cdot} = \bar{\mu}_{\cdot} + \bar{\varepsilon}_{\cdot\cdot}$$

where $\bar{\varepsilon}_{i\cdot}$ and $\bar{\varepsilon}_{\cdot\cdot}$ are defined in (16.44) and (16.47), respectively, and:

$$\bar{\mu}_{\cdot} = \frac{\sum_{i=1}^{r} \mu_i}{r}$$

(Note the use of a different notation for the mean of the μ_i here than for ANOVA model I to emphasize the random nature of the mean of the r values μ_i for ANOVA model II.) Corresponding to (16.49), we obtain:

$$\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot} = (\mu_i - \bar{\mu}_{\cdot}) + (\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})$$

so that:

$$\sum(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 = \sum(\mu_i - \bar{\mu}_{\cdot})^2 + \sum(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})^2 + 2\sum(\mu_i - \bar{\mu}_{\cdot})(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon}_{\cdot\cdot})$$

When we take the expectation, the cross-product term drops out because of the independence of the μ_i and the ε_{ij} and because the deviations $\mu_i - \bar{\mu}.$ and $\bar{\varepsilon}_{i}. - \bar{\varepsilon}..$ all have expectations zero. From (16.52) we know that:

$$E\left\{\sum(\bar{\varepsilon}_{i}. - \bar{\varepsilon}..)^2\right\} = \frac{(r-1)\sigma^2}{n}$$

Lastly, since $\sum(\mu_i - \bar{\mu}.)^2$ is the numerator of an ordinary sample variance for r independent μ_i values, it follows from the unbiasedness of the sample variance that:

$$E\left\{\sum(\mu_i - \bar{\mu}.)^2\right\} = (r-1)\sigma_\mu^2$$

Hence, we obtain:

$$E\left\{\frac{n}{r-1}\sum(\bar{Y}_{i}. - \bar{Y}..)^2\right\} = \frac{n}{r-1}\left[(r-1)\sigma_\mu^2 + \frac{r-1}{n}\sigma^2\right] = n\sigma_\mu^2 + \sigma^2$$

which is the result in (25.7).

2. The F^* test statistic in (25.8) and the decision rule in (25.9) are also appropriate when the factor level sample sizes are not equal. The degrees of freedom associated with MSE are then denoted, as usual, by $n_T - r$, where $n_T = \sum n_i$. The expected value of $MSTR$ becomes:

$$E\{MSTR\} = \sigma^2 + n'\sigma_\mu^2 \tag{25.10}$$

where:

$$n' = \frac{1}{r-1}\left[\left(\sum n_i\right) - \frac{\sum n_i^2}{\sum n_i}\right] \tag{25.10a}$$

■

Estimation of $\mu.$

When ANOVA model II is applicable, there is frequent interest in estimating the overall mean $\mu..$ We now develop an interval estimate for $\mu.$ when all factor level sample sizes are equal. We know from (25.2a) that:

$$E\{Y_{ij}\} = \mu.$$

Hence, an unbiased estimator of $\mu.$ is:

$$\hat{\mu}. = \bar{Y}.. \tag{25.11}$$

It can be shown that the variance of this estimator is:

$$\sigma^2\{\bar{Y}..\} = \frac{\sigma_\mu^2}{r} + \frac{\sigma^2}{rn} = \frac{n\sigma_\mu^2 + \sigma^2}{rn} \tag{25.12}$$

Formula (25.12) shows that the variance of $\bar{Y}..$ is made up of two components. The first corresponds to the variance of a sample mean based on r values when sampling from the population of the μ_i, and it reflects the contribution due to sampling the factor levels. The second component corresponds to the variance of a sample mean based on rn observations when sampling from the populations of the Y_{ij}, given the μ_i, and it reflects the contribution due to variation within factor levels.

An unbiased estimator of $\sigma^2\{\bar{Y}..\}$ is:

$$s^2\{\bar{Y}..\} = \frac{MSTR}{rn} \tag{25.13}$$

This estimator is unbiased because we know from (25.7) that $E\{MSTR\} = n\sigma_\mu^2 + \sigma^2$. Dividing the result in (25.7) by rn yields (25.12).

It can be shown that:

$$\frac{\bar{Y}_{..} - \mu_.}{s\{\bar{Y}_{..}\}} \text{ is distributed as } t(r-1) \text{ for ANOVA model (25.1)} \tag{25.14}$$

Hence, we obtain in usual fashion the confidence limits for $\mu_.$:

$$\bar{Y}_{..} \pm t(1-\alpha/2; r-1)s\{\bar{Y}_{..}\} \tag{25.15}$$

Example

Management of Apex Enterprises wishes to estimate the mean rating for all prospective employees by all personnel officers with a 90 percent confidence interval. We have from Tables 25.1 and 25.2:

$$\bar{Y}_{..} = 71.45 \qquad MSTR = 394.9 \qquad rn = 20$$

We require $t(.95; 4) = 2.132$ and:

$$s^2\{\bar{Y}_{..}\} = \frac{394.9}{20} = 19.75$$

Hence, $s\{\bar{Y}_{..}\} = 4.44$, the confidence limits are $71.45 \pm 2.132(4.44)$, and the desired 90 percent confidence interval is:

$$62 \leq \mu_. \leq 81$$

Thus, with a 90 percent confidence coefficient, we conclude that the mean rating assigned by all personnel officers to all prospective employees is between 62 and 81. The interval estimate is not very precise because of the relatively small samples of personnel officers and potential employees.

Comment

The variance of $\bar{Y}_{..}$ in (25.12) can be derived readily. First, we consider:

$$\bar{Y}_{i.} = \mu_i + \bar{\varepsilon}_{i.}$$

where $\bar{\varepsilon}_{i.}$ is defined in (16.44). Because of the independence of μ_i and the ε_{ij}, we have:

$$\sigma^2\{\bar{Y}_{i.}\} = \sigma_\mu^2 + \frac{\sigma^2}{n}$$

Remember that $\bar{\varepsilon}_{i.}$ is just an ordinary mean of n independent ε_{ij} values.

For the case $n_i \equiv n$ that we are considering here, we have:

$$\bar{Y}_{..} = \frac{\sum_{i=1}^{r} \bar{Y}_{i.}}{r}$$

In view of the independence of the μ_i and the ε_{ij} among themselves and between each other, it follows that the $\bar{Y}_{i.}$ are independent so that:

$$\sigma^2\{\bar{Y}_{..}\} = \frac{\sigma^2\{\bar{Y}_{i.}\}}{r} = \frac{\sigma_\mu^2}{r} + \frac{\sigma^2}{rn} = \frac{n\sigma_\mu^2 + \sigma^2}{rn}$$

■

Estimation of $\sigma_\mu^2/(\sigma_\mu^2 + \sigma^2)$

As noted earlier, the ratio $\sigma_\mu^2/(\sigma_\mu^2 + \sigma^2)$ reveals meaningfully the effect of the extent of variation between the μ_i. We shall develop an interval estimate for this ratio by first obtaining confidence limits for the ratio σ_μ^2/σ^2. It can be shown that *MSTR* and *MSE* are independent random variables for ANOVA model II, just as for ANOVA model I. When $n_i \equiv n$, the case considered here, it can be shown further that:

$$\frac{MSTR}{n\sigma_\mu^2 + \sigma^2} \div \frac{MSE}{\sigma^2} \sim F[r-1, r(n-1)] \tag{25.16}$$

Hence, we can write the probability statement:

$$P\{F[\alpha/2; r-1, r(n-1)] \leq \frac{MSTR}{n\sigma_\mu^2 + \sigma^2} \div \frac{MSE}{\sigma^2} \leq F[1-\alpha/2; r-1, r(n-1)]\} = 1-\alpha \tag{25.17}$$

Rearranging the inequalities, we obtain the following confidence limits L and U for σ_μ^2/σ^2:

$$L = \frac{1}{n}\left[\frac{MSTR}{MSE}\left(\frac{1}{F[1-\alpha/2; r-1, r(n-1)]}\right) - 1\right] \tag{25.18a}$$

$$U = \frac{1}{n}\left[\frac{MSTR}{MSE}\left(\frac{1}{F[\alpha/2; r-1, r(n-1)]}\right) - 1\right] \tag{25.18b}$$

where L is the lower confidence limit and U the upper.

The confidence limits L^* and U^* for $\sigma_\mu^2/(\sigma_\mu^2 + \sigma^2)$ can now be obtained and are as follows:

$$L^* = \frac{L}{1+L} \qquad U^* = \frac{U}{1+U} \tag{25.19}$$

Example

Management of Apex Enterprises wishes to obtain a 90 percent confidence interval for $\sigma_\mu^2/(\sigma_\mu^2 + \sigma^2)$. From previous work, we have:

$$MSTR = 394.9 \qquad MSE = 73.3 \qquad n = 4 \qquad r = 5$$

For a 90 percent confidence interval, we require:

$$F(.05; 4, 15) = .170 \qquad F(.95; 4, 15) = 3.06$$

Hence, the 90 percent confidence limits for σ_μ^2/σ^2 are by (25.18):

$$L = \frac{1}{4}\left[\frac{394.9}{73.3}\left(\frac{1}{3.06}\right) - 1\right] = .19 \qquad U = \frac{1}{4}\left[\frac{394.9}{73.3}\left(\frac{1}{.170}\right) - 1\right] = 7.7$$

and the confidence interval for σ_μ^2/σ^2 is:

$$.19 \leq \frac{\sigma_\mu^2}{\sigma^2} \leq 7.7$$

Finally, the confidence limits for $\sigma_\mu^2/(\sigma_\mu^2 + \sigma^2)$ are obtained by (25.19); they are $L^* = .19/1.19 = .16$ and $U^* = 7.7/8.7 = .89$. Hence, the 90 percent confidence interval is:

$$.16 \leq \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2} \leq .89$$

With confidence coefficient .90, we conclude that the variability of the mean ratings for the different personnel officers accounts for somewhere between 16 and 89 percent of the total variability of the ratings. Note that this interval estimate is not precise, partly the result of relatively small sample sizes and partly because variance components are much more difficult to estimate precisely than means. The confidence interval does indicate, though, that the variability among personnel officers is not negligible since it accounts for at least 16 percent of the total variability.

Comments

1. It may happen occasionally that the lower limit of the confidence interval for σ_μ^2/σ^2 is negative. Since this ratio cannot be negative, the usual practice is to consider the lower limit L in (25.18a) to be zero in that case.

2. If one-sided or two-sided tests concerning the relative magnitudes of σ_μ^2 and σ^2 are desired, such as the following (where c is a specified constant):

$$H_0\colon \sigma_\mu^2 \leq c\sigma^2 \qquad H_0\colon \sigma_\mu^2 = c\sigma^2$$
$$H_a\colon \sigma_\mu^2 > c\sigma^2 \qquad H_a\colon \sigma_\mu^2 \neq c\sigma^2$$

a decision rule can be constructed by utilizing (25.16). Alternatively, one-sided or two-sided confidence intervals can be used to draw the appropriate conclusion.

3. The ratio σ_μ^2/σ^2 is of relevance in planning investigations. In the Apex Enterprises example dealing with the personnel officers, suppose that the mean rating $\mu.$ is to be estimated, and that the costs of including in the study a personnel officer and a candidate are c_1 and c_2, respectively. For a given total budget C, the ratio σ_μ^2/σ^2 is the determining variable for finding the optimum balance between the number of personnel officers and the number of candidates to include in the study so as to minimize the variance of the estimator. If the populations are not large, the model will need to take account of their finite nature. ■

Estimation of σ^2

At times, it is desired to estimate σ^2 and σ_μ^2 separately. According to (25.6), an unbiased estimator of σ^2 is MSE. An interval estimate for σ^2 is easily constructed. We make use of the fact that $[r(n-1)MSE]/\sigma^2$ is distributed as a χ^2 random variable with $r(n-1)$ degrees of freedom:

$$\frac{r(n-1)MSE}{\sigma^2} \sim \chi^2[r(n-1)] \tag{25.20}$$

It follows that a $1-\alpha$ confidence interval for σ^2 is:

$$\frac{r(n-1)MSE}{\chi^2[1-\alpha/2; r(n-1)]} \leq \sigma^2 \leq \frac{r(n-1)MSE}{\chi^2[\alpha/2; r(n-1)]} \tag{25.21}$$

Example

To construct a 90 percent confidence interval for σ^2 for the Apex Enterprises example, we require:

$$MSE = 73.3 \qquad \chi^2(.05; 15) = 7.26 \qquad \chi^2(95; 15) = 25.0$$

The desired confidence interval by (25.21) then is:

$$44.0 = \frac{15(73.3)}{25.0} \leq \sigma^2 \leq \frac{15(73.3)}{7.26} = 151.4$$

An approximate 90 percent confidence interval for σ is obtained by taking the square roots of the confidence limits for σ^2:

$$6.6 \leq \sigma \leq 12.3$$

With 90 percent confidence, we conclude that the standard deviation of the ratings of prospective employees for each personnel officer is between 6.6 and 12.3 points.

Comment

Confidence interval (25.21) is also appropriate when the factor level sample sizes are not equal. The degrees of freedom associated with *MSE* are then denoted by $n_T - r$. ■

Point Estimation of σ_μ^2

An unbiased estimator of σ_μ^2 is available by noting that we have from (25.6) and (25.7):

$$E\{MSE\} = \sigma^2$$
$$E\{MSTR\} = \sigma^2 + n\sigma_\mu^2$$

It follows that:

$$\sigma_\mu^2 = \frac{E\{MSTR\} - E\{MSE\}}{n} \tag{25.22}$$

An unbiased estimator of σ_μ^2 is obtained by substituting the observed mean squares for the corresponding expected mean squares:

$$s_\mu^2 = \frac{MSTR - MSE}{n} \tag{25.23}$$

Occasionally, this point estimator will turn out to be negative. Since a variance cannot be negative, the usual practice is to consider the estimator to be zero in that event.

Comment

An unbiased estimator of σ_μ^2 when the factor level sample sizes are not equal can be obtained by slightly modifying the expression in (25.23). The denominator n is simply replaced by n' as defined in (25.10a). ■

Interval Estimation of σ_μ^2

It is not possible to construct exact confidence intervals for σ_μ^2. However, several approximate confidence intervals have been developed. We shall now describe two approximate confidence intervals for σ_μ^2, assuming as before that the study is balanced; that is, $n_i \equiv n$.

Procedures for constructing confidence intervals for σ_μ^2 when the factor level sample sizes are not equal are presented in Section 25.6 and in Reference 25.2.

Satterthwaite Procedure. The Satterthwaite procedure (Ref. 25.3) is a general procedure for constructing approximate confidence intervals for linear combinations of expected mean squares. Note that σ_μ^2 is such a linear combination since we can express (25.22) as follows:

$$\sigma_\mu^2 = \left(\frac{1}{n}\right) E\{MSTR\} + \left(-\frac{1}{n}\right) E\{MSE\} \tag{25.24}$$

In general, we shall state a linear combination of expected mean squares as follows:

$$L = c_1 E\{MS_1\} + \cdots + c_h E\{MS_h\} \tag{25.25}$$

where the c_i are coefficients.

An unbiased estimator of L is:

$$\hat{L} = c_1 MS_1 + \cdots + c_h MS_h \tag{25.26}$$

Let df_i denote the degrees of freedom associated with mean square MS_i. Satterthwaite has suggested that the distribution of the statistic:

$$\frac{(df)\hat{L}}{L} \tag{25.27}$$

can be approximated by a χ^2 distribution whose degrees of freedom, denoted by df, are given by:

$$df = \frac{(c_1 MS_1 + \cdots + c_h MS_h)^2}{\dfrac{(c_1 MS_1)^2}{df_1} + \cdots + \dfrac{(c_h MS_h)^2}{df_h}} \tag{25.28}$$

An approximate $1 - \alpha$ confidence interval for L therefore is:

$$\frac{(df)\hat{L}}{\chi^2(1-\alpha/2; df)} \le L \le \frac{(df)\hat{L}}{\chi^2(\alpha/2; df)} \tag{25.29}$$

where df is given by (25.28).

For the single-factor random ANOVA model (25.1) for a balanced study ($n_i \equiv n$), we have the following correspondences:

$$\begin{aligned}
MS_1 &= MSTR & MS_2 &= MSE \\
df_1 &= r - 1 & df_2 &= n_T - r = r(n-1) \\
c_1 &= \frac{1}{n} & c_2 &= -\frac{1}{n} \\
L = \sigma_\mu^2 &= \left(\frac{1}{n}\right) E\{MSTR\} + \left(-\frac{1}{n}\right) E\{MSE\} \\
\hat{L} = s_\mu^2 &= \left(\frac{1}{n}\right) MSTR + \left(-\frac{1}{n}\right) MSE
\end{aligned} \tag{25.30}$$

Hence, an approximate $1 - \alpha$ confidence interval for σ_μ^2 by the Satterthwaite procedure (25.29) is:

$$\frac{(df)s_\mu^2}{\chi^2(1-\alpha/2; df)} \leq \sigma_\mu^2 \leq \frac{(df)s_\mu^2}{\chi^2(\alpha/2; df)} \tag{25.31}$$

where:

$$df = \frac{\left(ns_\mu^2\right)^2}{\dfrac{(MSTR)^2}{r-1} + \dfrac{(MSE)^2}{r(n-1)}} \tag{25.31a}$$

Usually, the degrees of freedom will not turn out to be an integer. Interpolation in the χ^2 table or rounding to the nearest integer may then be used.

While the Satterthwaite procedure is general and easy to carry out, the accuracy of the approximation can be quite limited when some of the coefficients c_i are negative and some are positive. Note that this is the case here in (25.30), since $c_1 = 1/n$ and $c_2 = -1/n$. More detailed guidelines as to when the Satterthwaite approximation is appropriate are given in Reference 25.4.

Example

For the Apex Enterprises example, we shall first obtain a point estimate of σ_μ^2 by means of (25.23). We require:

$$MSE = 73.3 \qquad MSTR = 394.9 \qquad n = 4$$

Hence we find:

$$s_\mu^2 = \frac{394.9 - 73.3}{4} = 80.4$$

and the estimated standard deviation of the mean ratings of all personnel officers is $\sqrt{80.4} = 9.0$ points.

Next, we obtain a 90 percent confidence interval for σ_μ^2 by the Satterthwaite procedure. Using the earlier results:

$$s_\mu^2 = 80.4 \qquad MSTR = 394.9 \qquad MSE = 73.3 \qquad n = 4 \qquad r = 5$$

we obtain the degrees of freedom df by means of (25.31a):

$$df = \frac{[4(80.4)]^2}{\dfrac{(394.9)^2}{5-1} + \dfrac{(73.3)^2}{5(4-1)}} = 2.63$$

which we shall round up to 3.0. Confidence limits (25.31) also require:

$$\chi^2(.05; 3) = .352 \qquad \chi^2(.95; 3) = 7.81$$

so that the Satterthwaite approximate 90 percent confidence interval for σ_μ^2 is:

$$30.9 = \frac{3(80.4)}{7.81} \leq \sigma_\mu^2 \leq \frac{3(80.4)}{.352} = 685.2$$

By taking square roots of the two limits, we obtain an approximate confidence interval for σ_μ:

$$5.6 \leq \sigma_\mu \leq 26.2$$

Hence, with approximate 90 percent confidence coefficient, we conclude that the standard deviation of the mean ratings of all of the personnel officers is between 5.6 and 26.2 points.

MLS Procedure. An improved procedure for obtaining an approximate confidence interval for σ_μ^2 is based on the modified large sample (MLS) procedure (Ref. 25.5). It involves somewhat greater computational complexity than the Satterthwaite procedure, and is designed to estimate a linear combination of two expected mean squares for balanced studies of the form:

$$L = c_1 E\{MS_1\} + c_2 E\{MS_2\} \qquad c_1 > 0, \qquad c_2 < 0 \tag{25.32}$$

where c_1 is positive and c_2 is negative. An unbiased estimator of L is:

$$\hat{L} = c_1 MS_1 + c_2 MS_2 \qquad c_1 > 0, \qquad c_2 < 0 \tag{25.33}$$

If $(df_1)MS_1/E\{MS_1\}$ and $(df_2)MS_2/E\{MS_2\}$ are independent χ^2 random variables with df_1 and df_2 degrees of freedom, respectively, an approximate $1-\alpha$ confidence interval for L is given by:

$$\hat{L} - H_L \leq L \leq \hat{L} + H_U \tag{25.34}$$

where $\hat{L}$ is defined in (25.33) and H_L and H_U are defined by the equations in Table 25.3.

TABLE 25.3 Computational Formulas for MLS Approximate $1-\alpha$ Confidence Limits in (25.34).

$F_1 = F(1-\alpha/2; df_1, \infty)$	(25.34a)
$F_2 = F(1-\alpha/2; df_2, \infty)$	(25.34b)
$F_3 = F(1-\alpha/2; \infty, df_1)$	(25.34c)
$F_4 = F(1-\alpha/2; \infty, df_2)$	(25.34d)
$F_5 = F(1-\alpha/2; df_1, df_2)$	(25.34e)
$F_6 = F(1-\alpha/2, df_2, df_1)$	(25.34f)
$G_1 = 1 - \dfrac{1}{F_1}$	(25.34g)
$G_2 = 1 - \dfrac{1}{F_2}$	(25.34h)
$G_3 = \dfrac{(F_5-1)^2 - (G_1F_5)^2 - (F_4-1)^2}{F_5}$	(25.34i)
$G_4 = F_6\left[\left(\dfrac{F_6-1}{F_6}\right)^2 - \left(\dfrac{F_3-1}{F_6}\right)^2 - G_2^2\right]$	(25.34j)
$H_L = \{[G_1c_1MS_1]^2 + [(F_4-1)c_2MS_2]^2 - G_3c_1c_2MS_1MS_2\}^{1/2}$	(25.34k)
$H_U = \{[(F_3-1)c_1MS_1]^2 + (G_2c_2MS_2)^2 - G_4c_1c_2MS_1MS_2\}^{1/2}$	(25.34l)

To obtain an approximate $1-\alpha$ confidence interval for σ_μ^2 with the MLS procedure, we simply observe that the correspondences in (25.30) for the Satterthwaite procedure apply here also and confidence interval (25.34) becomes:

$$s_\mu^2 - H_L \leq \sigma_\mu^2 \leq s_\mu^2 + H_U \tag{25.35}$$

Example

For the Apex Enterprises example, we shall obtain a 90 percent confidence interval for σ_μ^2 by means of the MLS procedure. From earlier, we have:

$$c_1 = 1/n = 1/4 = .25 \qquad MS_1 = MSTR = 394.9 \qquad df_1 = r - 1 = 4$$

$$c_2 = -1/n = -1/4 = -.25 \qquad MS_2 = MSE = 73.3 \qquad df_2 = r(n-1) = 15$$

$$\hat{L} = s_\mu^2 = 80.4$$

We first determine the six percentiles (25.34a) to (25.34f):

$$F_1 = F(.95; 4, \infty) = 2.37 \qquad F_2 = F(.95; 15, \infty) = 1.67$$

$$F_3 = F(.95; \infty, 4) = 5.63 \qquad F_4 = F(.95; \infty, 15) = 2.07$$

$$F_5 = F(.95; 4, 15) = 3.06 \qquad F_6 = F(.95; 15, 4) = 5.86$$

Intermediate calculations required are:

$$G_1 = 1 - \frac{1}{2.37} = .5781$$

$$G_2 = 1 - \frac{1}{1.67} = .4012$$

$$G_3 = \frac{(3.06-1)^2 - [(.5781)3.06]^2 - (2.07-1)^2}{3.06} = -.0100$$

$$G_4 = 5.86\left[\left(\frac{5.86-1}{5.86}\right)^2 - \left(\frac{5.63-1}{5.86}\right)^2 - (.4012)^2\right] = -.5708$$

H_L and H_U are then computed as follows:

$$\begin{aligned} H_L &= \{[(.5781)(.25)394.9]^2 + [(2.07-1)(-.25)73.3]^2 \\ &\quad - (-.0100)(.25)(-.25)(394.9)73.3\}^{1/2} \\ &= 60.2 \end{aligned}$$

$$\begin{aligned} H_U &= \{[(5.63-1)(.25)394.9]^2 + [(.4012)(-.25)73.3]^2 \\ &\quad - (-.5708)(.25)(-.25)(394.9)73.3\}^{1/2} \\ &= 456.0 \end{aligned}$$

The approximate 90 percent confidence interval for σ_μ^2 therefore is:

$$20.2 = 80.4 - 60.2 \leq \sigma_\mu^2 \leq 80.4 + 456.0 = 536.4$$

Taking the square roots of the confidence limits, we obtain an approximate confidence interval for σ_μ:

$$4.5 \leq \sigma_\mu \leq 23.2$$

Notice that in this instance the confidence limits obtained by the Satterthwaite procedure (5.6 and 26.2) are quite similar to the ones just obtained by the more accurate MLS procedure. Note also the impreciseness of the MLS confidence interval here, a result of the small sample sizes and the difficulty in estimating variance components precisely.

Random Factor Effects Model

We can express the single-factor random cell means model (25.1) in an equivalent random factor effects fashion, just as we did for fixed factor levels in Chapter 16. We do this by expressing each factor level mean μ_i as a deviation from its expected value, $E\{\mu_i\} = \mu.$, as follows:

$$\tau_i = \mu_i - \mu. \tag{25.36}$$

Then we simply replace μ_i in ANOVA model (25.1) by its equivalent expression from (25.36):

$$\mu_i = \mu. + \tau_i \tag{25.37}$$

The random factor effects model therefore is expressed as follows:

$$Y_{ij} = \mu. + \tau_i + \varepsilon_{ij} \tag{25.38}$$

where:

$\mu.$ is a constant component common to all observations
τ_i are independent $N(0, \sigma_\mu^2)$
ε_{ij} are independent $N(0, \sigma^2)$
τ_i and ε_{ij} are independent
$i = 1, \ldots, r;\ j = 1, \ldots, n$

Note that the τ_i are random variables in ANOVA model (25.38). With reference to the personnel officers in the Apex Enterprises example, τ_i represents the effect of the ith personnel officer who is selected at random. Specifically, τ_i measures by how much the mean rating of all potential employees by the ith personnel officer differs from the overall mean rating by all personnel officers.

25.2 Two-Factor Studies—ANOVA Models II and III

ANOVA Model II—Random Factor Effects

Consider an investigation of the effects of machine operators (factor A) and machines (factor B) on the number of pieces produced in a day. Five operators and three machines are used in the study. Yet the inferences are not to be confined to the particular five operators and three machines participating in the study, but rather they are to pertain to all operators

and all machines available to the company. Here a random factor effects ANOVA model (model II) would be appropriate for the two-factor study, since each of the two sets of factor levels may be considered the result of sampling a population (all operators, all machines) about which inferences are to be drawn.

In the random factor effects version of ANOVA model II for a two-factor study, we assume analogously to a single-factor study that both the factor A main effects α_i and the factor B main effects β_j are independent random variables. Further, we assume that the interaction effects $(\alpha\beta)_{ij}$ are independent random variables. Thus, the random factor level effects version of ANOVA model II for a two-factor study with equal sample sizes n is:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \tag{25.39}$$

where:

$\mu_{..}$ is a constant

α_i, β_j, $(\alpha\beta)_{ij}$ are independent normal random variables with expectations zero and respective variances σ^2_α, σ^2_β, $\sigma^2_{\alpha\beta}$

ε_{ijk} are independent $N(0, \sigma^2)$

α_i, β_j, $(\alpha\beta)_{ij}$, and ε_{ijk} are pairwise independent

$i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n$

Meaning of Model Terms. We shall explain the meaning of the terms in random ANOVA model (25.39) with reference to the production example involving the two factors, machine operators and machines. The main effect of operator i in the study (selected at random from the population of operators) is α_i. Similarly, the main effect of machine j in the study (selected at random from the population of machines) is β_j. Further, the interaction effect between operator i and machine j on the number of pieces produced per day is $(\alpha\beta)_{ij}$. ANOVA model (25.39) assumes that the main effects of operators on output per day are normally distributed with zero mean and variance σ^2_α. Similarly, the main effects of machines are normally distributed with zero mean and variance σ^2_β. Finally, the operator-machine interaction effects are normally distributed with mean zero and variance $\sigma^2_{\alpha\beta}$. Since random factor effects ANOVA model (25.39) assumes these three effects to be independent random variables, the mean output for operator i–machine j, namely, $\mu_{ij} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij}$, may be viewed as the sum of independent selections of α_i, β_j, and $(\alpha\beta)_{ij}$ from three different normal distributions.

Comment

We caution that random factor effects ANOVA model (25.39) should only be used if the factor levels of the two factors do indeed represent random samples from populations of interest. Also, when a study involves only a few levels of each random factor, precise estimation of the factor variance components will usually be very difficult because of the small number of factor levels sampled. ■

Important Features of Model

1. For ANOVA model (25.39), the expected value of response Y_{ijk} is:

$$E\{Y_{ijk}\} = \mu_{..} \tag{25.40a}$$

2. The variance of Y_{ijk}, denoted by σ_Y^2, is:

$$\sigma^2\{Y_{ijk}\} = \sigma_Y^2 = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2 \tag{25.40b}$$

The Y_{ijk} thus have constant variance. They are normally distributed because they are linear combinations of independent normal random variables.

3. In advance of the random trials, different responses Y_{ijk} are independent except for responses from the same factor A level and/or from the same factor B level, which are correlated because they contain some common random terms. The covariances are as follows:

$$\sigma\{Y_{ijk}, Y_{ij'k'}\} = \sigma_\alpha^2 \qquad j \neq j' \tag{25.41a}$$

$$\sigma\{Y_{ijk}, Y_{i'jk'}\} = \sigma_\beta^2 \qquad i \neq i' \tag{25.41b}$$

$$\sigma\{Y_{ijk}, Y_{ijk'}\} = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma_{\alpha\beta}^2 \qquad k \neq k' \tag{25.41c}$$

$$\sigma\{Y_{ijk}, Y_{i'j'k'}\} = 0 \qquad i \neq i', j \neq j' \tag{25.41d}$$

ANOVA Model III—Mixed Factor Effects

When one of the two factors has fixed factor levels while the other has random factor levels, a mixed factor effects ANOVA model (model III) is applicable. An instance where this model may be appropriate is an investigation of the effects of four different training methods (factor A) and five instructors (factor B) upon learning in a company training program. The four levels for training methods may be considered fixed, since interest centers in these particular training methods. In contrast, the levels for instructors may be viewed as random, since inferences are to be made about a population of instructors of which the five used in the study are viewed as a sample.

Two mixed factor effects ANOVA models are widely used. They are related to each other and are called the *restricted* and *unrestricted* mixed models. The restricted model is somewhat more general, and will be the mixed model that we shall present. When factor A has fixed factor levels and factor B has random factor levels, the α_i effects are constants and the β_j effects are random variables. The interaction effects $(\alpha\beta)_{ij}$ are also random variables because the factor B levels are random. As for the fixed effects ANOVA model for two-factor studies, the fixed effects α_i in the restricted mixed model will be subject to the restriction that their sum is zero; i.e., $\sum \alpha_i = 0$. Similarly, the interaction terms $(\alpha\beta)_{ij}$ will be subject to a restriction related to the fact that all fixed factor A levels are included in the study; the restriction is that $\sum_i (\alpha\beta)_{ij} = 0$ for each level j of random factor B. Any two interaction terms will be independent, as in the random effects model (25.39), except if they come from the same level of random factor B in which case they will be correlated. The correlation is related to the restriction that $\sum_i (\alpha\beta)_{ij} = 0$ for each level j of random factor B.

The restricted mixed ANOVA model for two-factor studies, where factor A is fixed and factor B is random, can now be stated as follows:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \tag{25.42}$$

where:

$\mu_{..}$ is a constant

α_i are constants subject to the restriction $\sum \alpha_i = 0$

β_j are independent $N(0, \sigma^2_\beta)$

$(\alpha\beta)_{ij}$ are $N\left(0, \frac{a-1}{a}\sigma^2_{\alpha\beta}\right)$, subject to the restrictions:

$$\sum_i (\alpha\beta)_{ij} = 0 \qquad \text{for all } j$$

$$\sigma\{(\alpha\beta)_{ij}, (\alpha\beta)_{i'j}\} = -\frac{1}{a}\sigma^2_{\alpha\beta} \qquad i \neq i'$$

ε_{ijk} are independent $N(0, \sigma^2)$

β_j, $(\alpha\beta)_{ij}$, and ε_{ijk} are pairwise independent

$i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n$

Comments

1. Note that $\sigma^2_{\alpha\beta}$ is not the variance of the interaction terms in model (25.42) but is proportional to their variance, the proportionality constant being $(a-1)/a$. The reason why the variance of the interaction terms in ANOVA model (25.42) is expressed as $(a-1)\sigma^2_{\alpha\beta}/a$ rather than simply as $\sigma^2_{\alpha\beta}$ is so that the expected mean squares will be relatively simple expressions. This facilitates the making of inferences for this model. Some texts denote the variance of $(\alpha\beta)_{ij}$ by $\sigma^2_{\alpha\beta}$.

2. The unrestricted mixed ANOVA model for two-factor studies is quite similar to the restricted model in (25.42). In the unrestricted model, there are no restrictions on the interaction effects $(\alpha\beta)_{ij}$ and they are pairwise independent. Denote the unrestricted random effects by β^*_j and $(\alpha\beta)^*_{ij}$. Also let $(\overline{\alpha\beta})^*_{\cdot j}$ denote the mean of the unrestricted interaction terms $(\alpha\beta)^*_{1j}, (\alpha\beta)^*_{2j}, \ldots, (\alpha\beta)^*_{aj}$ for the fixed factor A levels for any factor level j of random factor B. Then the terms β_j and $(\alpha\beta)_{ij}$ in restricted model (25.42) are related to the unrestricted terms as follows:

$$\beta_j = \beta^*_j + (\overline{\alpha\beta})^*_{\cdot j} \qquad (\alpha\beta)_{ij} = (\alpha\beta)^*_{ij} - (\overline{\alpha\beta})^*_{\cdot j} \tag{25.43}$$

The restrictions on the $(\alpha\beta)_{ij}$ in model (25.42) follow from the relation in (25.43). References 25.6 and 25.7 contain detailed discussions of the restricted and unrestricted mixed ANOVA models. ■

Important Features of Model. The expected value of response Y_{ijk} for mixed ANOVA model (25.42) is:

$$E\{Y_{ijk}\} = \mu_{..} + \alpha_i \tag{25.44}$$

The variance of Y_{ijk} follows directly from the pairwise independence of α_i, β_j, and $(\alpha\beta)_{ij}$:

$$\sigma^2\{Y_{ijk}\} = \sigma^2_Y = \sigma^2_\beta + \frac{a-1}{a}\sigma^2_{\alpha\beta} + \sigma^2 \tag{25.45}$$

Notice that the Y_{ijk} have constant variance. Further, they are normally distributed because each is a linear combination of independent normal random variables.

In advance of the random trials, different responses Y_{ijk} are independent if they are not from the same random factor B level. Responses from the same random factor B level are

correlated; their covariances are as follows:

$$\sigma\{Y_{ijk}, Y_{ijk'}\} = \sigma_\beta^2 + \frac{a-1}{a}\sigma_{\alpha\beta}^2 \qquad k \neq k' \tag{25.46a}$$

$$\sigma\{Y_{ijk}, Y_{i'jk'}\} = \sigma_\beta^2 - \frac{1}{a}\sigma_{\alpha\beta}^2 \qquad i \neq i' \tag{25.46b}$$

$$\sigma\{Y_{ijk}, Y_{i'j'k'}\} = 0 \qquad j \neq j' \tag{25.46c}$$

Covariance Structure of Observations. We shall illustrate the form of the variance-covariance matrix of the responses Y_{ijk} for mixed ANOVA model (25.42) for a simple example. Here, A is a fixed factor with $a = 2$ levels, B is a random factor with $b = 2$ levels, and $n = 2$ responses are obtained for each of the six treatments. The variance of response Y_{ijk} is according to (25.45) for $a = 2$:

$$\sigma^2\{Y_{ijk}\} = \sigma_Y^2 = \sigma_\beta^2 + \sigma_{\alpha\beta}^2/2 + \sigma^2$$

The covariance in (25.46a) will be denoted by $\sigma_{kk'}$ to indicate that the two Y_{ijk} observations only differ for the replication. Similarly, the covariance in (25.46b) will be denoted by $\sigma_{ii'}$ to indicate that the two observations come from different factor A levels but not from different factor B levels. In this notation, the two pairwise covariances are for $a = 2$:

$$\sigma_{kk'} = \sigma_\beta^2 + \sigma_{\alpha\beta}^2/2$$

$$\sigma_{ii'} = \sigma_\beta^2 - \sigma_{\alpha\beta}^2/2$$

The response vector $\mathbf{Y}$ for this example is shown in Table 25.4a. Note that the observations are listed in the vector with i varying within j. This permits a simple block structure

TABLE 25.4 Illustration of Variance-Covariance Matrix for Mixed Model (25.42)—$a = 2$, $b = 2$, $n = 2$.

(a) Observations Vector

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{211} \\ Y_{212} \\ Y_{121} \\ Y_{122} \\ Y_{221} \\ Y_{222} \end{bmatrix}$$

(b) Variance-Covariance Matrix in Block Form

$$\underset{8\times 8}{\boldsymbol{\sigma}^2\{\mathbf{Y}\}} = \begin{bmatrix} \underset{4\times 4}{\boldsymbol{\Sigma}} & \underset{4\times 4}{\mathbf{0}} \\ \underset{4\times 4}{\mathbf{0}} & \underset{4\times 4}{\boldsymbol{\Sigma}} \end{bmatrix}$$

where:

$\mathbf{0}$ = 4 × 4 matrix containing all 0s

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_Y^2 & \sigma_{kk'} & \sigma_{ii'} & \sigma_{ii'} \\ \sigma_{kk'} & \sigma_Y^2 & \sigma_{ii'} & \sigma_{ii'} \\ \sigma_{ii'} & \sigma_{ii'} & \sigma_Y^2 & \sigma_{kk'} \\ \sigma_{ii'} & \sigma_{ii'} & \sigma_{kk'} & \sigma_Y^2 \end{bmatrix}$$

$$\sigma_Y^2 = \sigma_\beta^2 + \sigma_{\alpha\beta}^2/2 + \sigma^2$$

$$\sigma_{kk'} = \sigma_\beta^2 + \sigma_{\alpha\beta}^2/2$$

$$\sigma_{ii'} = \sigma_\beta^2 - \sigma_{\alpha\beta}^2/2$$

presentation of the variance-covariance matrix in Table 25.4b. In this presentation, four rows and four columns are represented by a block matrix. Because of the symmetry of the blocks, only two different block matrices are required. These are shown in Table 25.4b. Note the correlations between pairs of observations on the block main diagonal and the uncorrelatedness elsewhere.

Comment

The reason why the restricted mixed model in (25.42) is somewhat more general than the unrestricted model is that two observations from the same random factor B level can be positively or negatively correlated for the restricted model according to (25.46b) but cannot be negatively correlated for the unrestricted model. ■

25.3 Two-Factor Studies—ANOVA Tests for Models II and III

For both the mixed and random ANOVA models for two-factor studies, the analysis of variance calculations for sums of squares are identical to those for the fixed ANOVA model. Thus, formulas (19.37) and (19.39) are entirely applicable for two-factor ANOVA models II and III. Similarly, the degrees of freedom and mean squares are exactly the same as those shown in Table 19.8 for the fixed two-factor ANOVA model. The random and mixed ANOVA models depart from the fixed ANOVA model only in the expected mean squares and the consequent choice of the appropriate test statistic.

Expected Mean Squares

The expected mean squares for the random and mixed ANOVA models for balanced two-factor studies can be worked out by utilizing the properties of the model and applying the usual expectation theorems. They are shown in Table 25.5, together with those for the fixed ANOVA model. The derivations are tedious, but simple rules have been developed for finding the expected mean squares. These rules are described in Appendix D.

TABLE 25.5 Expected Mean Squares for Balanced Two-Factor ANOVA Models.

Mean Square	*df*	Fixed ANOVA Model (A and B fixed)	Random ANOVA Model (A and B random)	Mixed ANOVA Model (A fixed, B random)
MSA	$a-1$	$\sigma^2 + nb\frac{\sum \alpha_i^2}{a-1}$	$\sigma^2 + nb\sigma_\alpha^2 + n\sigma_{\alpha\beta}^2$	$\sigma^2 + nb\frac{\sum \alpha_i^2}{a-1} + n\sigma_{\alpha\beta}^2$
MSB	$b-1$	$\sigma^2 + na\frac{\sum \beta_j^2}{b-1}$	$\sigma^2 + na\sigma_\beta^2 + n\sigma_{\alpha\beta}^2$	$\sigma^2 + na\sigma_\beta^2$
MSAB	$(a-1)(b-1)$	$\sigma^2 + n\frac{\sum\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$	$\sigma^2 + n\sigma_{\alpha\beta}^2$	$\sigma^2 + n\sigma_{\alpha\beta}^2$
MSE	$(n-1)ab$	σ^2	σ^2	σ^2

TABLE 25.6 Test Statistics for Balanced Two-Factor ANOVA Models.

Test for Presence of Effects of:	Fixed ANOVA Model (*A* and *B* fixed)	Random ANOVA Model (*A* and *B* random)	Mixed ANOVA Model (*A* fixed, *B* random)
Factor *A*	*MSA/MSE*	*MSA/MSAB*	*MSA/MSAB*
Factor *B*	*MSB/MSE*	*MSB/MSAB*	*MSB/MSE*
Interactions	*MSAB/MSE*	*MSAB/MSE*	*MSAB/MSE*

Construction of Test Statistics

As usual, each statistic for testing factor effects is constructed by comparing two mean squares that have the properties:

1. Under H_0, both mean squares have the same expectation.
2. Under H_a, the numerator mean square has a larger expectation than the denominator mean square.

It can be shown that such a test statistic follows the F distribution if H_0 holds. The decision rule is constructed in the ordinary fashion, with large values of the test statistic leading to H_a.

For instance, to test for the presence of factor A main effects in random ANOVA model (25.39), namely:

$$\begin{aligned} H_0&: \sigma_\alpha^2 = 0 \\ H_a&: \sigma_\alpha^2 > 0 \end{aligned} \tag{25.47}$$

we see from Table 25.5 that MSA and $MSAB$ both have the same expectation if $\sigma_\alpha^2 = 0$, that is, if factor A has no main effects. If $\sigma_\alpha^2 > 0$, $E\{MSA\}$ is greater than $E\{MSAB\}$. Hence, the appropriate test statistic is:

$$F^* = \frac{MSA}{MSAB} \tag{25.48}$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; a-1, (a-1)(b-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; a-1, (a-1)(b-1)], \text{ conclude } H_a \end{aligned} \tag{25.49}$$

Note that the denominator for testing for factor A main effects in the random ANOVA model is $MSAB$, whereas it is MSE in the fixed ANOVA model.

We summarize the appropriate test statistics for mixed and random ANOVA models in Table 25.6. For comparison purposes, we also present the test statistics for the fixed ANOVA model there. As may be seen from Table 25.6, the denominator of the test statistic for mixed and random ANOVA models in a number of instances differs from that for the fixed ANOVA model. Hence, it is important that the expected mean squares be known when random or mixed models are utilized so that the appropriate test statistics can be determined.

Example

We return to our earlier mixed ANOVA model example of four different training methods (factor A, fixed) and five instructors (factor B, random). Four classes were assigned to each training method–instructor combination. The response variable of interest was the mean

TABLE 25.7 ANOVA Table for Mixed ANOVA Model—Training Example (A fixed, B random, $a = 4$, $b = 5$, $n = 4$).

Source of Variation	*SS*	*df*	*MS*	F^*
Factor A (training methods, fixed)	42.1	3	14.0	$14.0/3.9 = 3.59$
Factor B (instructors, random)	53.9	4	13.5	$13.5/2.1 = 6.43$
AB interactions	46.7	12	3.9	$3.9/2.1 = 1.86$
Error	126.4	60	2.1	
Total	269.1	79		

$F(.95; 3, 12) = 3.49$ $F(.95; 4, 60) = 2.53$

$F(.95; 12, 60) = 1.92$

improvement per student in the class at the end of the training program. The data are not shown, but the ANOVA table is presented in Table 25.7. To test whether or not training methods and instructors interact:

$$H_0: \sigma^2_{\alpha\beta} = 0$$

$$H_a: \sigma^2_{\alpha\beta} > 0$$

we utilize according to Table 25.6 the test statistic:

$$F^* = \frac{MSAB}{MSE}$$

Using the results from Table 25.7, we obtain:

$$F^* = \frac{3.9}{2.1} = 1.86$$

For level of significance $\alpha = .05$, we require $F(.95; 12, 60) = 1.92$. Since $F^* = 1.86 \leq 1.92$, we conclude that training methods and instructors do not interact. The P-value of this test is .06.

The test statistics for testing training method main effects and instructor main effects are shown in Table 25.7. By comparing the test statistics with the appropriate percentiles of the F distribution shown at the bottom of Table 25.7 for level of significance $\alpha = .05$ each, we find that both training methods and instructors differ in effectiveness.

Comment

When there is only one case per treatment ($n = 1$) with the fixed two-factor ANOVA model, we know from Section 20.1 that no exact tests are possible unless the model can be modified. The reason is that $MSE = 0$ always in that case so that no estimate of σ^2 can be obtained. In contrast, Table 25.5 indicates that exact tests for both factor A and factor B main effects are possible with the random two-factor ANOVA model when $n = 1$ without any restrictive assumptions about the interactions. This is because $MSAB$ is the appropriate denominator of the test statistic here, and $MSAB$ can be determined regardless of sample size. With the mixed ANOVA model where factor A is the fixed factor, the presence of factor A main effects can also be tested when $n = 1$ without the need

for restrictive assumptions about the interactions. However, an exact test for factor B main effects would require the assumption that all interactions are zero or some other modification of the ANOVA model. ■

5.4 Two-Factor Studies—Estimation of Factor Effects for Models II and III

stimation of Variance Components

When a random factor has significant main effects, we often wish to estimate the magnitude of the variance component. Unbiased estimators can readily be derived from appropriate linear combinations of the expected mean squares in Table 25.5. For instance, the variance component σ_β^2 in mixed ANOVA model (25.42) can be estimated by noting that:

$$E\{MSB\} - E\{MSE\} = \sigma^2 + na\sigma_\beta^2 - \sigma^2 = na\sigma_\beta^2$$

Hence, we have:

$$\sigma_\beta^2 = \frac{E\{MSB\} - E\{MSE\}}{na} \tag{25.50}$$

and an unbiased estimator of σ_β^2 is:

$$s_\beta^2 = \frac{MSB - MSE}{na} \tag{25.50a}$$

Approximate confidence intervals for the variance components in balanced two-factor studies can be obtained by either the Satterthwaite procedure in (25.29) or the MLS procedure in (25.34). For example, the MLS procedure can be used to estimate the variance component σ_β^2 in mixed ANOVA model (25.42) by noting from (25.50a) that s_β^2 can be expressed in the form (25.33):

$$\hat{L} = s_\beta^2 = \left(\frac{1}{na}\right) MSB + \left(-\frac{1}{na}\right) MSE$$

The correspondences are $MS_1 = MSB$, $MS_2 = MSE$, $c_1 = 1/na$, and $c_2 = -1/na$. The approximate $1 - \alpha$ MLS confidence limits therefore are:

$$s_\beta^2 - H_L \leq \sigma_\beta^2 \leq s_\beta^2 + H_U \tag{25.51}$$

where H_L and H_U are determined using the formulas in Table 25.3, with $df_1 = b - 1$ and $df_2 = (n - 1)ab$.

Example

In the training example of Table 25.7 with one fixed and one random factor, random factor B (instructors) had significant effects. To estimate σ_β^2, we utilize the estimator in (25.50a). Substituting, we obtain:

$$s_\beta^2 = \frac{13.5 - 2.1}{16} = .71$$

To construct an approximate 95 percent confidence interval for σ_β^2 by the MLS procedure, we first note that the correspondences to the form in (25.33) are:

$$c_1 = \frac{1}{na} = \frac{1}{4(4)} = .0625 \qquad MS_1 = MSB = 13.5$$

$$c_2 = -\frac{1}{na} = -\frac{1}{4(4)} = -.0625 \qquad MS_2 = MSE = 2.1$$

$$df_1 = b - 1 = 4 \qquad df_2 = (n-1)ab = 60$$

Carrying out the calculations indicated in Table 25.3, we first obtain the percentiles:

$$F_1 = F(.975; 4, \infty) = 2.79 \qquad F_2 = (.975; 60, \infty) = 1.39$$

$$F_3 = F(.975; \infty, 4) = 8.26 \qquad F_4 = (.975; \infty, 60) = 1.48$$

$$F_5 = F(.975; 4, 60) = 3.01 \qquad F_6 = F(.975; 60, 4) = 8.36$$

and then:

$$G_1 = .6416 \qquad G_4 = -.4834$$

$$G_2 = .2806 \qquad H_L = .55$$

$$G_3 = .0266 \qquad H_U = 6.12$$

The desired confidence interval is obtained from (25.51):

$$.16 = .71 - .55 \leq \sigma_\beta^2 \leq .71 + 6.12 = 6.83$$

Hence, an approximate 95 percent confidence interval for σ_β, the standard deviation measuring the variability among instructors, is:

$$.4 \leq \sigma_\beta \leq 2.6$$

Estimation of Fixed Effects in Mixed Model

Point Estimators. We now consider point and interval estimation of fixed effect parameters for balanced mixed model (25.42), where factor A is fixed and factor B is random. The situation is more complicated than for fixed ANOVA model I because certain pairs of observations are correlated for the mixed model, as we have seen in (25.46). When the responses Y are correlated, the method of generalized least squares must be used to obtain minimum variance unbiased estimators. Weighted least squares, discussed in Chapter 11, is a special case of generalized least squares. It turns out, however, that the generalized least squares estimators of the fixed effects α_i for the balanced case are the same as the ones obtained by the method of ordinary least squares:

$$\hat{\alpha}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot} \tag{25.52}$$

Frequently, the marginal mean $\mu_{i\cdot}$ is also of interest. Since $\mu_{i\cdot} = \mu_{\cdot\cdot} + \alpha_i$, it follows from (25.52) that a best linear unbiased estimator of $\mu_{i\cdot}$ for balanced studies is:

$$\hat{\mu}_{i\cdot} = \bar{Y}_{\cdot\cdot\cdot} + (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}) = \bar{Y}_{i\cdot\cdot} \tag{25.53}$$

Often a contrast of the fixed effects α_i is also of interest:

$$L = \sum c_i \alpha_i \qquad \text{where} \qquad \sum c_i = 0 \tag{25.54}$$

An unbiased estimator of L is:

$$\hat{L} = \sum c_i \hat{\alpha}_i = \sum c_i (\bar{Y}_{i..} - \bar{Y}_{...}) = \sum c_i \bar{Y}_{i..} \tag{25.55}$$

Variances of Estimators. For mixed ANOVA model (25.42) for balanced studies, it can be shown that the variance of $\hat{\alpha}_i$ is as follows:

$$\sigma^2\{\hat{\alpha}_i\} = \frac{\sigma^2 + n\sigma^2_{\alpha\beta}}{bn} = \frac{E\{MSAB\}}{bn} \tag{25.56}$$

It can also be shown that the variance of a contrast $\hat{L}$ of the estimated fixed factor A effects $\hat{\alpha}_i$, defined in (25.55), is as follows:

$$\sigma^2\{\hat{L}\} = \sum c_i^2 \sigma^2\{\hat{\alpha}_i\} \tag{25.57}$$

where $\sigma^2\{\hat{\alpha}_i\}$ is given in (25.56).

Since $\sigma^2\{\hat{\alpha}_i\}$ is a constant multiple of an expected mean square, it can be estimated unbiasedly and exact confidence intervals for α_i and for contrasts of the α_i can be obtained. An unbiased estimator of the variance of $\hat{\alpha}_i$ is:

$$s^2\{\hat{\alpha}_i\} = \frac{MSAB}{bn} \tag{25.58}$$

and of a contrast of the $\hat{\alpha}_i$ is:

$$s^2\{\hat{L}\} = \frac{MSAB}{bn} \sum c_i^2 \tag{25.59}$$

Comment

The variances in (25.56) and (25.57) are obtained by recognizing that $\hat{\alpha}_i$ and $\hat{L}$ are linear combinations of the responses Y_{ijk}. For instance, consider an experiment with $a = b = n = 2$. Then $\hat{\alpha}_1$ in (25.52) is as follows:

$$\begin{aligned}
\hat{\alpha}_1 = \bar{Y}_{1..} - \bar{Y}_{...} &= \frac{1}{4}(Y_{111} + Y_{112} + Y_{121} + Y_{122}) - \frac{1}{8}(Y_{111} + \cdots + Y_{222}) \\
&= \left(\frac{1}{8}\right)Y_{111} + \left(\frac{1}{8}\right)Y_{112} + \left(\frac{1}{8}\right)Y_{121} + \left(\frac{1}{8}\right)Y_{122} + \left(-\frac{1}{8}\right)Y_{211} \\
&\quad + \left(-\frac{1}{8}\right)Y_{212} + \left(-\frac{1}{8}\right)Y_{221} + \left(-\frac{1}{8}\right)Y_{222}
\end{aligned}$$

Let the coefficients of the responses Y be denoted by c and define the row vector of the coefficients as follows:

$$\mathbf{c}' = (c_1 \quad c_2 \quad \cdots \quad c_{n_T})$$

and let $\mathbf{Y}$ as always denote the vector of the responses. We can then represent the estimator ($\hat{\alpha}_i$ or $\hat{L}$) as $\mathbf{c}'\mathbf{Y}$.

We know the variances and covariances of the responses Y_{ijk} from (25.45) and (25.46). Let $\boldsymbol{\sigma}^2\{\mathbf{Y}\}$, as usual, denote the variance-covariance matrix containing these variances and covariances. We then

utilize (5.46) to obtain the variance of the estimator, namely $\mathbf{c}'\boldsymbol{\sigma}^2\{\mathbf{Y}\}\mathbf{c}$. The resulting variance will be expressed in terms of the variance components σ^2, σ^2_{β}, and $\sigma^2_{\alpha\beta}$. We then use the expected mean squares in Table 25.5 for the mixed model to express the variance, if possible, in terms of expected mean squares. ■

Confidence Intervals for Fixed Effects Contrasts. It is not always possible to obtain exact confidence intervals for the fixed effects in mixed models. Exact confidence intervals are available only when the variance of the estimated parameter or contrast of interest is proportional to an expected mean square from the analysis of variance table. In cases where the variance is not directly proportional to an expected mean square, Satterthwaite's method can sometimes be used to construct approximate confidence intervals. For mixed ANOVA model (25.42), it is possible to obtain exact confidence intervals for contrasts of the fixed effects α_i because $\sigma^2\{\hat{\alpha}_i\}$ in (25.56) is a constant multiple of $E\{MSAB\}$. It can be shown that:

$$\frac{\hat{L} - L}{s\{\hat{L}\}} \text{ is distributed as } t[(a-1)(b-1)] \tag{25.60}$$

As a result, the $1 - \alpha$ confidence limits for L are:

$$\hat{L} \pm t[1 - \alpha/2; (a-1)(b-1)]s\{\hat{L}\} \tag{25.61}$$

where $\hat{L}$ is given by (25.55) and $s^2\{\hat{L}\}$ is given by (25.59).

Notice that confidence limits (25.61) are identical to those in (19.65) for the fixed ANOVA model, except that:

1. $MSAB$ replaces MSE in the estimated variance of the contrast.
2. The degrees of freedom now are $(a-1)(b-1)$ instead of $(n-1)ab$ since a different mean square is utilized.

Example

In the training example of Table 25.7, no interaction effects were found to be present. We now wish to estimate the difference $L = \alpha_1 - \alpha_2$ in the mean improvements between training methods 1 and 2, using a 95 percent confidence interval. The relevant sample results are:

$$\bar{Y}_{1..} = 43.1 \qquad \bar{Y}_{2..} = 40.8$$

Hence, our point estimate of $L = \alpha_1 - \alpha_2 = \mu_{1.} - \mu_{2.}$ is:

$$\hat{L} = \bar{Y}_{1..} - \bar{Y}_{2..} = 43.1 - 40.8 = 2.3$$

From (25.59), the estimated variance is:

$$s^2\{\hat{L}\} = \frac{MSAB}{bn}(1+1) = \frac{2(3.9)}{20} = .39$$

or $s\{\hat{L}\} = .62$. There are 12 degrees of freedom associated with $MSAB$; hence, we require $t(.975; 12) = 2.179$. The confidence limits (25.61) therefore are $2.3 \pm 2.179(.62)$ and the desired confidence interval is:

$$.9 \leq \mu_{1.} - \mu_{2.} \leq 3.7$$

Thus, we conclude with confidence coefficient .95 that training method 1 is more effective than training method 2, its mean improvement being somewhere between .9 and 3.7 units larger.

Multiple Comparison Procedures. Multiple comparison procedures can be utilized for the main effects of the fixed factor in a mixed two-factor ANOVA model in the same way as for the fixed ANOVA model. For example, suppose we wish to obtain all pairwise comparisons between the different training methods in the training example in Table 25.7 by means of the Tukey procedure. We would calculate $s^2\{\hat{L}\}$ as in the previous example. The t multiple in (25.61) now would be:

$$T = \frac{1}{\sqrt{2}} q[1-\alpha; a, (a-1)(b-1)] \tag{25.62}$$

With specific reference to the training example in Table 25.7, we would require for constructing 95 percent family confidence coefficient intervals for all pairwise comparisons between training methods:

$$q(.95; 4, 12) = 4.20 \qquad T = \frac{1}{\sqrt{2}}(4.20) = 2.97$$

Confidence Intervals for Marginal Means. An exact confidence interval for a marginal mean $\mu_{i\cdot}$ in mixed ANOVA model (25.42) cannot be obtained because the variance of the marginal mean $\hat{\mu}_{i\cdot}$ in (25.53) is not a multiple of a single expected mean square. Rather, the variance is a linear combination of two expected mean squares, as follows:

$$\sigma^2\{\hat{\mu}_{i\cdot}\} = c_1 E\{MSAB\} + c_2 E\{MSB\} \tag{25.63}$$

where:

$$c_1 = \frac{a-1}{nab} \tag{25.63a}$$

$$c_2 = \frac{1}{nab} \tag{25.63b}$$

An unbiased estimator of $\sigma^2\{\hat{\mu}_{i\cdot}\}$ is:

$$s^2\{\hat{\mu}_{i\cdot}\} = c_1 MSAB + c_2 MSB \tag{25.64}$$

Since the form of the variance of estimated marginal mean $\hat{\mu}_{i\cdot}$ is that in (25.25), the Satterthwaite approximation can be employed, where the degrees of freedom associated with the estimated variance $s^2\{\hat{\mu}_{i\cdot}\}$ are according to (25.28):

$$df = \frac{\left(\frac{a-1}{nab}MSAB + \frac{1}{nab}MSB\right)^2}{\frac{\left(\frac{a-1}{nab}MSAB\right)^2}{(a-1)(b-1)} + \frac{\left(\frac{1}{nab}MSB\right)^2}{b-1}} \tag{25.65}$$

Approximate $1-\alpha$ confidence limits for $\mu_{i\cdot}$ therefore are:

$$\hat{\mu}_{i\cdot} \pm t(1-\alpha/2; df) s\{\hat{\mu}_{i\cdot}\} \tag{25.66}$$

where $s^2\{\hat{\mu}_{i\cdot}\}$ is given in (25.64) and df is given in (25.65).

Example

Referring again to the training example of Table 25.7, a 95 percent confidence interval for $\mu_{1\cdot}$ is desired. As noted previously, the estimated mean improvement for training method 1 is:

$$\hat{\mu}_{1\cdot} = \bar{Y}_{1\cdot\cdot} = 43.1$$

Using (25.64) and noting that $nab = 4(4)5 = 80$, we obtain:

$$s^2\{\hat{\mu}_{1\cdot}\} = \frac{3}{80}(3.9) + \frac{1}{80}(13.5) = .315$$

or $s\{\hat{\mu}_{1\cdot}\} = .561$. From (25.65) we find:

$$df = \frac{\left[\frac{3}{80}(3.9) + \frac{1}{80}(13.5)\right]^2}{\frac{\left[\frac{3}{80}(3.9)\right]^2}{3(4)} + \frac{\left[\frac{1}{80}(13.5)\right]^2}{4}} = 11.1$$

Using $df = 11$, the required t percentile is $t(.975; 11) = 2.201$. The confidence limits are therefore $43.1 \pm 2.201(.561)$ and the desired confidence interval is:

$$41.9 \leq \mu_{1\cdot} \leq 44.3$$

We conclude with approximate confidence coefficient .95 that the mean improvement for training method 1 averaged over all instructors is between 41.9 and 44.3.

25.5 Randomized Complete Block Design: Random Block Effects

In our discussion of randomized complete block designs in Chapter 21, we assumed that block effects were fixed. However, when blocks are a random sample from a population, the block effects in the randomized complete block design model should be considered to be random variables, as in the following two examples.

1. A researcher investigated the improvement in learning in third-grade classes by augmenting the teacher with one or two teaching assistants. Ten schools were selected at random, and three third-grade classes in each school were utilized in the study. In each school, one class was randomly chosen to have no teaching assistant, one class was randomly chosen to have one teaching assistant, and the third class was assigned two teaching assistants. The amount of learning by the class at the end of the school year, suitably measured, was the response variable. Here the blocks are schools, which may be viewed as a random sample from the population of all schools eligible for the study.

2. In a study of the effectiveness of four different dosages of a drug, 20 litters of mice, each consisting of four mice, were utilized. The 20 litters (blocks) here may be viewed as a random sample from the population of all litters that could have been used for the study.

When blocks are considered to be a random sample from a population of blocks, either an additive (i.e., no-interaction) or a nonadditive (i.e., interaction) model can be employed. The choice can be assisted by the diagnostics discussed in Section 21.4. In particular, plots of the responses Y_{ij} for each block, such as in Figure 21.2, can be helpful in examining

whether blocks and treatments interact. A severe lack of parallelism in such a plot would be a clear indication that the interaction model may be preferable. The Tukey test statistic for interactions in (20.11) may also be utilized, with the interpretation here that the test applies to the given blocks that have been selected. Finally, the nature of the correlations between the experimental units within a block may be examined because the two models make different assumptions about these correlations.

When the primary emphasis of the analysis is on testing and estimating treatment effects, which is the usual case, the choice between the two models actually is not critical because the inference procedures for fixed treatment effects, as we shall see, are exactly the same for the two models.

We first explain the additive, no-interaction model for randomized block designs with fixed treatment effects and random block effects, and then we will take up the interaction model. Both of these models are special cases of two-factor mixed model (25.42). We shall repeat the principal results here because the notation for randomized block designs is slightly different.

Comment

A special case of random blocks occurs when the blocks are experimental units such as persons, stores, or cities, where each receives all of the treatments over time or where the effect of a given treatment (e.g., advertising) is evaluated at different points of time. These repeated measures designs are discussed in Chapter 27. ■

Additive Model

The additive model for random block effects and fixed treatment effects is a special case of mixed two-factor model (25.42), with $n = 1$, the interaction term dropped, and fixed factor A effects now being the treatment effects denoted by τ_j and random factor B effects now being the block effects denoted by ρ_i:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \varepsilon_{ij} \tag{25.67}$$

where:

$\mu_{..}$ is a constant
ρ_i are independent $N(0, \sigma_\rho^2)$
τ_j are constants subject to the restriction $\sum \tau_j = 0$
ε_{ij} are independent $N(0, \sigma^2)$, and independent of the ρ_i
$i = 1, \ldots, n_b; j = 1, \ldots, r$

Properties of Model. The important properties of mixed two-factor model (25.42) were given in (25.44)–(25.46). These properties for randomized complete block design model (25.67) are:

$$E\{Y_{ij}\} = \mu_{..} + \tau_j \tag{25.68a}$$

$$\sigma^2\{Y_{ij}\} = \sigma_Y^2 = \sigma_\rho^2 + \sigma^2 \tag{25.68b}$$

$$\sigma\{Y_{ij}, Y_{ij'}\} = \sigma_\rho^2 \qquad j \neq j' \tag{25.68c}$$

$$\sigma\{Y_{ij}, Y_{i'j'}\} = 0 \qquad i \neq i' \tag{25.68d}$$

Thus, the variance of Y_{ij}, again denoted by σ_Y^2, is a constant for all observations; any two observations from different blocks are independent; and any two observations from the same block are correlated for this model. Note that the covariance for any two observations from the same block must be positive in advance of the random trials and that the covariance is the same for all blocks. A positive covariance is reasonable for many applications. For example, class learning in different classes in the same school will tend to be more similar than for classes in different schools because of similar facilities, similar quality of teachers, and the like.

The coefficient of correlation between any two observations from the same block for model (25.67) is constant for all blocks and will be denoted by ω:

$$\omega = \frac{\sigma_\rho^2}{\sigma_Y \sigma_Y} = \frac{\sigma_\rho^2}{\sigma_Y^2} \tag{25.69}$$

This follows from the definition of a coefficient of correlation in (2.76) and the fact that $\sigma\{Y_{ij}\} = \sigma\{Y_{ij'}\} = \sigma_Y$. Note also that the covariance in (25.68c) can be expressed as follows, using (25.69):

$$\sigma\{Y_{ij}, Y_{ij'}\} = \omega\sigma_Y^2 \qquad j \neq j' \tag{25.70}$$

Covariance Structure of Observations. Since any two Y_{ij} observations within a given block in advance of the random trials are correlated in the same fashion, the variance-covariance matrix of the observations in a given block is of a particular form. We illustrate this variance-covariance matrix for the observations in a block for a randomized block study with $r = 3$ treatments, using the covariance expression in (25.70):

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \begin{bmatrix} \sigma_Y^2 & \omega\sigma_Y^2 & \omega\sigma_Y^2 \\ \omega\sigma_Y^2 & \sigma_Y^2 & \omega\sigma_Y^2 \\ \omega\sigma_Y^2 & \omega\sigma_Y^2 & \sigma_Y^2 \end{bmatrix} = \sigma_Y^2 \begin{bmatrix} 1 & \omega & \omega \\ \omega & 1 & \omega \\ \omega & \omega & 1 \end{bmatrix} \tag{25.71}$$

where:

$$\mathbf{Y} = \begin{bmatrix} Y_{i1} \\ Y_{i2} \\ Y_{i3} \end{bmatrix}$$

Note that the main diagonal of the matrix contains the constant variance of the Y_{ij}, σ_Y^2, and the entries off the main diagonal are the constant covariances, $\omega\sigma_Y^2$. The particular pattern of the variance-covariance matrix in (25.71) is called *compound symmetry*.

While any two observations in a given block are correlated in advance of the random trials, once a block has been selected, additive model (25.67) assumes that the observations in that block are independent. The only remaining random variation in an observation Y_{ij} then is the error term ε_{ij}, and additive model (25.67) assumes that these are independent. Thus, in the teacher assistant study, model (25.67) assumes that once the schools have been selected, any one class performance is independent of that of another class in each selected school, given all of the common conditions for the classes in that school as reflected in the block effect ρ_i.

Comments

1. The variance of Y_{ij} in (25.68b) can be expressed as follows using (25.69):

$$\sigma_Y^2 = \omega\sigma_Y^2 + \sigma^2$$

Hence, we obtain:

$$\sigma_Y^2 = \frac{\sigma^2}{1-\omega} \tag{25.72}$$

2. The assumption of compound symmetry in additive model (25.67) is restrictive. While this assumption is sufficient so that the F^* statistic for testing treatment effects will follow the F distribution when H_0 holds (i.e., when no treatment effects are present), the assumption is not necessary. For this purpose, it suffices that the condition of *sphericity* be met. This condition requires that the variance of the difference between any two estimated treatment means be constant; that is:

$$\sigma^2\{\bar{Y}_j - \bar{Y}_{j'}\} = \text{constant} \qquad j \neq j' \tag{25.73}$$

This condition can be met without the compound symmetry requirement. For example, consider the following variance-covariance matrix for the Y_{ij} observations in any block for a randomized complete block study with $r = 3$ treatments:

$$\boldsymbol{\sigma}^2\{\mathbf{Y}\} = \begin{bmatrix} 2 & 2 & 4 \\ 2 & 4 & 5 \\ 4 & 5 & 8 \end{bmatrix}$$

This matrix does not exhibit compound symmetry. Yet the requirement for sphericity in (25.73) is met because $\sigma^2\{\bar{Y}_j - \bar{Y}_{j'}\} = 2/n_b$ always. For example, we have:

$$\sigma^2\{\bar{Y}_1 - \bar{Y}_3\} = \frac{2}{n_b} + \frac{8}{n_b} - 2\left(\frac{4}{n_b}\right) = \frac{2}{n_b}$$

■

Analysis of Variance. Table 25.8 contains the analysis of variance for additive model (25.67). The sums of squares are the same as in (21.6) for the fixed effects model. Table 25.8 also contains the expected mean squares for model (25.67). The expected mean

TABLE 25.8 ANOVA for Randomized Complete Block Design—Block Effects Random, Treatment Effects Fixed.

				$E\{MS\}$	
Source of Variation	*SS*	*df*	*MS*	Additive Model (25.67)	Interaction Model (25.74)
Blocks	*SSBL*	$n_b - 1$	*MSBL*	$\sigma^2 + r\sigma_\rho^2$	$\sigma^2 + r\sigma_\rho^2$
Treatments	*SSTR*	$r - 1$	*MSTR*	$\sigma^2 + n_b\frac{\sum \tau_j^2}{r-1}$	$\sigma^2 + \sigma_{\rho\tau}^2 + n_b\frac{\sum \tau_j^2}{r-1}$
Error	*SSBL.TR*	$(n_b - 1)(r - 1)$	*MSBL.TR*	σ^2	$\sigma^2 + \sigma_{\rho\tau}^2$
Total	*SSTO*	$n_b r - 1$			

squares correspond to those in Table 25.5 for the mixed two-factor model, with $n = 1$, no interaction effects, and change of notation associated with fixed factor A being treatments and random factor B being blocks. The statistic for testing for treatment effects is $F^* = MSTR/MSBL.TR$, as may be seen from the $E\{MS\}$ column in Table 25.8. Thus, the test statistic is the same as when block effects are fixed. Confidence intervals for treatment contrasts also present no new issues. Again, $MSBL.TR$ will be used as the mean square in the estimated variance of the contrast.

Interaction Model

When blocks are a random sample from a population of blocks, the presence of interactions between blocks and treatments can be accommodated by a model including these interaction effects:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + (\rho\tau)_{ij} + \varepsilon_{ij} \tag{25.74}$$

where:

$\mu_{..}$ is a constant

ρ_i are independent $N(0, \sigma_\rho^2)$

τ_j are constants subject to the restriction $\sum \tau_j = 0$

$(\rho\tau)_{ij}$ are $N\left(0, \frac{r-1}{r}\sigma_{\rho\tau}^2\right)$, subject to the restrictions:

$$\sum_j (\rho\tau)_{ij} = 0 \qquad \text{for all } i$$

$$\sigma\{(\rho\tau)_{ij}, (\rho\tau)_{ij'}\} = -\frac{1}{r}\sigma_{\rho\tau}^2 \qquad \text{for } j \neq j'$$

$(\rho\tau)_{ij}$ are independent of the ρ_i

ε_{ij} are independent $N(0, \sigma^2)$ and independent of the ρ_i and of the $(\rho\tau)_{ij}$

$i = 1, \ldots, n_b; j = 1, \ldots, r$

This model is a special case of mixed two-factor model (25.42), with $n = 1$ and with some changes in notation to recognize that fixed factor A now is treatments and random factor B now is blocks.

Properties of Model. The properties of interaction model (25.74) are obtained directly from those in (25.44)–(25.46) for the mixed two-factor model:

$$E\{Y_{ij}\} = \mu_{..} + \tau_j \tag{25.75a}$$

$$\sigma^2\{Y_{ij}\} = \sigma_Y^2 = \sigma_\rho^2 + \frac{r-1}{r}\sigma_{\rho\tau}^2 + \sigma^2 \tag{25.75b}$$

$$\sigma\{Y_{ij}, Y_{ij'}\} = \sigma_\rho^2 - \frac{1}{r}\sigma_{\rho\tau}^2 \qquad j \neq j' \tag{25.75c}$$

$$\sigma\{Y_{ij}, Y_{i'j'}\} = 0 \qquad i \neq i' \tag{25.75d}$$

Note again that the Y_{ij} have constant variance, that observations from different blocks are assumed to be independent, and that any two observations Y_{ij} and $Y_{ij'}$ from the same block are correlated, the covariance being the same for all blocks. Unlike for additive model

(25.67), the covariance between any two observations from the same block can be negative or positive for interaction model (25.74).

The coefficient of correlation between any two observations in the same block, denoted by ω^*, is:

$$\omega^* = \frac{\sigma_\rho^2 - \frac{1}{r}\sigma_{\rho\tau}^2}{\sigma_Y^2} \tag{25.76}$$

Interaction model (25.74) assumes, just like additive model (25.67), that, once the blocks have been selected, any two observations from a given block are uncorrelated.

Analysis of Variance. The sums of squares and degrees of freedom for interaction model (25.74) are the same as those for additive model (25.67). The principal difference in the use of the two models occurs because of the difference in the expected mean squares, as shown in Table 25.8. No exact test for block effects is possible with the interaction model, whereas an exact test is possible with the additive model. This distinction is unimportant whenever blocks are used primarily to reduce the experimental error variability and are not of intrinsic interest themselves.

The F^* test statistic for treatment effects is the same for the two models, namely $F^* = MSTR/MSBL.TR$, which is exactly the same as test statistic (21.7b) for randomized block model (21.1) with fixed block effects. Similarly, estimation of fixed treatment effects for both models with random block effects is carried out in the manner described in Section 21.3 for fixed block effects.

Comments

1. Table 25.8 indicates that when the block effects are random, *MSBL.TR* estimates σ^2 for additive model (25.67). For interaction model (25.74), however, *MSBL.TR* estimates the sum of the error term variance σ^2 and the interaction variance $\sigma_{\rho\tau}^2$. Separate estimation of these two components is not possible for this latter model, and the two components are said to be confounded. This problem can be avoided by utilizing replication within blocks described in Section 21.7.

2. When the assumption of compound symmetry, which underlies both additive model (25.67) and interaction model (25.74), and the less restrictive requirement of sphericity are not met, the usual F test becomes biased. Some computer packages provide the user with the option of formally testing for compound symmetry or sphericity.

When these conditions are violated, an approximate conservative test procedure is as follows:

a. Conduct the usual F test; if it leads to conclusion H_0, accept this conclusion.

b. If the usual F test leads to conclusion H_a, replace $F[1-\alpha; r-1, (n_b-1)(r-1)]$ in decision rule (21.7c) by $F(1-\alpha; 1, n_b-1)$. If this modified decision rule leads to H_a, accept this conclusion.

c. If the modified decision rule leads to H_0, revise the degrees of freedom in the modified decision rule by one of the *epsilon adjustment procedures,* as described in References 25.8 and 25.9.

Alternatively, multivariate analysis of variance techniques may be employed provided that $n_b > r$. See Reference 25.10 for further discussions of these issues.

3. Mixed models based on less restrictive assumptions regarding the variance-covariance matrix and the parameters in the ANOVA model have also been proposed. See Reference 25.7 for a discussion of these models. ■

25.6 Three-Factor Studies—ANOVA Models II and III

Just as for single-factor and two-factor studies, the analysis of variance sums of squares and degrees of freedom for random and mixed multi-factor models are the same as those for the corresponding fixed ANOVA model. The principal issue with random and mixed multi-factor models, as we saw for two-factor models, is the determination of the expected mean squares. Once these are known, the proper test statistics and confidence intervals can be constructed. Rules for finding expected mean squares for random and mixed models are given in Appendix D for balanced studies with any number of factors. We now present model II (random factor levels) and model III (mixed factor levels) for three-factor studies and show how appropriate tests are conducted. We consider again the balanced case where all treatment sample sizes are equal.

ANOVA Model II—Random Factor Effects

In a study of the effects of operators, machines, and batches of raw material on daily output, all three factors may be considered to have random factor levels. The random ANOVA model for such a three-factor study is:

$$Y_{ijkm} = \mu_{...} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkm} \quad \textbf{(25.77)}$$

where:

$\mu_{...}$ is a constant

$\alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{ij}, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}, (\alpha\beta\gamma)_{ijk}, \varepsilon_{ijkm}$ are independent normal random variables with expectations zero and respective variances $\sigma^2_\alpha, \sigma^2_\beta, \sigma^2_\gamma, \sigma^2_{\alpha\beta}, \sigma^2_{\alpha\gamma}, \sigma^2_{\beta\gamma}, \sigma^2_{\alpha\beta\gamma}, \sigma^2$

$i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, c; m = 1, \ldots, n$

Just as for two-factor random ANOVA model (25.39), the responses Y_{ijkm} for three-factor random ANOVA model (25.77) are normally distributed with constant variance. The expected value and variance of response Y_{ijkm} are:

$$E\{Y_{ijkm}\} = \mu_{...} \quad \textbf{(25.78a)}$$

$$\sigma^2\{Y_{ijkm}\} = \sigma^2_Y = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_\gamma + \sigma^2_{\alpha\beta} + \sigma^2_{\alpha\gamma} + \sigma^2_{\beta\gamma} + \sigma^2_{\alpha\beta\gamma} + \sigma^2 \quad \textbf{(25.78b)}$$

Any two responses are independent except when they have one or more common factor levels; these latter are correlated because they contain some common random terms.

Table 25.9 contains the degrees of freedom and the expected mean squares for all components of the ANOVA table for random ANOVA model (25.77).

ANOVA Model III—Mixed Factor Effects

Consider a three-factor study where factors B and C have random factor levels while factor A has fixed factor levels. The restricted mixed ANOVA model for such a three-factor balanced study is:

$$Y_{ijkm} = \mu_{...} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \varepsilon_{ijkm} \quad \textbf{(25.79)}$$

LE 25.9
ected Mean
ares for
anced
dom
e-Factor
OVA Model
25.77).

Mean Square	*df*	Expected Mean Square
MSA	$a-1$	$\sigma^2 + nbc\sigma^2_{\alpha} + nc\sigma^2_{\alpha\beta} + nb\sigma^2_{\alpha\gamma} + n\sigma^2_{\alpha\beta\gamma}$
MSB	$b-1$	$\sigma^2 + nac\sigma^2_{\beta} + nc\sigma^2_{\alpha\beta} + na\sigma^2_{\beta\gamma} + n\sigma^2_{\alpha\beta\gamma}$
MSC	$c-1$	$\sigma^2 + nab\sigma^2_{\gamma} + nb\sigma^2_{\alpha\gamma} + na\sigma^2_{\beta\gamma} + n\sigma^2_{\alpha\beta\gamma}$
MSAB	$(a-1)(b-1)$	$\sigma^2 + nc\sigma^2_{\alpha\beta} + n\sigma^2_{\alpha\beta\gamma}$
MSAC	$(a-1)(c-1)$	$\sigma^2 + nb\sigma^2_{\alpha\gamma} + n\sigma^2_{\alpha\beta\gamma}$
MSBC	$(b-1)(c-1)$	$\sigma^2 + na\sigma^2_{\beta\gamma} + n\sigma^2_{\alpha\beta\gamma}$
MSABC	$(a-1)(b-1)(c-1)$	$\sigma^2 + n\sigma^2_{\alpha\beta\gamma}$
MSE	$(n-1)abc$	σ^2

where:

$\mu_{...}$ is a constant

α_i are constants

β_j, γ_k, $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$, $(\alpha\beta\gamma)_{ijk}$ are pairwise independent normal random variables with expectations zero and constant variances

ε_{ijkm} are independent $N(0, \sigma^2)$, and are independent of the other random components

$\sum_i \alpha_i = \sum_i (\alpha\beta)_{ij} = \sum_i (\alpha\gamma)_{ik} = \sum_i (\alpha\beta\gamma)_{ijk} = 0$

$i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, c; m = 1, \ldots, n$

Note that all interaction terms in this model are random, since at least one of the factors contained in each has random factor levels. Note also that all sums of effects over the fixed factor levels are zero. Various correlations exist between the random effects terms, which we shall not detail.

The responses Y_{ijkm} for three-factor mixed ANOVA model (25.79) are normally distributed with constant variance. The expected value of observation Y_{ijkm} is:

$$E\{Y_{ijkm}\} = \mu_{...} + \alpha_i \tag{25.80}$$

In advance of the random trials, any two responses are independent except for those that contain common and/or correlated random effects terms; these observations are correlated.

Table 25.10 contains all the expected mean squares for mixed ANOVA model (25.79).

Other mixed ANOVA models can be developed in similar fashion. The expected mean squares for these mixed models can be found by employing the rules presented in Appendix D

Appropriate Test Statistics

From the expected mean squares, we seek to determine the appropriate F^* statistic for a given test. An exact test statistic can often be found for random and mixed multi-factor models, but not always.

Exact *F* Test. Suppose we wish to determine whether or not *BC* interactions are present in random ANOVA model (25.77). We see from Table 25.9 that the appropriate test statistic is *MSBC*/*MSABC*. If we wish to study the same question for mixed ANOVA model (25.79),

TABLE 25.10 Expected Mean Squares for Balanced Mixed Three-Factor ANOVA Model (25.79) (*A* fixed, *B* and *C* random).

Mean Square	*df*	Expected Mean Square
MSA	$a-1$	$\sigma^2 + nbc\frac{\sum \alpha_i^2}{a-1} + nc\sigma^2_{\alpha\beta} + nb\sigma^2_{\alpha\gamma} + n\sigma^2_{\alpha\beta\gamma}$
MSB	$b-1$	$\sigma^2 + nac\sigma^2_{\beta} + na\sigma^2_{\beta\gamma}$
MSC	$c-1$	$\sigma^2 + nab\sigma^2_{\gamma} + na\sigma^2_{\beta\gamma}$
MSAB	$(a-1)(b-1)$	$\sigma^2 + nc\sigma^2_{\alpha\beta} + n\sigma^2_{\alpha\beta\gamma}$
MSAC	$(a-1)(c-1)$	$\sigma^2 + nb\sigma^2_{\alpha\gamma} + n\sigma^2_{\alpha\beta\gamma}$
MSBC	$(b-1)(c-1)$	$\sigma^2 + na\sigma^2_{\beta\gamma}$
MSABC	$(a-1)(b-1)(c-1)$	$\sigma^2 + n\sigma^2_{\alpha\beta\gamma}$
MSE	$(n-1)abc$	σ^2

we see from Table 25.10 that an appropriate test statistic is available, but this time it is $MSBC/MSE$. We thus note that the two test statistics are not the same, even though the same factor effects are being studied, because of the differences between the two models.

It is not always possible to find an exact F test for mixed and random multi-factor ANOVA models. For instance, we cannot directly test for the presence of factor A main effects in random ANOVA model (25.77). Note from Table 25.9 that there is no expected mean square that consists of the components of $E\{MSA\}$ except for the $nbc\sigma^2_{\alpha}$ term.

Sometimes it is possible to assume that certain interactions are zero, and then proceed in the usual way with an exact F test. For example, to test for factor A main effects in random ANOVA model (25.77) (see Table 25.9), it may be possible to assume that $\sigma^2_{\alpha\gamma} = 0$ (indeed, this can be tested with $MSAC/MSABC$). If this assumption is appropriate, we can use the test statistic $MSA/MSAB$ to test for factor A main effects.

Satterthwaite Approximate F Test. Often, it is not known whether certain interactions are zero. In that case, an approximate F test may be employed that utilizes a *pseudo* F or *quasi* F test statistic. This approximate test, called the *Satterthwaite test,* involves developing a linear combination of mean squares that has the same expectation when H_0 holds as the factor effects mean square. As noted in our discussion of the Satterthwaite procedure for constructing approximate confidence limits for variance components, this linear combination is expressed in the form:

$$\hat{L} = c_1 MS_1 + \cdots + c_h MS_h$$

where the c_i are constants. The approximate number of degrees of freedom associated with this linear combination of mean squares is given by (25.28). The test statistic is then set up in the usual way and follows approximately the F distribution when H_0 holds.

We illustrate this procedure for testing factor A main effects in random ANOVA model (25.77):

$$H_0: \sigma^2_{\alpha} = 0 \qquad H_a: \sigma^2_{\alpha} > 0 \tag{25.81}$$

BLE 25.11
OVA Table
or Random
ee-Factor
udy ($a = 3$,
$= 2$, $c = 5$,
$= 3$).

Source of Variation	*SS*	*df*	*MS*
Factor *A* (operators)	17.3	2	8.65
Factor *B* (machines)	4.2	1	4.20
Factor *C* (batches)	24.8	4	6.20
AB interactions	4.8	2	2.40
AC interactions	31.7	8	3.96
BC interactions	12.5	4	3.13
ABC interactions	11.9	8	1.49
Error	137.7	60	2.30
Total	244.9	89	

Note from Table 25.9 that:

$$E\{MSAB\} + E\{MSAC\} - E\{MSABC\} = \sigma^2 + nc\sigma^2_{\alpha\beta} + nb\sigma^2_{\alpha\gamma} + n\sigma^2_{\alpha\beta\gamma} \quad \textbf{(25.82)}$$

This equals precisely $E\{MSA\}$ when $\sigma^2_\alpha = 0$. Hence, the suggested test statistic is:

$$F^{**} = \frac{MSA}{MSAB + MSAC - MSABC} \quad \textbf{(25.83)}$$

where we denote the test statistic as F^{**} as a reminder that a pseudo F test is involved.

Example

Table 25.11 contains the analysis of variance for a study of the effects of operators, machines, and batches on the daily output of a highly automated process. Each factor is assumed to be a random factor. To test whether operators (factor A) have a main effect on output, we use test statistic (25.83):

$$F^{**} = \frac{8.65}{2.40 + 3.96 - 1.49} = \frac{8.65}{4.87} = 1.78$$

The approximate number of degrees of freedom associated with the denominator is, from (25.28):

$$df = \frac{(4.87)^2}{\dfrac{(2.40)^2}{2} + \dfrac{(3.96)^2}{8} + \dfrac{(-1.49)^2}{8}} = 4.63$$

which we round to 5. For level of significance $\alpha = .05$, we require $F(.95; 2, 5) = 5.79$. Since $F^{**} = 1.78 \leq 5.79$, we conclude H_0, that operators do not have a main effect on daily output.

Comment

Since the Satterthwaite pseudo F test is an approximate one, it must be employed with caution. Some alternative procedures are provided in References 25.2 and 25.11. ■

Estimation of Effects

No new problems arise in the estimation of variance components for random factors or in the estimation of contrasts for fixed factors in mixed models, when three or more factors

are studied at one time. Confidence limits for contrasts of the factor level means of a fixed factor are obtained by using the mean square utilized in the denominator of the test statistic for examining the presence of main effects for that factor. The degrees of freedom are those associated with the mean square utilized.

25.7 ANOVA Models II and III with Unequal Sample Sizes

We noted in Chapter 23 for the fixed two-factor ANOVA model that unequal treatment sample sizes make the analysis of variance more complicated because the sums of squares no longer are orthogonal. Tests of hypotheses must then be based on the general linear test approach. When sample sizes are unequal for studies involving random effects, the level of complexity increases in a similar fashion. Most of the methods described thus far for two-factor and multi-factor ANOVA models II and III do not apply to unbalanced studies. For example, in unbalanced studies typically neither exact nor Satterthwaite approximate F tests exist.

A number of alternative approaches have been developed for making inferences for ANOVA models II and III in the presence of unequal sample sizes. We shall discuss an approach based on the method of maximum likelihood. This approach has the advantage of conceptual simplicity and is a general procedure that possesses a number of optimality properties. Detailed discussions of this and alternative approaches can be found in References 25.2, 25.7, and 25.12.

We shall illustrate the maximum likelihood approach using an example involving a two-factor experiment where one factor has fixed factor levels and the second factor has random factor levels.

Example

The Sheffield Foods Company markets a variety of dairy products, including milk, ice cream, and yogurt. Recently, the company received a complaint from a government agency that the actual levels of milkfat in its yogurt exceeded the labeled amount. Company personnel were concerned that the government's laboratory method for measuring fat content in yogurt might be unreliable because it is primarily designed for use with milk and ice cream. To study the reliability of Sheffield's and the government's laboratory methods, a small interlaboratory study was carried out. Four testing laboratories were randomly selected from the population of laboratories in the United States. Each laboratory was sent 12 samples of yogurt, with instructions to evaluate six of the samples using the government's method and six by the company's method. The yogurt had been mixed under carefully controlled conditions and the fat content of each sample was known to be 3.0 percent.

In this study, measurement method is a fixed factor with $a = 2$ levels ($i = 1$: Government method; $i = 2$: Sheffield method) and laboratories is a random factor with $b = 4$ levels. Because of technical difficulties with the Government method, none of the laboratories was able to obtain fat content determinations for all of the six samples assigned to that method in the time available. The results of the study are given in Table 25.12. Figure 25.3 contains dot plots of the data. The variability of the sample fat determinations appears to be reasonably constant for all measurement method–laboratory combinations. Figure 25.4 contains a MINITAB estimated treatment means plot. For the four laboratories included in the study, no major interaction effects between laboratory and measurement method on fat content determination appear to be present. The plot suggests

LE 25.12
oratory Fat ontent Deter- ations— effield Foods pany ample.

Measurement Method	k	Laboratory $j=1$	$j=2$	$j=3$	$j=4$
$i=1$ Government	1	5.19	4.09	4.62	3.71
	2	5.09	3.99	4.32	3.86
	3		3.75	4.35	3.79
	4		4.04	4.59	3.63
	5		4.06		
$i=2$ Sheffield	1	3.26	3.02	3.08	2.98
	2	3.48	3.32	2.95	2.89
	3	3.24	2.83	2.98	2.75
	4	3.41	2.96	2.74	3.04
	5	3.35	3.23	3.07	2.88
	6	3.04	3.07	2.70	3.20

FIGURE 25.3
Dot Plots of Fat Content eterminations by Laboratory and Measurement Method— Sheffield Foods Company Example.

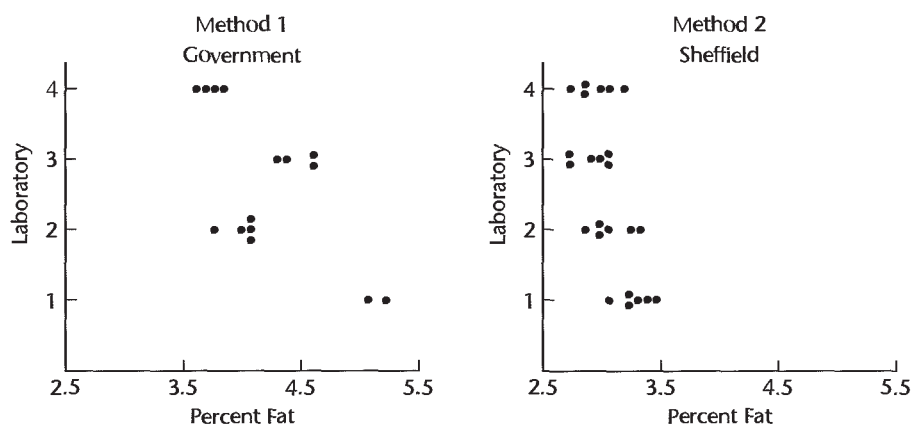
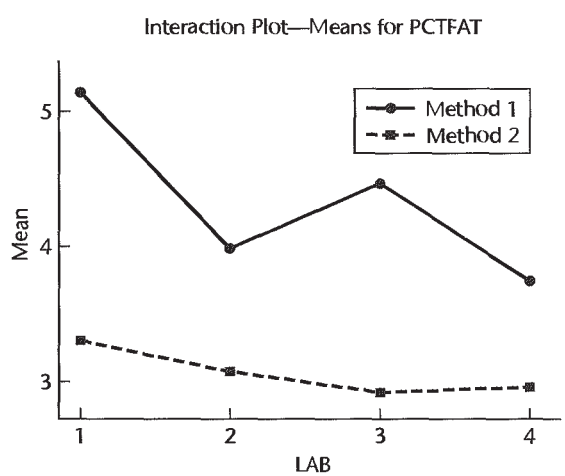


FIGURE 25.4
MINITAB Estimated Treatment Means Plot—Sheffield Foods Company Example.



a definite measurement method effect and possibly also some differences between laboratories. We shall now analyze the data formally by means of the maximum likelihood approach.

Maximum Likelihood Approach

The maximum likelihood approach that we will utilize for the Sheffield Foods Company example makes somewhat stronger assumptions than mixed ANOVA model (25.42), which we would use if the study were balanced. We first review mixed ANOVA model (25.42) as it applies to the Sheffield Foods Company example.

Mixed ANOVA Model (25.42). This model for the Sheffield Foods Company example is as follows:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \tag{25.84}$$

$$\sigma^2\{\beta_j\} = \sigma_\beta^2$$

$$\sigma^2\{(\alpha\beta)_{ij}\} = \frac{2-1}{2}\sigma_{\alpha\beta}^2 = \frac{\sigma_{\alpha\beta}^2}{2}$$

$$\sigma^2\{\varepsilon_{ijk}\} = \sigma^2$$

$$i = 1, 2;\ j = 1, \ldots, 4;\ k = 1, \ldots, n_{ij}$$

For this model, the expected value and variance of Y_{ijk} are according to (25.44) and (25.45):

$$E\{Y_{ijk}\} = \mu_{..} + \alpha_i \tag{25.85}$$

$$\sigma^2\{Y_{ijk}\} = \sigma_Y^2 = \sigma_\beta^2 + \frac{\sigma_{\alpha\beta}^2}{2} + \sigma^2 \tag{25.86}$$

Also, the responses Y_{ijk} are correlated as follows according to (25.46):

$$\sigma\{Y_{ijk}, Y_{ijk'}\} = \sigma_\beta^2 + \frac{\sigma_{\alpha\beta}^2}{2} \qquad k \neq k' \tag{25.87a}$$

$$\sigma\{Y_{ijk}, Y_{i'jk'}\} = \sigma_\beta^2 - \frac{\sigma_{\alpha\beta}^2}{2} \qquad i \neq i' \tag{25.87b}$$

$$\sigma\{Y_{ijk}, Y_{i'j'k'}\} = 0 \qquad j \neq j' \tag{25.87c}$$

We also know that the responses Y_{ijk} for mixed ANOVA model (25.42) are normally distributed.

Since the expected value of Y_{ijk} depends only on the fixed effects $\mu_{..}$ and α_i (the random effects have expectations zero), we can represent the vector of expected values, $\mathbf{E}\{\mathbf{Y}\}$, in the matrix form $\mathbf{X\beta}$. We illustrate this in Table 25.13 for the Sheffield Foods Company example. This table contains the vector of responses $\mathbf{Y}$, the vector of parameters $\boldsymbol{\beta}$, and the $\mathbf{X}$ matrix containing the usual column of 1s associated with $\mu_{..}$ and an indicator variable taking on the values 1 and -1 associated with α_1. Recall that $\alpha_2 = -\alpha_1$ since $\sum \alpha_i = 0$.

The variance-covariance matrix of the responses Y_{ijk}, $\boldsymbol{\sigma}^2\{\mathbf{Y}\}$, has on the main diagonal the constant variance from (25.86) and off the main diagonal the covariances from (25.87). We illustrated such a variance-covariance matrix in Table 25.4 for a study in which $a = b = n = 2$.

TABLE 25.13 Matrix Formulation—Sheffield Foods Company Example.

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ \vdots \\ Y_{144} \\ Y_{211} \\ Y_{212} \\ Y_{213} \\ Y_{214} \\ \vdots \\ Y_{246} \end{bmatrix} = \begin{bmatrix} 5.19 \\ 5.09 \\ 4.09 \\ 3.99 \\ \vdots \\ 3.63 \\ 3.26 \\ 3.48 \\ 3.24 \\ 3.41 \\ \vdots \\ 3.20 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ \vdots & \vdots \\ 1 & -1 \end{bmatrix} \qquad \boldsymbol{\beta} = \begin{bmatrix} \mu_{..} \\ \alpha_1 \end{bmatrix}$$

Density Function. To employ the method of maximum likelihood, we make a somewhat stronger assumption than with ANOVA model (25.42). We assume all of the properties of model (25.42) and in addition assume that the Y_{ijk} are jointly normally distributed. The density function of the multivariate normal distribution is given in (5.50). The mean vector $\boldsymbol{\mu}$ in (5.50) corresponds here to $\mathbf{X}\boldsymbol{\beta}$, and the variance-covariance matrix $\boldsymbol{\Sigma}$ in (5.50) corresponds to $\boldsymbol{\sigma}^2\{\mathbf{Y}\}$. We shall continue to use $\boldsymbol{\Sigma}$ to represent the variance-covariance matrix of the responses Y_{ijk}. The number of Y variables p in (5.50) corresponds here to n_T. We can then express the joint density function of the responses Y_{ijk} as follows:

$$f(\mathbf{Y}) = \frac{1}{(2\pi)^{n_T/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right] \tag{25.88}$$

Viewing the joint density as a function of the unknown parameters (for the Sheffield Foods Company example, $\mu_{..}$ and α_1 in $\boldsymbol{\beta}$ and σ^2, σ^2_β, and $\sigma^2_{\alpha\beta}$ in $\boldsymbol{\Sigma}$), given the observations Y_{ijk}, the function in (25.88) is called the likelihood function and denoted by L.

Maximum Likelihood Estimates. To obtain the maximum likelihood estimates of the unknown parameters, it is easiest to work with the logarithm of the likelihood function:

$$\log_e L = -\frac{n_T}{2}\log_e(2\pi) - \frac{1}{2}\log_e|\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) \tag{25.89}$$

The maximum likelihood estimates of $\mu_{..}$, α_1, σ^2, σ^2_β, and $\sigma^2_{\alpha\beta}$ for the Sheffield Foods Company example are those values of these parameters that maximize the log-likelihood function in (25.89), subject to the constraints that the variance components are nonnegative. For unbalanced studies, numerical search procedures are generally required to obtain the maximum likelihood estimates. We shall rely on standard statistical software programs to carry out the numerical search procedures.

Inference Procedures. Inference procedures are analogous to those explained in Chapter 14 for maximum likelihood estimation of the regression parameters in logistic regression. The estimated approximate variance-covariance matrix of the estimated parameters

is obtained through the Hessian matrix in (14.50), which contains the second-order partial derivatives of the logarithm of the likelihood function with respect to the parameters. This estimated variance-covariance matrix is usually provided by a statistical package in conjunction with the numerical search for maximum likelihood estimates.

Large-sample inference procedures are described in Chapter 14. In the Sheffield Foods Company example, for instance, the following approximate result for estimating the fixed laboratory method effect α_1 is obtained from (14.52):

$$\frac{\hat{\alpha}_1 - \alpha_1}{s\{\hat{\alpha}_1\}} \sim z \tag{25.90}$$

An approximate confidence interval for α_1 or a test concerning α_1 can then be developed readily. Simultaneous estimation of several parameters can be done as usual by means of the Bonferroni procedure. Tests whether several parameters equal zero (e.g., $\sigma^2_\beta = \sigma^2_{\alpha\beta} = 0$) are carried out by fitting the full and reduced models and obtaining the likelihood ratio test statistic (14.60). This test should not be used if any of the estimated variance components equals zero.

Often, there is interest in a linear combination of the parameters. For instance, the marginal mean $\mu_{1\cdot}$ may be of interest in the Sheffield Foods Company example. Since $\mu_{1\cdot} = \mu_{\cdot\cdot} + \alpha_1$, the maximum likelihood estimator of this quantity is the following linear combination of the estimated parameters:

$$\hat{\mu}_{1\cdot} = \hat{\mu}_{\cdot\cdot} + \hat{\alpha}_1 = (1 \quad 1 \quad 0 \quad 0 \quad 0) \begin{bmatrix} \hat{\mu}_{\cdot\cdot} \\ \hat{\alpha}_1 \\ \hat{\sigma}^2 \\ \hat{\sigma}^2_\beta \\ \hat{\sigma}^2_{\alpha\beta} \end{bmatrix} \tag{25.91}$$

Denoting the row vector of coefficients by $\mathbf{c}'$, we use (5.46) to obtain the estimated variance of $\hat{\mu}_{1\cdot}$:

$$s^2\{\hat{\mu}_{1\cdot}\} = \mathbf{c}'\mathbf{s}^2\{\mathbf{b}\}\mathbf{c} \tag{25.92}$$

where $\mathbf{s}^2\{\mathbf{b}\}$ is the estimated approximate variance-covariance matrix of the parameter estimates. Large-sample inferences are then conducted in the usual manner, utilizing the standard normal distribution.

We caution again that the inference procedures discussed here require large sample sizes. In studies with random factor levels, the number of factor levels frequently is not large. For instance, in the Sheffield Foods Company example only four laboratories were employed in the study. Use of a much larger number of laboratories would have been much too costly. An estimate of interlaboratory variability based on four randomly selected laboratories is likely not to be precise and use of a large-sample approximation for obtaining an interval estimate may not be appropriate.

Bootstrapping, as explained in Chapter 11, may be used to examine the appropriateness of large-sample inference procedures for maximum likelihood estimates in unbalanced studies. However, in some cases bootstrapping for variance components may not perform properly, which could be an indication that large-sample inference procedures are not appropriate.

Example

In the Sheffield Foods Company example, the investigators were primarily interested in determining whether the two different measurement methods yield systematic differences in the determination of fat content. The BMDP3V computer package was used, together with transformations (25.43) to go from the unrestricted to the restricted model, to obtain the maximum likelihood estimates of the parameters in the log-likelihood function (25.89) for the mixed ANOVA model. Table 25.14a contains the maximum likelihood estimates of the parameters and the estimated approximate standard deviations of these estimates. Table 25.14b contains the estimated approximate variance-covariance matrix of the maximum likelihood estimates obtained through the Hessian matrix in (14.50).

Since the sample sizes are not large here, bootstrapping was employed to examine whether the large-sample inference procedures for maximum likelihood estimates described in Chapter 14 are appropriate. Five hundred bootstrap samples were generated, the maximum likelihood estimates were obtained for each using SAS PROC MIXED, and a bootstrap distribution of the parameter estimates was created for each parameter. Table 25.15 contains the means and standard deviations of these bootstrap distributions, together with the maximum likelihood estimates and the approximate standard deviations repeated from Table 25.14a.

Before examining whether the two measurement methods differ in their fat content determinations, we need to consider whether measurement method-laboratory interactions are present. The large-sample test statistic (14.52) for testing H_0: $\sigma^2_{\alpha\beta} = 0$ is, using the results in Table 25.14a, $z^* = .086/.064 = 1.34$. This small value of the test statistic supports H_0, that there are no interaction effects. However, the bootstrap distribution of $\hat{\sigma}^2_{\alpha\beta}$ is highly

TABLE 25.14 Maximum Likelihood Estimates and Estimated Variance-Covariance Matrix—Sheffield Foods Company Example.

(a) Estimated Parameters and Standard Deviations

Parameter	Estimated Parameter	Estimated Standard Deviation
$\mu_{..}$	3.694	.158
α_1	.633	.107
σ^2	.023	.006
σ^2_β	.097	.071
$\sigma^2_{\alpha\beta}$	.086	.064

(b) Estimated Approximate Variance-Covariance Matrix

$$
\mathbf{s}^2\{\mathbf{b}\} = \begin{array}{c} \\ \hat{\mu}_{..} \\ \hat{\alpha}_1 \\ \hat{\sigma}^2 \\ \hat{\sigma}^2_\beta \\ \hat{\sigma}^2_{\alpha\beta} \end{array}
\begin{array}{c} \begin{array}{ccccc} \hat{\mu}_{..} & \hat{\alpha}_1 & \hat{\sigma}^2 & \hat{\sigma}^2_\beta & \hat{\sigma}^2_{\alpha\beta} \end{array} \\ \begin{bmatrix} .0250 & .0002 & .0000 & .0000 & .0000 \\ .0002 & .0114 & .0000 & .0000 & .0000 \\ .0000 & .0000 & .0000 & .0000 & -.0000 \\ .0000 & .0000 & .0000 & .0050 & -.0001 \\ .0000 & .0000 & -.0000 & -.0001 & .0041 \end{bmatrix} \end{array}
$$

TABLE 25.15 Means and Standard Deviations of Bootstrap Distributions and Maximum Likelihood Estimates—Sheffield Foods Company Example.

Parameter	Bootstrap Mean	Maximum Likelihood Estimate	Standard Deviation	
			Bootstrap	Maximum Likelihood
$\mu_{..}$	3.69	3.69	.157	.158
α_1	.637	.633	.110	.107
σ^2_β	.092	.097	.128	.071
$\sigma^2_{\alpha\beta}$	.078	.086	.190	.064
σ^2	.023	.023	.006	.006

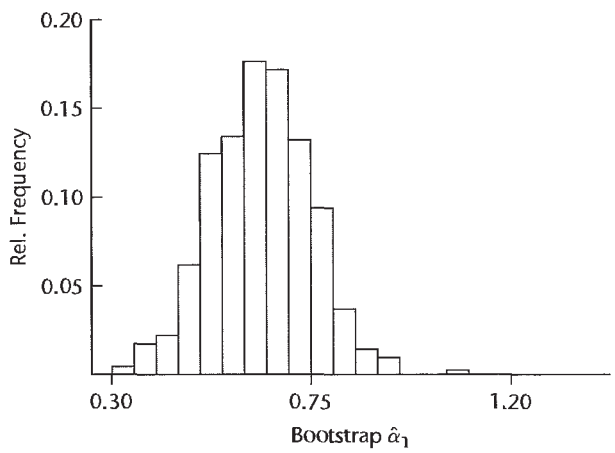


FIGURE 25.5 Bootstrap Distribution for $\hat{\alpha}_1$—Sheffield Foods Company Example.

skewed, with a large concentration at zero. Furthermore, the bootstrap standard deviation according to Table 25.15, $s^*\{\hat{\sigma}^2_{\alpha\beta}\} = .190$, is much larger than the large-sample estimate. Thus, use of large-sample inference procedures may not be appropriate here. Nevertheless, the bootstrap results are consistent with the large-sample results, suggesting even more strongly that there are no interaction effects between measurement methods and laboratories.

We therefore examine next the measurement method main effects. The bootstrap distribution for $\hat{\alpha}_1$ is shown in Figure 25.5. It is approximately normal. Also, Table 25.15 shows that the bootstrap standard deviation for $\hat{\alpha}_1$ and the large-sample standard deviation are very similar. These findings support the use of large-sample inference procedures for α_1. Hence, we use the large-sample confidence interval in (14.54) to estimate $\alpha_1 - \alpha_2 = 2\alpha_1$. For a 95 percent confidence interval, we require:

$$z(.975) = 1.960 \qquad 2\hat{\alpha}_1 = 2(.633) = 1.266 \qquad 2s\{\hat{\alpha}_1\} = 2(.107) = .214$$

The confidence limits therefore are $1.266 \pm 1.960(.214)$ and the approximate 95 percent confidence interval is:

$$.85 \leq \alpha_1 - \alpha_2 \leq 1.69$$

We conclude, with approximate confidence coefficient .95, that the mean government method fat determination is between .85 and 1.69 percent points higher than that for the Sheffield method. Since the true fat content in the samples was 3 percent, Figure 25.4 indicates that the government method is biased upward and that the Sheffield method is more accurate.

Comment

Mixed effects models are sometimes estimated by means of *restricted maximum likelihood* (REML). Using this approach, the variance-covariance components are estimated via maximum likelihood (ML) averaging over all possible values of the fixed effects. The fixed effects are estimated using generalized least squares given their variance-covariance estimates. Under full maximum likelihood, the variance-covariance parameters and the fixed effects are estimated by maximizing their joint likelihood. The fixed effect estimates using REML generally exhibit less bias than ML estimates whereas both REML and ML variance component estimates are identical. See Reference 25.7 for further details of these estimation methods. ■

Cited References

25.1. Searle, S. R., G. Casella, and C. E. McCulloch. *Variance Components*. New York: John Wiley & Sons, 1992.

25.2. Burdick, R. K., and F. A. Graybill. *Confidence Intervals on Variance Components*. New York: Marcel Dekker, Inc., 1992.

25.3. Satterthwaite, F. E. "An Approximate Distribution of Estimates of Variance Components," *Biometrics Bulletin* 2 (1946), pp. 110–14.

25.4. Gaylor, D. W., and F. N. Hopper. "Estimating the Degrees of Freedom for Linear Combinations of Mean Squares by Satterthwaite's Formula," *Technometrics* 11 (1969), pp. 691–706.

25.5. Ting, N., R. K. Burdick, F. A. Graybill, S. Jeyaratnam, and T. F. C. Lu. "Confidence Intervals on Linear Combinations of Variance Components That Are Unrestricted in Sign," *Journal of Statistical Computation and Simulation* 35 (1990), pp. 135–43.

25.6. Schwarz, C. J. "The Mixed-Model ANOVA: The Truth, the Computer Packages, the Books. Part I: Balanced Data." *The American Statistician* 47 (1993), pp. 48–59.

25.7. Hocking, R. R. *Methods and Applications of Linear Models: Regression and the Analysis of Variance*. 2nd ed. New York: John Wiley & Sons, 2003.

25.8. Greenhouse, S. W., and S. Geisser. "On Methods in the Analysis of Profile Data," *Psychometrika* 24 (1959), pp. 95–112.

25.9. Huynh, H., and L. Feldt. "Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Split-Plot Designs," *Journal of Educational Statistics* 1 (1976), pp. 69–82.

25.10. Winer, B. J., D. R. Brown, and K. M. Michels. *Statistical Principles in Experimental Design*. 3rd ed. New York: McGraw-Hill, 1991.

25.11. Burdick, R. K. "Using Confidence Intervals to Test Variance Components." *Journal of Quality Technology* 26 (1994), pp. 30–38.

25.12. Searle, S. R. *Linear Models for Unbalanced Data*. New York: John Wiley & Sons, 1987.

Problems

25.1. A student asks why ε_{ij} is shown as a separate term in random cell means model (25.1) in view of μ_i being a random variable in this model. Respond.

25.2. Refer to Figure 25.1. Here, the situation portrayed is one where the variance σ^2 is larger than the variance σ_μ^2. Is this always the case? Explain.

25.3. In each of the following cases, indicate whether ANOVA model I or model II is more appropriate and state your reasons:

a. In a study of absenteeism at a plant, the treatments are the three 8-hour shifts.

b. In a study of employee productivity, the treatments are 10 production employees selected at random from all production employees in a large company.

c. In a study of anticipated annual income at retirement, the treatments are the four types of retirement plans available to employees.

d. In a study of tire wear in 18-wheel trucks, the treatments are four tire locations selected at random.

25.4. Refer to the Apex Enterprises personnel officers example on page 1036. Explain with reference to this example over what the expectation in (25.2a) is taken. Over what is the variance in (25.2b) taken? Over what is the covariance in (25.2c) taken?

*25.5. Refer to **Filling machines** Problem 16.11. Suppose that the company uses a large number of filling machines and the six machines studied were selected randomly. Assume that ANOVA model (25.1) is applicable.

a. Interpret the following with reference to this example: (1) $\mu_.$, (2) σ_μ^2, (3) σ^2, (4) $\sigma^2\{Y_{ij}\}$.

b. Test whether or not all machines in the population have the same mean fill; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Estimate the mean fill for all machines in the population with a 95 percent confidence interval.

*25.6. Refer to **Filling machines** Problems 16.11 and 25.5.

a. Estimate the proportion of the total variability in carton fills that reflects the differences in mean fills between machines; use a 95 percent confidence interval.

b. Estimate σ^2 with a 95 percent confidence interval. Interpret your interval estimate.

c. Obtain a point estimate of σ_μ^2.

d. Obtain separate approximate 95 percent confidence intervals for σ_μ^2 using the Satterthwaite procedure and the MLS procedure. Are these intervals similar? Comment.

25.7. **Sodium content.** A researcher studied the sodium content in lager beer by selecting at random six brands from the large number of brands of U.S. and Canadian beers sold in a metropolitan area. The researcher then chose eight 12-ounce cans or bottles of each selected brand at random from retail outlets in the area and measured the sodium content (in milligrams) of each can or bottle. The observations follow.

	j							
i	1	2	3	4	5	6	7	8
1	24.4	22.6	23.8	22.0	24.5	22.3	25.0	24.5
2	10.2	12.1	10.3	10.2	9.9	11.2	12.0	9.5
...	...	...	...	...		.	..	...
6	21.3	20.2	20.7	20.8	20.1	18.8	21.1	20.3

Assume that ANOVA model (25.1) is applicable.

a. Test whether or not the mean sodium content is the same in all brands sold in the metropolitan area; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Estimate the mean sodium content for all brands; use a 99 percent confidence interval.

25.8. Refer to **Sodium content** Problem 25.7.

a. Estimate $\sigma_\mu^2/(\sigma_\mu^2 + \sigma^2)$ with a 99 percent confidence interval. Interpret your interval estimate.

b. Obtain point estimates of σ^2 and σ_μ^2.

c. Estimate σ^2 with a 99 percent confidence interval.

d. It has been conjectured that the variance of sodium content between brands is more than twice as great as that within brands. Conduct an appropriate test using $\alpha = .01$. State the alternatives, decision rule, and conclusion.

e. Obtain an approximate 99 percent confidence interval for σ_μ^2 using the MLS procedure. Interpret your confidence interval.

25.9. **Coil winding machines.** A plant contains a large number of coil winding machines. A production analyst studied a certain characteristic of the wound coils produced by these machines by selecting four machines at random and then choosing 10 coils at random from the day's output of each selected machine. The results follow.

	j									
i	1	2	3	4	5	6	7	8	9	10
1	205	204	207	202	208	206	209	205	207	206
2	201	204	198	203	209	207	199	206	205	204
3	198	204	196	201	199	203	202	198	202	197
4	210	209	214	215	211	208	210	209	211	210

Assume that ANOVA model (25.1) is appropriate.

a. Test whether or not the mean coil characteristic is the same for all machines in the plant; use $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Estimate the mean coil characteristic for all coil winding machines in the plant; use a 90 percent confidence interval.

25.10. Refer to **Coil winding machines** Problem 25.9.

a. Estimate $\sigma_\mu^2/(\sigma_\mu^2 + \sigma^2)$ with a 90 percent confidence interval. Interpret your interval estimate.

b. Test whether or not σ_μ^2 and σ^2 are equal; use $\alpha = .10$. State the alternatives, decision rule, and conclusion.

c. Estimate σ^2 with a 90 percent confidence interval. Interpret your interval estimate.

d. Obtain a point estimate of σ_μ^2.

e. Obtain separate approximate 90 percent confidence intervals for σ_μ^2 using the Satterthwaite procedure and the MLS procedure. Are these intervals similar? Comment.

25.11. For mixed effects model (25.42), why is $\sum_i (\alpha\beta)_{ij} = 0$ while usually $\sum_j (\alpha\beta)_{ij} \neq 0$?

25.12. A marketing consultant is designing several experiments involving a newly developed low-cost food processor. The initial experiment has the objectives (1) to compare the effects on unit sales of three possible prices recommended by the sales department ($23.99, $25.49, $25.95) and (2) to determine whether the color scheme used for the appliance affects unit sales. A great many color schemes are feasible; three (white, green, pink) have been selected for the initial experiment to represent the range of possible colors. If the experiment suggests that color scheme does have an effect, this aspect of the product design will be investigated in

detail in a follow-up study. Which ANOVA model would you employ for analyzing the initial experiment? Discuss.

25.13. In a two-factor ANOVA study with $a = 3$, $b = 2$, and $n = 5$, the two factor effects are both random with $\sigma^2 = 5.0$, $\sigma_\alpha^2 = 8.0$, $\sigma_\beta^2 = 10.0$, and $\sigma_{\alpha\beta}^2 = 6.0$. Assume that ANOVA model (25.39) is applicable.

a. Obtain $E\{MSA\}$, $E\{MSB\}$, and $E\{MSAB\}$.

b. What would be the expected mean squares if $\sigma_{\alpha\beta}^2 = 0$, all other parameters remaining the same?

25.14. A survey statistician has commented: "I am rather suspicious of uses of random effects and mixed effects ANOVA models. Seldom are the factor levels chosen by a random mechanism from a known population." Discuss.

25.15. **Miles per gallon.** An automobile manufacturer wished to study the effects of differences between drivers (factor A) and differences between cars (factor B) on gasoline consumption. Four drivers were selected at random; also five cars of the same model with manual transmission were randomly selected from the assembly line. Each driver drove each car twice over a 40-mile test course and the miles per gallon were recorded. The data follow.

Factor A (driver)	Factor B (car) $j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
$i = 1$	25.3	28.9	24.8	28.4	27.1
	25.2	30.0	25.1	27.9	26.6
$i = 2$	33.6	36.7	31.7	35.6	33.7
	32.9	36.5	31.9	35.0	33.9
$i = 3$	27.7	30.7	26.9	29.7	29.2
	28.5	30.4	26.3	30.2	28.9
$i = 4$	29.2	32.4	27.7	31.8	30.3
	29.3	32.4	28.9	30.7	29.9

Assume that random ANOVA model (25.39) is applicable.

a. Test whether or not the two factors interact; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Test separately whether or not factor A and factor B main effects are present. For each test, use $\alpha = .05$ and state the alternatives, decision rule, and conclusion. What is the P-value for each test?

c. Obtain point estimates of σ_α^2 and σ_β^2. Which factor appears to have the greater effect on gasoline consumption?

d. Use the MLS procedure to obtain an approximate 95 percent confidence interval for σ_α^2. Interpret your interval estimate.

e. Use the Satterthwaite procedure to obtain an approximate 95 percent confidence interval for σ_β^2. Is your interval estimate reasonably precise? Comment.

*25.16. Refer to **Disk drive service** Problem 19.16. Suppose that the service center employs a large number of technicians and that the three included in the study were selected at random. Assume that the conditions of mixed ANOVA model (25.42) are applicable, except that here factor A effects are random and factor B effects are fixed. Under current conditions, all technicians service each of the three makes with approximately equal frequency.

a. Test whether or not the two factors interact; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Obtain a point estimate of $\sigma^2_{\alpha\beta}$. Does $\sigma^2_{\alpha\beta}$ appear to be large relative to σ^2? Explain.

c. Test whether or not factor A main effects are present; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. Why is it meaningful here to test for factor A main effects?

d. Test whether or not factor B main effects are present; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. Why is it meaningful here to test for factor B main effects?

e. It is desired to obtain all pairwise comparisons between the means for the three disk drive makes. Use the Tukey procedure and a 95 percent family confidence coefficient to make these comparisons. State your findings.

f. Use the Satterthwaite procedure to obtain an approximate 99 percent confidence interval for $\mu_{\cdot 1}$. Interpret your interval estimate.

g. Obtain an approximate 99 percent confidence interval for σ^2_α using the MLS procedure. Does the variability between technicians appear to be large? Explain.

Imitation pearls. Preliminary research on the production of imitation pearls entailed studying the effect of the number of coats of a special lacquer (factor A) applied to an opalescent plastic bead used as the base of the pearl on the market value of the pearl. Four batches of 12 beads (factor B) were used in the study, and it is desired to also consider their effect on the market value. The three levels of factor A (6, 8, and 10 coats) were fixed in advance, while the four batches can be regarded as a random sample of batches from the bead production process. The market value of each pearl was determined by a panel of experts. The market value data (coded) follow.

Factor A (number of coats)		Factor B (batch) $j = 1$	$j = 2$	$j = 3$	$j = 4$
$i = 1$	6	72.0	72.1	75.2	70.4
		...	...	...	...
		72.8	73.3	77.8	72.4
$i = 2$	8	76.9	80.3	80.2	74.3
		...	...	...	...
		74.2	77.2	79.9	72.9
$i = 3$	10	76.3	80.9	79.2	71.6
		...	...	...	...
		75.0	80.2	81.2	74.4

Assume that mixed ANOVA model (25.42) is applicable.

a. Test for interaction effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Test for factor A and factor B main effects. For each test, use $\alpha = .05$ and state the alternatives, decision rule, and conclusion. What is the P-value for each test?

c. Estimate $D_1 = \mu_{2\cdot} - \mu_{1\cdot}$ and $D_2 = \mu_{3\cdot} - \mu_{2\cdot}$ by means of the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

d. Use the Satterthwaite procedure to obtain an approximate 95 percent confidence interval for $\mu_{2\cdot}$. Interpret your confidence interval.

e. Use the MLS procedure to obtain an approximate 90 percent confidence interval for σ^2_β. Does σ^2_β appear to be large compared to σ^2?

25.18. Refer to **Coin-operated terminals** Problem 20.2. Suppose that the weeks (factor B) had been selected intentionally but the locations (factor A) had been selected at random from a large number of possible locations. Assume that the conditions for additive random block effects in ANOVA model (25.67) are appropriate, except that here factor A effects (blocks) are random and factor B effects are fixed.

a. Test for factor B main effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Why can you not test for factor A main effects here?

*25.19 **Road paint wear.** A state highway department studied the wear characteristics of five different paints at eight locations in the state. The standard, currently used paint (paint 1) and four experimental paints (paints 2, 3, 4, 5) were included in the study. The eight locations were randomly selected, thus reflecting variations in traffic densities throughout the state. At each location, a random ordering of the paints to the chosen road surface was employed. After a suitable period of exposure to weather and traffic, a combined measure of wear, considering both durability and visibility, was obtained. The data on wear follow (the higher the score, the better the wearing characteristics).

Location	Paint (j)					Location	Paint (j)				
i	1	2	3	4	5	i	1	2	3	4	5
1	11	13	10	18	15	5	14	16	13	22	16
2	20	28	15	30	18	6	25	27	26	33	25
3	8	10	8	16	12	7	43	46	41	55	42
4	30	35	27	41	28	8	13	14	12	20	13

a. Obtain the residuals for additive randomized block model (25.67) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. Summarize your findings about the appropriateness of model (25.67).

b. Plot the responses by location in the format of Figure 21.2 on page 896. What does this plot suggest about the appropriateness of the no-interaction assumption here?

c. Conduct the Tukey test for additivity of location and treatment effects, conditional on the locations selected; use $\alpha = .005$. State the alternatives, decision rule, and conclusion.

*25.20 Refer to **Road paint wear** Problem 25.19. Assume that additive randomized block model (25.67) is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not the mean wear differs for the five paints; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Compare the mean wear of each experimental paint against that of the standard paint; use the most efficient multiple comparison procedure with a 90 percent family confidence coefficient. Summarize your findings.

d. Paints 1, 3, and 5 are white, whereas paints 2 and 4 are yellow. Estimate the difference in the mean wear for the two groups of paints with a 95 percent confidence interval. Interpret your findings.

25.21. **Muscle tissue.** A physiologist studied the effects of three reagents on muscle tissue in dogs. Ten litters of three dogs each were randomly selected and the three reagents were randomly assigned to the three dogs in each litter. The data on the effects of the reagents follow (the

higher the value, the higher the activity level):

Litter i	Reagent (j) 1	2	3	Litter i	Reagent (j) 1	2	3
1	10	15	14	6	7	9	10
2	8	12	13	7	24	30	27
3	21	27	25	8	16	18	20
4	14	17	17	9	23	29	32
5	12	18	16	10	18	22	21

a. Obtain the residuals for additive randomized block model (25.67) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. Summarize your findings.

b. Plot the responses by litter in the format of Figure 21.2 on page 896. What does this plot suggest about the appropriateness of the no-interaction assumption here?

c. Conduct the Tukey test for additivity of litter and reagent effects, conditional on the litters selected; use $\alpha = .025$. State the alternatives, decision rule, and conclusion.

d. Based on parts (b) and (c), would interaction randomized block model (25.74) be more appropriate here? What practical differences exist in using models (25.67) and (25.74)?

25.22. Refer to **Muscle tissue** Problem 25.21. Assume that additive randomized block model (25.67) is applicable.

a. Obtain the analysis of variance table.

b. Test whether or not the mean activity level differs for the three reagents; use significance level $\alpha = .025$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Reagents 2 and 3 were expected to be similar to each other but to differ from reagent 1. Use the most efficient multiple comparison procedure with a 95 percent family confidence coefficient to estimate:

$$L_1 = \mu_{\cdot 2} - \mu_{\cdot 3}$$
$$L_2 = \frac{\mu_{\cdot 2} + \mu_{\cdot 3}}{2} - \mu_{\cdot 1}$$

Summarize your findings.

*25.23. Refer to Table 25.11 on page 1069. All three factors in this study have random effects.

a. Test whether or not $\sigma^2_{\alpha\beta\gamma}$ equals zero; use $\alpha = .025$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Test whether or not AB interactions are present. Use significance level $\alpha = .01$. State the alternatives, decision rule, and conclusion.

c. Test whether machines (factor B) have main effects. Use significance level $\alpha = .01$. State the alternatives, decision rule, and conclusion.

d. Use the Satterthwaite procedure to obtain an approximate 95 percent confidence interval for σ^2_α. Interpret your interval estimate.

25.24. Refer to **Electronics assembly** Problem 24.12. Suppose that the number of feasible sequences in which the components can be attached to the board is very large and that the three sequences studied were selected randomly from the set of operationally feasible sequences. Assume that a normal error ANOVA model is applicable where factors A and C have fixed effects and

factor B has random effects. Some relevant expected mean squares for this model are:

$$E\{MSA\} = \sigma^2 + bcn\frac{\sum \alpha_i^2}{a-1} + cn\sigma_{\alpha\beta}^2 \qquad E\{MSABC\} = \sigma^2 + n\sigma_{\alpha\beta\gamma}^2$$

$$E\{MSB\} = \sigma^2 + acn\sigma_\beta^2 \qquad E\{MSE\} = \sigma^2$$

$$E\{MSAC\} = \sigma^2 + bn\frac{\sum\sum(\alpha\gamma)_{ik}^2}{(a-1)(c-1)} + n\sigma_{\alpha\beta\gamma}^2$$

a. What is the appropriate test statistic for testing for AC interactions? For testing for factor B main effects?

b. Test whether or not AC interactions are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

c. Test whether or not factor B main effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

d. Estimate σ_β^2 using the MLS procedure with a 95 percent confidence coefficient. Interpret your interval estimate.

25.25. Consider mixed ANOVA model (25.79) where factor A has fixed effects and the other two factors have random effects. Find the Satterthwaite test statistic F^{**} for testing for factor A main effects. What is the approximate number of degrees of freedom associated with the denominator of this test statistic?

*25.26. Refer to **Disk drive service** Problems 19.16 and 25.16. Suppose that observations $Y_{114} = 57$, $Y_{221} = 61$, and $Y_{224} = 66$ are missing because the time recording instrument malfunctioned. Assume that the conditions of mixed ANOVA model (25.42) are applicable (except that here factor A effects are random, factor B effects are fixed, and unequal sample sizes exist) and that the observations Y_{ijk} are jointly normally distributed. Use the maximum likelihood approach to answer the following.

a. Obtain maximum likelihood estimates of all unknown parameters. Are any of the estimated variances of the random effects equal to zero? If so, what would this imply about the applicability of the likelihood ratio statistic (14.60)?

b. Revise the model by dropping the main factor A effect and obtain maximum likelihood estimates of the unknown parameters in the revised model. Do these estimates differ from the ones obtained in part (a)?

c. Use the z^* test statistic to test whether or not the two factors interact; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Use the likelihood ratio test statistic (14.60) to test whether or not factor B main effects are present; control the risk of Type I error at $\alpha = .01$. State the alternatives, decision rule, and conclusion.

e. Obtain an approximate 99 percent confidence interval for $\sigma_{\alpha\beta}^2$. Interpret your confidence interval.

25.27. Refer to **Imitation pearls** Problem 25.17. Suppose that observations $Y_{113} = 67.4$ and $Y_{322} = 73.7$ are missing because of flaws in the beads. Assume that the conditions of mixed ANOVA model (25.42) are applicable (except that unequal sample sizes are present here) and that the observations Y_{ijk} are jointly normally distributed. Use the maximum likelihood approach to answer the following.

a. Obtain maximum likelihood estimates of all unknown parameters. Are any of the estimated variances of the random effects equal to zero? If so, what would this imply about the applicability of the likelihood ratio statistic (14.60)?

b. Revise the model by dropping the interaction term and obtain maximum likelihood estimates of the unknown parameters in the revised model. Do these estimates differ from the ones obtained in part (a)?

c. Use the likelihood ratio test statistic (14.60) to test for factor B main effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Use the likelihood ratio test statistic (14.60) to test whether factor A main effects are present; control the risk of Type I error at $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Obtain an approximate 95 percent confidence interval for σ_β^2. Interpret your interval estimate.

Exercises

25.28. Show that n' defined in (25.10a) equals n when $n_i \equiv n$.

25.29. What are the values r and n that minimize $\sigma^2\{\bar{Y}_{\cdot}\}$ in (25.12) for a given total sample size n_T?

25.30. Derive the confidence limits in (25.19) from those in (25.18).

25.31. For random ANOVA model (25.39), derive $\sigma^2\{\bar{Y}_{i\cdot\cdot}\}$.

25.32. Consider randomized block model (21.1), but with random treatment effects. Derive $\sigma^2\{Y_{ij}\}$ and $\sigma^2\{\bar{Y}_j\}$.

25.33. Refer to **Dental pain** Problem 21.9. Suppose that the subjects in the study had been randomly selected from eight towns (blocks), and that the towns were randomly selected from a population of towns. Assume that additive randomized block model (25.67) is applicable, except that the factorial structure of the fixed treatment effects needs to be recognized.

a. State the randomized block model for this case.

b. What is the appropriate test statistic for testing whether or not the two factors interact? What are the appropriate test statistics for testing for main effects? [*Hint*: Consider the test for treatment effects in model (25.67).]

25.34. Derive (25.68c).

25.35. For random ANOVA model (25.77), find the variance of the estimated mean $\bar{Y}_{i\cdots}$.

Projects

25.36. Consider a two-factor study with $a = 3$, $b = 2$, and $n = 5$. Random ANOVA model (25.39) is applicable with $\mu_{\cdot\cdot} = 92$, $\sigma_\alpha^2 = 24$, $\sigma_\beta^2 = 11$, $\sigma_{\alpha\beta}^2 = .1$, and $\sigma^2 = 8$.

a. Using a normal random number generator, obtain a value for each of the main effects α_i $(i = 1, 2, 3)$ and β_j $(j = 1, 2)$ and for each interaction effect $(\alpha\beta)_{ij}$.

b. Generate five error terms for each treatment.

c. Combine the parameter values obtained in part (a), the error terms obtained in part (b), and $\mu_{\cdot\cdot} = 92$ to yield five observations Y_{ijk} for each treatment.

d. For the observations obtained in part (c), calculate the F^* test statistic for testing whether or not factor A main effects are present. What is your conclusion using $\alpha = .05$?

e. Repeat the steps in parts (a)–(d) 100 times. Calculate the mean of the 100 numerator mean squares and the mean of the 100 denominator mean squares. Are these means close to theoretical expectations?

f. In what proportion of the 100 trials did the test lead to the conclusion of the presence of factor A main effects? Does the test have good power for the case considered here?

25.37. Refer to **Road paint wear** Problem 25.19.

a. Estimate the variance-covariance matrix of the treatment observations in a block; use (27.8) on page 1135 to obtain the entries in the matrix.

b. Does the compound symmetry property of (25.71) appear to be reasonable here? Explain.

c. Does the sphericity property of (25.73) appear to be reasonable here? Explain.

25.38. Refer to **Muscle tissue** Problem 25.21.

a. Estimate the variance-covariance matrix of the treatment observations in a block; use (27.8) on page 1135 to obtain the entries in the matrix.

b. Does the compound symmetry property of (25.71) appear to be reasonable here? Explain.

c. Does the sphericity property of (25.73) appear to be reasonable here? Explain.

25.39. Refer to **Miles per gallon** Problem 25.15. Suppose that observation $Y_{232} = 31.9$ is missing because the record was lost for this experimental trial. Assume that random ANOVA model (25.39) is applicable (except that the sample sizes are unequal here) and that the oberservations Y_{ijk} are jointly normally distributed.

a. Use the method of maximum likelihood to estimate $\mu_{..}$ and the variance components σ_α^2, σ_β^2, $\sigma_{\alpha\beta}^2$, and σ^2. Which variance component appears to be largest? Also obtain the estimated standard deviation for each of the estimated variance components.

b. Obtain a bootstrap sample by using a normal random number generator to provide normal values with means zero and variances equal to the estimates of the variance components in part (a) for (1) the α_i $(i = 1, \ldots, 4)$, (2) the β_j $(j = 1, \ldots, 5)$, (3) the $(\alpha\beta)_{ij}$, and (4) the n_{ij} error terms ε_{ijk} for each treatment. Combine these with $\hat{\mu}_{..}$ obtained in part (a) to create the n_{ij} bootstrap outcomes Y_{ijk} for each treatment.

c. Use the method of maximum likelihood to estimate σ_α^2, σ_β^2, and $\sigma_{\alpha\beta}^2$ for the bootstrap sample obtained in part (b).

d. Repeat parts (b) and (c) 250 times.

e. Obtain histograms of the bootstrap distributions for the 250 bootstrap estimates of σ_α^2, σ_β^2, and $\sigma_{\alpha\beta}^2$. Also obtain the mean and standard deviation for each of the bootstrap distributions. Based on these results and the results in part (a), does it appear that large-sample inference procedures are appropriate here? Explain.

Part VI

pecialized
tudy Designs

Chapter

Nested Designs, Subsampling, and Partially Nested Designs

In this chapter, we take up the basic elements of nested designs, including the use of subsampling. We begin by considering the general concept of nested designs and describe how these designs differ from crossed designs. We then take up in detail two-factor nested designs and their analysis. We conclude by considering subsampling designs and partially nested designs.

26.1 Distinction between Nested and Crossed Factors

In the factorial studies considered so far, where every level of one factor appears with each level of every other factor, the factors are said to be crossed. A different situation occurs when factors are nested. The distinction between nested and crossed factors will now be illustrated by some examples involving two-factor studies.

Example 1

A large manufacturing company operates three regional training schools for mechanics, one in each of its operating districts. The schools have two instructors each, who teach classes of about 15 mechanics in three-week sessions. The company was concerned about the effect of school (factor A) and instructor (factor B) on the learning achieved. To investigate these effects, classes in each district were formed in the usual way and then randomly assigned to one of the two instructors in the school. This was done for two sessions, and at the end of each session a suitable summary measure of learning for the class was obtained. The results are presented in Table 26.1.

The layout of Table 26.1 appears identical to an ordinary two-factor investigation, with two observations per cell (see, e.g., Table 19.7). In fact, however, the study is not an ordinary two-factor study. The reason is that the instructors in the Atlanta school did not also teach in the other two schools, and similarly for the other instructors. Thus, six different instructors were involved. An ordinary two-factor investigation with six different instructors would have consisted of 18 treatments, as shown in Figure 26.1a. In the training school example, however, only six treatments were included, as shown in Figure 26.1b, where

TABLE 26.1 Sample Data for Nested Two-Factor Study—Training School Example (class learning scores, coded).

Factor *A* (school) *i*		Factor *B* (instructor) *j* = 1	Factor *B* (instructor) *j* = 2	Average
Atlanta		25 29	14 11	
	Average	$\bar{Y}_{11\cdot} = 27$	$\bar{Y}_{12\cdot} = 12.5$	$\bar{Y}_{1\cdot\cdot} = 19.75$
Chicago		11 6	22 18	
	Average	$\bar{Y}_{21\cdot} = 8.5$	$\bar{Y}_{22\cdot} = 20$	$\bar{Y}_{2\cdot\cdot} = 14.25$
San Francisco		17 20	5 2	
	Average	$\bar{Y}_{31\cdot} = 18.5$	$\bar{Y}_{32\cdot} = 3.5$	$\bar{Y}_{3\cdot\cdot} = 11.00$
			Average	$\bar{Y}_{\cdot\cdot\cdot} = 15$

FIGURE 26.1 Illustration of Crossed and Nested Factors—Training School Example.

(a) Crossed Factors

School (factor A)	Instructor (factor B)					
	1	2	3	4	5	6
Atlanta						
Chicago						
San Francisco						

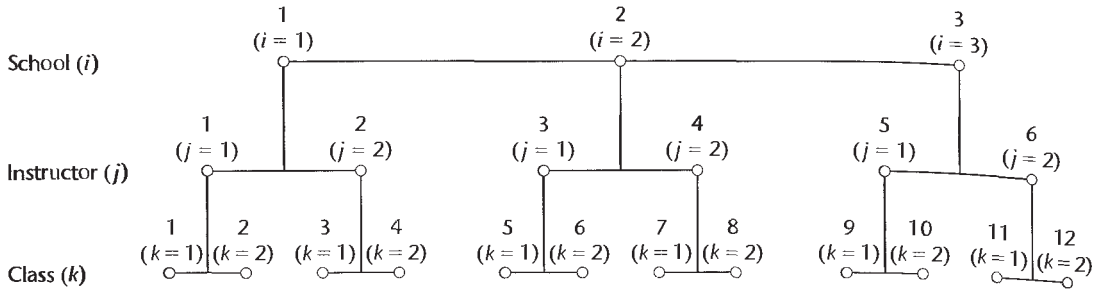
(b) Nested Factors

School (factor A)	Instructor (factor B)					
	1	2	3	4	5	6
Atlanta			X	X	X	X
Chicago	X	X			X	X
San Francisco	X	X	X	X		

the crossed-out cells represent treatments not studied. Figure 26.2 contains an alternative graphic representation of the nested design for the training school example, including the two replications of the study.

It is clear from Figure 26.1b that the experimental design for the training school example involves an incomplete factorial arrangement of a special type, where each level of factor *B* (instructor) occurs with only one level of factor *A* (school). Specifically here, each instructor

FIGURE 26.2 Graphic Representation of Two-Factor Nested Design—Training School Example.



teaches in only one school. Factor B is therefore said to be *nested* within factor A. As noted earlier, in an ordinary factorial study where every factor level of A appears with every factor level of B, factors A and B are said to be *crossed*.

There is another way to look at the distinction between nested and crossed designs. Let μ_{ij} denote the mean response when factor A is at the ith level and factor B is at the jth level. If the factors are crossed, the jth level of B is the same for all levels of A. If, on the other hand, factor B is nested within factor A, the jth level of B when A is at level 1 has nothing in common with the jth level of B when A is at level 2, and so on. For instance, in a crossed factorial study of the effects of price ($1.99, $2.49) and advertising level (high, low), a particular advertising level is the same no matter with which price it appears, and similarly for the price levels. On the other hand, in the nested design for the training school example, the first instructor in school 1 is not the same as the first instructor in school 2, and so on.

Example 2

An analyst was interested in the effects of community (factor A) and neighborhood (factor B) on the spread of information about new products. Information was obtained from samples of families in various neighborhoods within selected communities. Since the neighborhood designated 1 in a given community is not the same as the neighborhoods designated 1 in the other communities, and similarly for the other neighborhoods, neighborhoods here are nested within communities.

Comments

1. The distinction between crossed and nested factors is often a fine one. In Example 2, if the neighborhoods of each community represented specified average income levels so that, say, the first neighborhoods in each community had an average income of $5,000–$9,999, the second neighborhoods an average income of $10,000–$19,999, and so on for the other neighborhoods, one could view the design as a crossed one. The factors would be community and economic level of neighborhood, and these would be crossed since a given economic level is the same for all communities, and vice versa.

2. Nested factors are frequently encountered in observational studies where the researcher cannot manipulate the factors under study, or in experiments where only some factors can be manipulated. Factors that cannot be manipulated, it will be recalled, are designated observational factors, in distinction to experimental factors that can be assigned at will to the experimental units. Example 2 is an observational study where both community and neighborhood are observational factors since families (the study units) were not randomly assigned to either community or neighborhood. In Example 1, school is an observational factor because the classes of a school (the experimental units) are made

up of mechanics from the district in which the school is located. Instructors in this example are an experimental factor since they are assigned randomly to a class, but a nested design results because the randomization of instructors is restricted to within a school. ■

6.2 Two-Factor Nested Designs

We now consider nested designs involving two factors, one of which is nested inside the other. For consistency, we always consider the case where factor B is nested within factor A. We initially assume that both factor effects are fixed, but later we also consider the case of random effects. We assume throughout that *all treatment means are of equal importance.*

Development of Model Elements

We shall use the customary notation for a two-factor study, and let μ_{ij} denote the mean response when factor A is at the ith level ($i = 1, \ldots, a$) and factor B is at the jth level ($j = 1, \ldots, b$). As usual, when all mean responses are of equal importance we define:

$$\mu_{i\cdot} = \frac{\sum_j \mu_{ij}}{b} \tag{26.1}$$

For the training school example of Table 26.1, $\mu_{1\cdot}$ represents the mean learning score for the Atlanta school, averaged over the instructors of that school, and $\mu_{2\cdot}$ and $\mu_{3\cdot}$ are interpreted similarly. Note once more that the $\mu_{i\cdot}$ here represent mean learning scores that have been averaged over *different* instructors.

We define the main effect of the ith level of factor A as usual:

$$\alpha_i = \mu_{i\cdot} - \mu_{\cdot\cdot} \tag{26.2}$$

where:

$$\mu_{\cdot\cdot} = \frac{\sum_i \sum_j \mu_{ij}}{ab} = \frac{\sum_i \mu_{i\cdot}}{a} \tag{26.2a}$$

is the overall mean response. It follows from (26.2a) that:

$$\sum_i \alpha_i = 0 \tag{26.3}$$

In a nested design, it is not meaningful to employ a model component for the main effect of the jth level of factor B. To see why, consider again the training school example. Since each school employs different instructors and the jth instructors in the various schools are not the same, it would be meaningless to consider the effect of the jth instructor, averaged over all schools. Instead, the individual effects of each instructor in each school need to be considered. We denote these individual effects by $\beta_{j(i)}$, where the subscript $j(i)$ indicates that the jth factor level of B is nested within the ith factor level of A. $\beta_{j(i)}$ is defined as follows:

$$\beta_{j(i)} = \mu_{ij} - \mu_{i\cdot} \tag{26.4}$$

which can be rewritten, utilizing (26.2):

$$\beta_{j(i)} = \mu_{ij} - \alpha_i - \mu_{\cdot\cdot} \tag{26.4a}$$

It follows from (26.4) and (26.1) that:

$$\sum_j \beta_{j(i)} = 0 \qquad i = 1, \ldots, a \tag{26.5}$$

The meaning of $\beta_{j(i)}$ can be seen most clearly from (26.4). With reference to the training school example, $\beta_{j(i)}$ is simply the difference in the mean learning score for the jth instructor of school i and the average of the mean learning scores for all instructors in that school. Thus, the effect of the jth instructor in the ith school is measured with respect to the overall mean learning score for the school in which the instructor teaches. We shall call $\beta_{j(i)}$ the *specific effect* of the jth level of factor B nested within the ith level of factor A.

We have now expressed the mean response μ_{ij} in terms of the overall mean, the main effect of the ith level of factor A, and the specific effect of the jth level of factor B nested within the ith level of factor A, as can be seen from (26.4a):

$$\mu_{ij} \equiv \mu_{..} + \alpha_i + \beta_{j(i)} \equiv \mu_{..} + (\mu_{i.} - \mu_{..}) + (\mu_{ij} - \mu_{i.}) \tag{26.6}$$

For the training school example, the mean learning score for the jth instructor in school i has been expressed in terms of the overall mean, the main effect of school i, and the specific effect of instructor j within school i.

To complete the model, we need only add a random error term ε_{ijk}.

Nested Design Model

Let Y_{ijk} denote the response for the kth trial when factor A is at the ith level and factor B is at the jth level. We assume that there are n replications for each factor level combination, i.e., $k = 1, \ldots, n$, and that $i = 1, \ldots, a$ and $j = 1, \ldots, b$. Such a study is said to be *balanced* because the same number of factor B levels is nested within each factor A level and the number of replications is the same throughout.

When both factors A and B have fixed effects, an appropriate nested design model is:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk} \tag{26.7}$$

where:

$\mu_{..}$ is a constant
α_i are constants subject to the restriction $\sum \alpha_i = 0$
$\beta_{j(i)}$ are constants subject to the restrictions $\sum_j \beta_{j(i)} = 0$ for all i
ε_{ijk} are independent $N(0, \sigma^2)$
$i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, n$

The expected value and variance of observation Y_{ijk} for nested design model (26.7) with fixed factor effects are:

$$E\{Y_{ijk}\} = \mu_{..} + \alpha_i + \beta_{j(i)} \tag{26.8a}$$

$$\sigma^2\{Y_{ijk}\} = \sigma^2 \tag{26.8b}$$

Thus, all observations have a constant variance. Further, the observations Y_{ijk} are independent and normally distributed for this model.

Comments

1. It is not necessary, as in model (26.7), that the study be balanced, that is, that the number of replications be equal for all factor combinations and that the number of levels of nested factor B (number of instructors in the training school example) be the same for each level of factor A (school in this example). We shall discuss the removal of some of these restrictions in Section 26.6. We only point out now that the computations become more complex when the study is unbalanced.

2. There is no interaction term in nested design model (26.7). There is no need for it since factor B is nested within factor A, not crossed with it. To put this somewhat differently, with reference to the training school example, it is not possible to estimate a school-instructor interaction when each instructor teaches in only one school. The teacher effect $\beta_{j(i)}$, since it is specific to a given school i, in a sense incorporates the interaction effect between the particular teacher j (in the ith school) and the ith school, but it is not possible in a nested design to disentangle this interaction effect.

3. The factor level means $\mu_{i\cdot}$ in a nested design are not generally the same as the corresponding means in a crossed design. Remember that in a nested design, the $\mu_{i\cdot}$ are obtained by averaging over only some of the distinctive levels of factor B. With reference to the training school example, the $\mu_{i\cdot}$ are obtained by averaging over only those teachers who instruct in the ith school. In a crossed design, on the other hand, the $\mu_{i\cdot}$ would be obtained by averaging over all instructors included in the study. ■

Random Factor Effects

If both factors A and B have random factor levels, nested design model (26.7) is modified with α_i, $\beta_{j(i)}$, and ε_{ijk} being independent normal random variables with expectations 0 and variances σ_α^2, σ_β^2, and σ^2, respectively. Thus, it is assumed that all $\beta_{j(i)}$ have the same variance σ_β^2. The assumption that all $\beta_{j(i)}$ have the same variance also is made if only factor B is random. It is important to check whether this assumption is appropriate, since it may well be that the mean responses $\mu_{i1}, \mu_{i2}, \ldots$, in one factor A level (plant, school, city, etc.) differ in variability from those in other factor A levels (other plants, schools, cities, etc.). Tests for equality of variances are discussed in Section 18.2.

26.3 Analysis of Variance for Two-Factor Nested Designs

Fitting of Model

The least squares and maximum likelihood estimators of the parameters in nested design model (26.7) are obtained in the usual fashion. Employing our customary notation for sample data in factorial studies, the estimators are:

Parameter	Estimator	
$\mu_{\cdot\cdot}$	$\hat{\mu}_{\cdot\cdot} = \bar{Y}_{\cdot\cdot}$	**(26.9a)**
α_i	$\hat{\alpha}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}$	**(26.9b)**
$\beta_{j(i)}$	$\hat{\beta}_{j(i)} = \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot}$	**(26.9c)**

The fitted values therefore are:

$$\hat{Y}_{ijk} = \bar{Y}_{\cdot\cdot} + (\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}) + (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot}) = \bar{Y}_{ij\cdot} \tag{26.10}$$

and the residuals are:

$$e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - \bar{Y}_{ij\cdot} \tag{26.11}$$

Sums of Squares

The analysis of variance for nested design model (26.7) is obtained by decomposing the total deviation $Y_{ijk} - \bar{Y}_{...}$ as follows:

$$\underbrace{Y_{ijk} - \bar{Y}_{...}}_{\text{Total deviation}} = \underbrace{\bar{Y}_{i..} - \bar{Y}_{...}}_{A \text{ main effect}} + \underbrace{\bar{Y}_{ij.} - \bar{Y}_{i..}}_{\substack{\text{Specific } B \\ \text{effect when } A \\ \text{at } i\text{th level}}} + \underbrace{Y_{ijk} - \bar{Y}_{ij.}}_{\text{Residual}} \tag{26.12}$$

When we square (26.12) and sum over all cases, all cross-product terms drop out and we obtain:

$$SSTO = SSA + SSB(A) + SSE \tag{26.13}$$

where:

$$SSTO = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2 \tag{26.13a}$$

$$SSA = bn \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2 \tag{26.13b}$$

$$SSB(A) = n \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..})^2 \tag{26.13c}$$

$$SSE = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2 = \sum_i \sum_j \sum_k e_{ijk}^2 \tag{26.13d}$$

SSTO is the usual total sum of squares, and *SSA* is the ordinary factor A sum of squares, reflecting the variability of the estimated factor level means $\bar{Y}_{i..}$.

SSB(A) is the factor B sum of squares, with the notation reflecting that factor B is nested within factor A. *SSB*(A) is made up of terms such as:

$$n \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..})^2 \tag{26.14}$$

The term in (26.14) is simply the ordinary factor B sum of squares when factor A is at level i. These terms are then summed over all levels of factor A.

Finally, the error sum of squares *SSE* is, as usual, the sum of the squared residuals and reflects the variability of each observation Y_{ijk} around the corresponding estimated treatment mean $\bar{Y}_{ij.}$. Alternatively, we can view *SSE* as being made up of terms such as:

$$\sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2 \tag{26.15}$$

The term in (26.15) is simply the ordinary error sum of squares within the ith level of factor A. These terms are then summed over all levels of factor A.

Thus, a nested two-factor design can be viewed as a series of single-factor investigations at the successive levels of the other factor. In terms of the training school example, a study of the effects of instructors (B) within any given school (A_i) leads to the usual sums of squares for instructors and errors in a single-factor analysis of variance within school A_i,

ABLE 26.2 Relation between Nested Two-Factor ANOVA and Single-Factor ANOVAs—Training School xample.

e of ation	Single-Factor ANOVAs: School 1 SS	School 1 df	School 2 SS	School 2 df	School 3 SS	School 3 df	Nested Two-Factor ANOVA SS	Nested Two-Factor ANOVA df
een instructors within schools)	$SSB(A_1)$	$2-1$	$SSB(A_2)$	$2-1$	$SSB(A_3)$	$2-1$	$SSB(A)$	$3(2-1)$
	$SSE(A_1)$	$2(2-1)$	$SSE(A_2)$	$2(2-1)$	$SSE(A_3)$	$2(2-1)$	SSE	$3(2)(2-1)$
e within schools	$SSTO(A_1)$	$2(2)-1$	$SSTO(A_2)$	$2(2)-1$	$SSTO(A_3)$	$2(2)-1$		
e veen schools							SSA	$3-1$
otal							$SSTO$	$3(2)(2)-1$

denoted by $SSB(A_i)$ and $SSE(A_i)$:

$$SSB(A_i) = n\sum_j(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot})^2 \qquad SSE(A_i) = \sum_j\sum_k(Y_{ijk} - \bar{Y}_{ij\cdot})^2$$

These are then aggregated to yield $SSB(A)$ and SSE, respectively. It is only the between-schools sum of squares SSA that introduces explicitly the other factor. Table 26.2 demonstrates this relation between the single-factor analyses of variance for each school and the two-factor analysis of variance for the nested design.

Degrees of Freedom

The degrees of freedom associated with the various sums of squares can be deduced directly from the known relationships already studied. Since there is a total of abn cases, the degrees of freedom associated with $SSTO$ are $abn - 1$. For any level of factor A, there are $b(n-1)$ degrees of freedom associated with the error sum of squares. Aggregating over all levels of factor A, there are $ab(n-1)$ degrees of freedom associated with SSE. Similarly, for any level of factor A, there are $b-1$ degrees of freedom associated with the factor B sum of squares. Hence, by aggregating over all levels of factor A, we find that there are $a(b-1)$ degrees of freedom associated with $SSB(A)$. Finally, since there are a levels of factor A, there are $a-1$ degrees of freedom associated with SSA.

Table 26.2 shows this aggregation of the degrees of freedom for the training school example, and Table 26.3 presents the general analysis of variance table for two-factor nested design model (26.7) where factor B is nested within factor A.

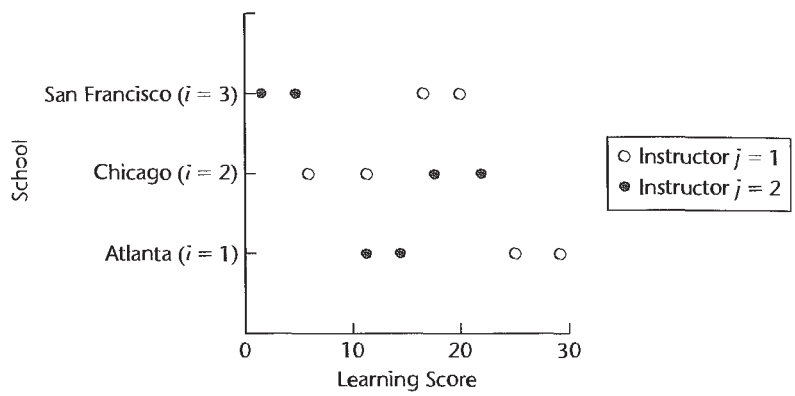
Example

In the training school example of Table 26.1, both schools and instructors were regarded as fixed factors; hence, model (26.7) was deemed appropriate. Figure 26.3 presents aligned dot plots of the class learning scores Y_{ijk} for each school. Note that different symbols are used for the two instructors within each school. Figure 26.3 suggests strongly that differences between instructors within a school are present and that there may be differences in the mean learning for the three schools. Note also from the dot plots that the variability of the class learning scores for the two classes taught by each of the six instructors appears to be reasonably constant, as required by model (26.7).

TABLE 26.3 ANOVA Table for Nested Balanced Two-Factor Fixed Effects Model (26.7) (B nested within A).

Source of Variation	SS	df	MS	$E\{MS\}$
Factor A	$SSA = bn\sum(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$a-1$	MSA	$\sigma^2 + bn\frac{\sum\alpha_i^2}{a-1}$
Factor B (within A)	$SSB(A) = n\sum\sum(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot})^2$	$a(b-1)$	$MSB(A)$	$\sigma^2 + n\frac{\sum\sum\beta_{j(i)}^2}{a(b-1)}$
Error	$SSE = \sum\sum\sum(Y_{ijk} - \bar{Y}_{ij\cdot})^2$	$ab(n-1)$	MSE	σ^2
Total	$SSTO = \sum\sum\sum(Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$	$abn-1$		

FIGURE 26.3 Dot Plots of Class Learning Scores—Training School Example.



To analyze the instructor and school effects formally, we begin by obtaining the analysis of variance. The sums of squares were obtained as follows using formulas (26.13):

$$SSTO = (25-15)^2 + (29-15)^2 + \cdots + (2-15)^2 = 766$$

$$SSA = 2(2)[(19.75-15)^2 + (14.25-15)^2 + (11.00-15)^2] = 156.5$$

$$SSB(A) = 2[(27-19.75)^2 + (12.5-19.75)^2 + \cdots + (3.5-11.00)^2] = 567.5$$

$$SSE = (25-27)^2 + (29-27)^2 + \cdots + (2-3.5)^2 = 42$$

Table 26.4a contains the analysis of variance.

Comment

Most analysis of variance computer packages provide an option for obtaining the ANOVA for nested designs. Should this option be unavailable, the ordinary ANOVA for crossed factors can be used with only slight inconvenience when the nested study is balanced. $SSTO$, SSA, and SSE with the crossed-factor analysis will be the same, and $SSB(A)$ is obtained from the relation:

$$\underbrace{SSB(A)}_{\text{Nested}} = \underbrace{SSB + SSAB}_{\text{Crossed}} \tag{26.16}$$

The same relation holds for the associated degrees of freedom. ■

ABLE 26.4 ANOVA for o-Factor ested esign—aining chool xample.

(a) ANOVA Table

Source of Variation	*SS*	*df*	*MS*
Schools (A)	$SSA = 156.5$	2	78.25
Instructors, within schools [$B(A)$]	$SSB(A) = 567.5$	3	189.17
Error (E)	$SSE = 42.0$	6	7.00
Total	$SSTO = 766.0$	11	

(b) Decomposition of SSB(A)

Source of Variation	$SSB(A_i)$	*df*	$MSB(A_i)$
Instructors, Atlanta	210.25	1	210.25
Instructors, Chicago	132.25	1	132.25
Instructors, San Francisco	225.00	1	225.00
Total	567.5	3	

Tests for Factor Effects

Tests for factor effects in a nested two-factor study are straightforward. The appropriate test statistics are determined, as for a crossed two-factor study, by comparing the expected values of the ANOVA mean squares. The expected mean squares for nested fixed effects model (26.7) are shown in Table 26.3. They can be obtained by somewhat tedious derivations. We do not illustrate these derivations because Appendix D describes a relatively simple method of finding expected mean squares for any balanced nested design. Also, many computer packages provide the expected mean squares for nested models.

The $E\{MS\}$ column in Table 26.3 indicates that for fixed effects model (26.7), the test for factor A main effects:

$$H_0\text{: all } \alpha_i = 0 \qquad H_a\text{: not all } \alpha_i \text{ equal zero} \tag{26.17a}$$

is based on the test statistic:

$$F^* = \frac{MSA}{MSE} \tag{26.17b}$$

and the decision rule to control the level of significance at α is:

$$\text{If } F^* \leq F[1-\alpha; a-1, (n-1)ab], \text{ conclude } H_0 \qquad \text{If } F^* > F[1-\alpha; a-1, (n-1)ab], \text{ conclude } H_a \tag{26.17c}$$

Similarly, to test for factor B specific effects:

$$H_0\text{: all } \beta_{j(i)} = 0 \qquad H_a\text{: not all } \beta_{j(i)} \text{ equal zero} \tag{26.18a}$$

the appropriate test statistic is:

$$F^* = \frac{MSB(A)}{MSE} \tag{26.18b}$$

and the appropriate decision rule is:

$$\text{If } F^* \leq F[1-\alpha; a(b-1), (n-1)ab], \text{ conclude } H_0$$
$$\text{If } F^* > F[1-\alpha; a(b-1), (n-1)ab], \text{ conclude } H_a \qquad \textbf{(26.18c)}$$

Example

For the analysis of variance in Table 26.4a for the training school example, we conduct the first test to determine whether or not main school effects exist. The alternatives are given in (26.17a), and test statistic (26.17b) here is:

$$F^* = \frac{78.25}{7.00} = 11.2$$

For level of significance $\alpha = .05$, we require $F(.95; 2, 6) = 5.14$. Since $F^* = 11.2 > 5.14$, we conclude that the three schools differ in mean learning effects. The P-value of the test is .0094.

Next is a test for differences in mean learning effects between instructors within each school. The alternatives are given in (26.18a), and test statistic (26.18b) here is:

$$F^* = \frac{189.17}{7.00} = 27.0$$

For $\alpha = .05$, we require $F(.95; 3, 6) = 4.76$. Since $F^* = 27.0 > 4.76$, we conclude that instructors within at least one school differ in terms of mean learning effects. The P-value of this test is .0007.

Comments

1. The alternative H_0 in (26.18a) can also be expressed in terms of the treatment means μ_{ij}:

$$H_0\colon \mu_{11} = \mu_{12} = \cdots = \mu_{1b}; \mu_{21} = \mu_{22} = \cdots = \mu_{2b}; \ldots \qquad \textbf{(26.19)}$$

In terms of the training school example, H_0 states that the mean learning scores for all instructors in Atlanta are the same, and similarly for the other schools. It does *not* state that the mean learning scores for all instructors in the different schools are the same.

2. If it is concluded that factor B effects are present, it is often desired to ascertain whether they are present in all levels of factor A or only in some. (In some cases, indeed, one may wish to proceed immediately to this analysis.) With reference to the training school example, the question would be whether the instructor effects differ in all schools or only in some schools. As noted earlier, $SSB(A)$ in Table 26.4a is made up of the instructor sums of squares within the individual schools. These component sums of squares can be used for testing instructor effects within each school. Table 26.4b contains the relevant component sums of squares. To test for instructor differences within the Atlanta school, for instance, we use test statistic $F^* = MSB(A_1)/MSE = 210.25/7.00 = 30.0$. For level of significance $\alpha = .05$, we need $F(.95; 1, 6) = 5.99$. Since $F^* = 30.0 > 5.99$, we conclude that the two instructors in Atlanta have different mean learning effects. Using the same level of significance each time, similar conclusions are reached for the other two schools. The family level of significance for the three tests according to the Bonferroni inequality is at most .15.

3. If the assumption of constant error variance were violated in the training school example through unequal variances for the different schools, it would still be possible to study instructor effects within each school by separate analyses of variance for each school.

4. The power of the tests for fixed factor A and factor B effects can be ascertained by using (24.49) together with the expected mean squares in Table 26.3. ■

TABLE 26.5 Expected Mean Squares for Nested Balanced Two-Factor Designs with Random Factor Effects (*B* nested within *A*).

	Expected Mean Square	
Mean Square	***A* Fixed, *B* Random**	***A* Random, *B* Random**
MSA	$\sigma^2 + bn\frac{\sum \alpha_i^2}{a-1} + n\sigma_\beta^2$	$\sigma^2 + bn\sigma_\alpha^2 + n\sigma_\beta^2$
MSB(A)	$\sigma^2 + n\sigma_\beta^2$	$\sigma^2 + n\sigma_\beta^2$
MSE	σ^2	σ^2
	Appropriate Test Statistic	
Test for	***A* Fixed, *B* Random**	***A* Random, *B* Random**
Factor *A*	*MSA/MSB(A)*	*MSA/MSB(A)*
Factor *B(A)*	*MSB(A)/MSE*	*MSB(A)/MSE*

Random Factor Effects

Test statistic (26.17b) for factor A main effects is not appropriate if either or both factor effects are random. Table 26.5 gives the expected mean squares for these cases and also the appropriate test statistics.

26.4 Evaluation of Appropriateness of Nested Design Model

The diagnostic procedures described earlier are entirely applicable for examining whether nested design model (26.7) is appropriate. The residuals in (26.11):

$$e_{ijk} = Y_{ijk} - \bar{Y}_{ij}. \tag{26.20}$$

may be examined as usual for normality, constancy of the error variance, and independence of the error terms. In particular, aligned dot plots of the residuals for each factor A level may be helpful in examining whether the variance of the error terms is constant for the different factor A levels within which factor B is nested.

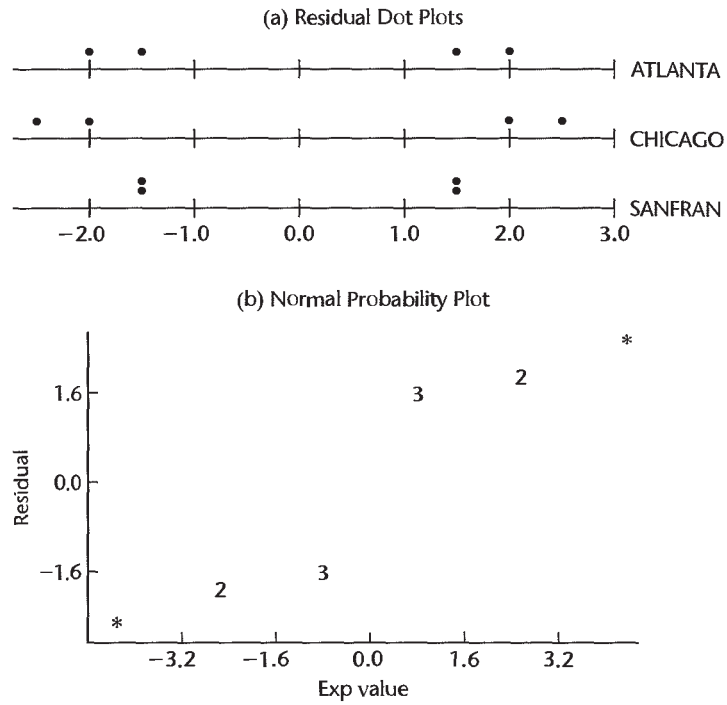
Example

Figure 26.4a contains MINITAB aligned dot plots of the residuals for each school for the training school example. These plots are affected by the rounded nature of the data, but they support the appropriateness of the assumption of constancy of the error variance. Figure 26.4b presents a normal probability plot of the residuals. This plot is also affected by the rounded nature of the observations, but does not indicate any gross departure from normality. This conclusion is supported by the coefficient of correlation between the ordered residuals and their expected values under normality, which is .927. These and other diagnostics (not shown here) support the appropriateness of nested design model (26.7) for the training school example.

Comment

Since there are numerous ties among the residuals in the training school example, the normal probability plot in Figure 26.4b is obtained by plotting each of the tied residuals against the expected value for the mean of the tied order positions and showing the number of tied residuals at that position. ■

FIGURE 26.4 MINITAB Diagnostic Residual Plots—Training School Example.



26.5 Analysis of Factor Effects in Two-Factor Nested Designs

When factor effects are present in a nested design, estimates and/or comparisons of these effects are usually desired.

Estimation of Factor Level Means $\mu_{i.}$

When factor A (fixed effects factor) has significant main effects, there is frequent interest in estimating the factor level means $\mu_{i.}$. The estimated factor level mean $\bar{Y}_{i..}$ is an unbiased estimator of $\mu_{i.}$. As usual for a fixed effects factor, the estimated variance of $\bar{Y}_{i..}$ is based on the mean square in the denominator of the statistic used for testing for factor A main effects, and on the number of cases on which $\bar{Y}_{i..}$ is based. Confidence limits for $\mu_{i.}$ are of the customary form:

$$\bar{Y}_{i..} \pm t(1-\alpha/2; df)s\{\bar{Y}_{i..}\} \tag{26.21}$$

where:

$$s^2\{\bar{Y}_{i..}\} = \frac{MSE}{bn} \qquad df = ab(n-1) \quad A \text{ and } B \text{ fixed} \tag{26.21a}$$

$$s^2\{\bar{Y}_{i..}\} = \frac{MSB(A)}{bn} \qquad df = a(b-1) \quad A \text{ fixed, } B \text{ random} \tag{26.21b}$$

Confidence limits for contrasts $L = \sum c_i \mu_{i.}$, where $\sum c_i = 0$, are set up in the usual way, utilizing the estimator $\hat{L} = \sum c_i \bar{Y}_{i..}$ and the t distribution with degrees of freedom

those associated with the appropriate mean square:

$$\hat{L} \pm t(1-\alpha/2; df)s\{\hat{L}\} \tag{26.22}$$

where:

$$s^2\{\hat{L}\} = \sum c_i^2 s^2\{\bar{Y}_{i..}\} \qquad \text{as given by (26.21a) or (26.21b)} \tag{26.22a}$$

The Tukey and Bonferroni simultaneous comparison procedures can be utilized in the usual way for making pairwise comparisons with family confidence coefficient $1-\alpha$, and the Scheffé and Bonferroni simultaneous comparison procedures can be employed for a family of contrasts.

Example

For the training school example in Table 26.1, it was desired to estimate the mean learning score for the Atlanta school with a 95 percent confidence coefficient. Using our earlier results in Tables 26.1 and 26.4a, we obtain for the fixed effects model:

$$\bar{Y}_{1..} = 19.75$$

$$s^2\{\bar{Y}_{1..}\} = \frac{MSE}{bn} = \frac{7.00}{4} = 1.75$$

$$s\{\bar{Y}_{1..}\} = 1.32$$

$$t(.975; 6) = 2.447$$

$$16.5 = 19.75 - 2.447(1.32) \leq \mu_{1.} \leq 19.75 + 2.447(1.32) = 23.0$$

In addition, pairwise comparisons of the three schools were to be made with family confidence coefficient .90. We shall utilize the Tukey procedure and require:

$$T = \frac{1}{\sqrt{2}}q[1-\alpha; a, ab(n-1)] = \frac{1}{\sqrt{2}}q(.90; 3, 6) = \frac{1}{\sqrt{2}}(3.56) = 2.52$$

The estimated variance is the same for all pairwise comparisons:

$$s^2\{\hat{L}\} = \frac{MSE}{bn} + \frac{MSE}{bn} = \frac{2(7.00)}{4} = 3.5$$

so that the estimated standard deviation is $s\{\hat{L}\} = 1.87$ and $Ts\{\hat{L}\} = 2.52(1.87) = 4.71$.

Using the results in Table 26.1, we have:

$$\bar{Y}_{1..} = 19.75 \qquad \bar{Y}_{2..} = 14.25 \qquad \bar{Y}_{3..} = 11.00$$

Hence, the 90 percent family of confidence intervals is:

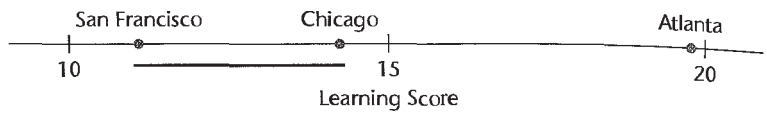
$$.8 = (19.75 - 14.25) - 4.71 \leq \mu_{1.} - \mu_{2.} \leq (19.75 - 14.25) + 4.71 = 10.2$$

$$4.0 = (19.75 - 11.00) - 4.71 \leq \mu_{1.} - \mu_{3.} \leq (19.75 - 11.00) + 4.71 = 13.5$$

$$-1.5 = (14.25 - 11.00) - 4.71 \leq \mu_{2.} - \mu_{3.} \leq (14.25 - 11.00) + 4.71 = 8.0$$

We conclude with 90 percent family confidence coefficient that the mean learning score is highest in Atlanta and that the difference in the observed mean scores for Chicago and San Francisco is not statistically significant. We summarize these results by the following

line plot:



Estimation of Treatment Means μ_{ij}

Confidence limits for μ_{ij} are set up in the usual fashion using the t distribution when both factors A and B have fixed effects:

$$\bar{Y}_{ij}. \pm t[1 - \alpha/2; (n-1)ab]s\{\bar{Y}_{ij}.\} \tag{26.23}$$

where:

$$s^2\{\bar{Y}_{ij}.\} = \frac{MSE}{n} \tag{26.23a}$$

To make a comparison within any factor A level, we estimate the contrast $L = \sum c_j \mu_{ij}$, where $\sum c_j = 0$, with the estimator $\hat{L} = \sum c_j \bar{Y}_{ij}.$ and employ the confidence limits:

$$\hat{L} \pm t[1 - \alpha/2; (n-1)ab]s\{\hat{L}\} \tag{26.24}$$

where:

$$s^2\{\hat{L}\} = \frac{MSE}{n} \sum c_j^2 \tag{26.24a}$$

The Bonferroni procedure may be used when several comparisons are to be made and the family confidence level is to be controlled. The Tukey procedure is also applicable for paired comparisons and the Scheffé procedure for contrasts, but these procedures often will not be efficient since ordinarily only comparisons within each factor level are of interest, whereas the Tukey and Scheffé families are based on comparisons among all ab treatments.

Example

In the training school example, we are to compare the mean scores for the two instructors in each school, using the Bonferroni procedure with a 90 percent family confidence coefficient. For $g = 3$ comparisons, we require $B = t[1 - .10/2(3); 6] = t(.983; 6) = 2.748$. The estimated variance in each case is:

$$s^2\{\hat{L}\} = \frac{7.00}{2}(2) = 7.0$$

Hence, $Bs\{\hat{L}\} = 2.748\sqrt{7.0} = 7.27$. Obtaining the estimated treatment means $\bar{Y}_{ij}.$ from Table 26.1, we find:

$$7.2 = (27 - 12.5) - 7.27 \leq \mu_{11} - \mu_{12} \leq (27 - 12.5) + 7.27 = 21.8$$

$$-18.8 = (8.5 - 20) - 7.27 \leq \mu_{21} - \mu_{22} \leq (8.5 - 20) + 7.27 = -4.2$$

$$7.7 = (18.5 - 3.5) - 7.27 \leq \mu_{31} - \mu_{32} \leq (18.5 - 3.5) + 7.27 = 22.3$$

It is evident that substantial differences between the two instructors exist at each school.

Estimation of Overall Mean $\mu_{..}$

Sometimes there is interest in estimating the overall mean $\mu_{..}$. For the training school example, $\mu_{..}$ is the overall mean learning score for all training schools and all instructors in these schools. The point estimator is $\bar{Y}_{...}$. The confidence limits are constructed utilizing the t distribution as follows:

$$\bar{Y}_{..} \pm t(1-\alpha/2; df)s\{\bar{Y}_{..}\} \tag{26.25}$$

where:

$$s^2\{\bar{Y}_{..}\} = \frac{MSE}{abn} \qquad df = ab(n-1) \qquad A \text{ and } B \text{ fixed} \tag{26.25a}$$

$$s^2\{\bar{Y}_{..}\} = \frac{MSA}{abn} \qquad df = a-1 \qquad A \text{ and } B \text{ random} \tag{26.25b}$$

$$s^2\{\bar{Y}_{..}\} = \frac{MSB(A)}{abn} \qquad df = a(b-1) \qquad A \text{ fixed, } B \text{ random} \tag{26.25c}$$

Example

For the training school example, we wish to estimate the overall mean $\mu_{..}$ with a 95 percent confidence interval. The estimated variance (26.25a) is appropriate here since the model involves fixed factor effects. Hence, we obtain:

$$s^2\{\bar{Y}_{..}\} = \frac{7.00}{12} = .583 \qquad s\{\bar{Y}_{..}\} = .764$$

For confidence coefficient .95, we require $t(.975; 6) = 2.447$. From Table 26.1, we find $\bar{Y}_{..} = 15$. The desired confidence interval therefore is:

$$13.1 = 15 - 2.447(.764) \leq \mu_{..} \leq 15 + 2.447(.764) = 16.9$$

Estimation of Variance Components

With random factor effects, estimates of the variance components may be of interest. No new problems arise for balanced nested designs. For instance, we see from Table 26.5 that when both factors A and B are random factors, the variance component σ_α^2 can be expressed as follows:

$$\sigma_\alpha^2 = \frac{E\{MSA\} - E\{MSB(A)\}}{bn} \tag{26.26}$$

Hence, an unbiased estimator of σ_α^2 is:

$$s_\alpha^2 = \frac{MSA - MSB(A)}{bn} \tag{26.27}$$

Approximate confidence intervals for variance components σ_α^2 or σ_β^2 can be obtained using the MLS interval (25.34). For example, to estimate σ_α^2 when both A and B are random factors, we see from (26.26) that the correspondences to (25.32) are:

$$c_1 = \frac{1}{bn} \qquad MS_1 = MSA$$

$$c_2 = -\frac{1}{bn} \qquad MS_2 = MSB(A)$$

Hence, the MLS confidence interval for σ_α^2 is:

$$s_\alpha^2 - H_L \leq \sigma_\alpha^2 \leq s_\alpha^2 + H_U \tag{26.28}$$

where H_L and H_U are given by the formulas in Table 25.3, $df_1 = a - 1$, $df_2 = a(b-1)$, and s_α^2 is given by (26.27).

26.6 Unbalanced Nested Two-Factor Designs

Up to this point, we have assumed that the nested study is balanced; that is, the same number of levels of factor B is nested within each of the levels of factor A, and the same number of replications is made for each factor level combination. There are occasions, however, when a study is unbalanced. For instance, in our earlier example dealing with the effects of school (factor A) and instructor (factor B) on the learning achieved by classes of mechanics, there might have been b_i instructors in the ith school and n_{ij} classes taught by the jth instructor in school i.

The ANOVA sums of squares formulas given earlier are not appropriate for unbalanced studies. Ordinarily, it is best to use the regression approach for unbalanced studies when the factor effects are fixed. Since no new principles are involved, we proceed directly to an example.

Example

The manufacturing company that conducted the training school study subsequently made a follow-up study involving only Atlanta and Chicago. At that time, three instructors were used in Atlanta and two in Chicago. All instructors were to train two classes, but one class for one of the instructors in Atlanta had to be canceled. The data for this follow-up study are presented in Table 26.6a. We shall again assume that a fixed effects nested design model is appropriate:

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk} \tag{26.29}$$

$$i = 1, 2; j = 1, \ldots, b_i; k = 1, \ldots, n_{ij}$$

$$b_1 = 3, \quad b_2 = 2 \quad n_{11} = n_{13} = 2, \quad n_{12} = 1, \quad n_{21} = n_{22} = 2$$

$$\sum_{i=1}^{2} \alpha_i = 0 \qquad \sum_{j=1}^{3} \beta_{j(1)} = 0 \qquad \sum_{j=1}^{2} \beta_{j(2)} = 0$$

Proceeding as usual, we shall incorporate the parameters α_1, $\beta_{1(1)}$, $\beta_{2(1)}$, and $\beta_{1(2)}$ into the regression model. The other parameters are not required since according to the constraints in (26.29) we have:

$$\alpha_2 = -\alpha_1 \qquad \beta_{3(1)} = -\beta_{1(1)} - \beta_{2(1)} \qquad \beta_{2(2)} = -\beta_{1(2)} \tag{26.30}$$

Thus, we require four indicator variables for our example, each taking on values 1, −1, or 0.

The equivalent regression model therefore is:

$$Y_{ijk} = \mu_{..} + \underbrace{\alpha_1 X_{ijk1}}_{\text{School main effect}} + \underbrace{\beta_{1(1)} X_{ijk2} + \beta_{2(1)} X_{ijk3} + \beta_{1(2)} X_{ijk4}}_{\text{Specific instructor within school effect}} + \varepsilon_{ijk} \qquad \text{Full model} \tag{26.31}$$

TABLE 26.6 Nested Unbalanced Two-Factor Study—Follow-up Training School Study.

(a) Data

Study Replication	Atlanta (A_1)			Chicago (A_2)	
k	B_1	B_2	B_3	B_1	B_2
1	20	8	9	4	16
2	22		13	8	20

(b) Y and X Variables for Regression Approach

			(1)	(2)	(3)	(4)	(5)
i	j	k	Y	X_1	X_2	X_3	X_4
1	1	1	20	1	1	0	0
1	1	2	22	1	1	0	0
1	2	1	8	1	0	1	0
1	3	1	9	1	−1	−1	0
1	3	2	13	1	−1	−1	0
2	1	1	4	−1	0	0	1
2	1	2	8	−1	0	0	1
2	2	1	16	−1	0	0	−1
2	2	2	20	−1	0	0	−1

where:

$$X_1 = \begin{cases} 1 & \text{if class from school 1} \\ -1 & \text{if class from school 2} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{if class for instructor 1 in school 1} \\ -1 & \text{if class for instructor 3 in school 1} \\ 0 & \text{otherwise} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{if class for instructor 2 in school 1} \\ -1 & \text{if class for instructor 3 in school 1} \\ 0 & \text{otherwise} \end{cases}$$

$$X_4 = \begin{cases} 1 & \text{if class for instructor 1 in school 2} \\ -1 & \text{if class for instructor 2 in school 2} \\ 0 & \text{otherwise} \end{cases}$$

The Y observations and X indicator variables for this example are shown in Table 26.6b.

To test for school main effects, we first fit full model (26.31) by regressing Y in Table 26.6b, column 1, on X_1, X_2, X_3, X_4 in columns 2–5, and obtain $SSE(F)$. We then fit the reduced model for H_0: $\alpha_1 = 0$:

$$Y_{ijk} = \mu_{..} + \beta_{1(1)}X_{ijk2} + \beta_{2(1)}X_{ijk3} + \beta_{1(2)}X_{ijk4} + \varepsilon_{ijk} \qquad \text{Reduced model} \qquad \textbf{(26.32)}$$

by regressing Y in column 1 on X_2, X_3, X_4 in columns 3–5, and obtain $SSE(R)$. The difference $SSE(R) - SSE(F)$ equals SSA. Test statistic (2.70) is then obtained in the usual fashion.

TABLE 26.7 ANOVA Table for Nested Unbalanced Two-Factor Study—Follow-up Training School Study.

Source of Variation	*SS*	*df*	*MS*	F^*
Schools (A)	3.76	1	3.76	3.76/6.5 = .58
Instructors [$B(A)$]	295.20	3	98.4	98.4/6.5 = 15.1
Error (E)	26.00	4	6.5	

To test for specific instructor effects, we employ the reduced model for H_0: $\beta_{1(1)} = \beta_{2(1)} = \beta_{1(2)} = 0$:

$$Y_{ijk} = \mu_{..} + \alpha_1 X_{ijk1} + \varepsilon_{ijk} \qquad \text{Reduced model} \tag{26.33}$$

We therefore regress Y in column 1 on X_1 in column 2, and obtain $SSE(R)$. The difference $SSE(R) - SSE(F)$ equals $SSB(A)$.

Table 26.7 contains the ANOVA table for the follow-up training school study. No total sum of squares is shown because the component sums of squares are not orthogonal.

The tests for school and instructor effects are carried out as before. Estimation of factor effects is done by means of the regression parameters. For instance, a comparison of the mean scores for the two schools involves:

$$\mu_{1.} - \mu_{2.} = \alpha_1 - \alpha_2$$

Since $\alpha_2 = -\alpha_1$ by (26.30), we need to estimate:

$$\mu_{1.} - \mu_{2.} = \alpha_1 - (-\alpha_1) = 2\alpha_1$$

An unbiased estimator is $2\hat{\alpha}_1$. Other desired estimates are obtained in a similar fashion.

26.7 Subsampling in Single-Factor Study with Completely Randomized Design

Up to this point in our discussion of experimental designs, we have considered only designs in which one observation of the response variable is made on an experimental unit. There are occasions, however, when more than one observation is desirable. Consider an experiment to study the effect of oven temperature on crustiness of bread. Three temperatures were utilized, and two experimental units (batches of flour mix) were randomly assigned to each treatment. It was not economical to use the entire batch to bake breads, nor was it technically feasible to use a batch as a block. Hence, three subsamples were selected from each batch to make three loaves, which were baked at a given temperature. Here, then, three observations (subsamples) were made on each experimental unit (batch).

Another instance of several observations on the response variable being made for each experimental unit occurred in an experiment on the effectiveness of three different training methods. The experimental units here were persons, and the experiment sought to measure the length of time required to perform a certain engine assembly operation after the given training program was completed. Ten consecutive assemblies were timed, and these constituted the subsamples of the experimental unit (person).

Formally, subsampling (i.e., repeated observations on the same experimental unit) is completely analogous to nested factors. We shall demonstrate this for a completely randomized design.

odel

Consider again the experiment to study the effect of oven temperature on the crustiness of bread. The model for this study can be written as follows:

$$Y_{ijk} = \mu_{..} + \tau_i + \varepsilon_{j(i)} + \eta_{ijk} \tag{26.34}$$

The meaning of the symbols is as follows:

1. $\mu_{..}$ is an overall constant.
2. τ_i is the temperature (i.e., treatment) effect (fixed effect, here).
3. $\varepsilon_{j(i)}$ is the experimental error associated with the particular batch (random effect, here). The experimental error is nested within the treatment, since the jth batch for treatment i was not used with any other treatment.
4. η_{ijk} is the error associated with the kth subsample or observation on the jth experimental unit for the ith treatment (random effect, here).

Note that subsampling model (26.34) appears the same as nested design model (26.7) for a nested two-factor design, except for changes in notation to reflect the fact that subsampling model (26.34) is a single-factor model and contains both experimental and observation errors. Specifically, the treatment effect τ_i here corresponds to α_i in the nested two-factor model, the batch effect $\varepsilon_{j(i)}$ corresponds to $\beta_{j(i)}$, and the observation error term η_{ijk} corresponds to ε_{ijk}. Consequently, the analysis of variance for the case of subsampling in a single-factor study with a completely randomized design parallels that for a nested two-factor study.

In general, the model for subsampling in a balanced single-factor study with a completely randomized design where the treatment effects are fixed is:

$$Y_{ijk} = \mu_{..} + \tau_i + \varepsilon_{j(i)} + \eta_{ijk} \tag{26.35}$$

where:

$\mu_{..}$ is a constant
τ_i are constants subject to the restriction $\sum \tau_i = 0$
$\varepsilon_{j(i)}$ are independent $N(0, \sigma^2)$
η_{ijk} are independent $N(0, \sigma_\eta^2)$
$\varepsilon_{j(i)}$ and η_{ijk} are independent
$i = 1, \ldots, r;\ j = 1, \ldots, n;\ k = 1, \ldots, m$

The mean and variance of observation Y_{ijk} for this model are:

$$E\{Y_{ijk}\} = \mu_{..} + \tau_i \tag{26.36a}$$

$$\sigma^2\{Y_{ijk}\} = \sigma_Y^2 = \sigma^2 + \sigma_\eta^2 \tag{26.36b}$$

Further, the observations Y_{ijk} are normally distributed for this model. Observations from different replications (i.e., from different subsamples) are independent, but any two

observations from the same replication are correlated in advance of the random trials because they contain the same random term $\varepsilon_{j(i)}$:

$$\sigma\{Y_{ijk}, Y_{ijk'}\} = \sigma^2 \qquad k \neq k' \tag{26.36c}$$

$$\sigma\{Y_{ijk}, Y_{i'j'k'}\} = 0 \qquad i \neq i' \text{ and/or } j \neq j' \tag{26.36d}$$

Analysis of Variance and Tests of Effects

The appropriate sums of squares for the analysis of variance for balanced subsampling model (26.35) are as follows:

$$SSTO = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2 \tag{26.37a}$$

$$SSTR = nm \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2 \tag{26.37b}$$

$$SSEE = m \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..})^2 \tag{26.37c}$$

$$SSOE = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2 \tag{26.37d}$$

Here, *SSEE* stands for the *experimental error sum of squares,* and *SSOE* stands for the *observation error sum of squares.* Note the correspondence of formulas (26.37) to formulas (26.13) for nested two-factor designs. The only difference is that we now have $i = 1, \ldots, r$, $j = 1, \ldots, n$, and $k = 1, \ldots, m$, whereas before i, j, and k ran to a, b, and n, respectively.

Table 26.8 contains the ANOVA for a single-factor completely randomized balanced experiment with subsampling. Also shown there are the expected mean squares for both fixed and random treatment effects. Note that regardless of whether treatment effects are fixed or random, the appropriate statistic for testing treatment effects is:

$$F^* = \frac{MSTR}{MSEE} \tag{26.38a}$$

TABLE 26.8 ANOVA for Single-Factor Completely Randomized Balanced Experiment with Subsampling.

Source of Variation	*SS*	*df*	*MS*	$E\{MS\}$	
				τ_i Fixed	τ_i Random
Treatments	*SSTR*	$r-1$	*MSTR*	$\sigma_\eta^2 + m\sigma^2 + nm\frac{\sum \tau_i^2}{r-1}$	$\sigma_\eta^2 + m\sigma^2 + nm\sigma_\tau^2$
Experimental error	*SSEE*	$r(n-1)$	*MSEE*	$\sigma_\eta^2 + m\sigma^2$	$\sigma_\eta^2 + m\sigma^2$
Observation error	*SSOE*	$rn(m-1)$	*MSOE*	σ_η^2	σ_η^2
Total	*SSTO*	$rnm-1$			

A test for the presence of experimental error effects, i.e., $H_0: \sigma^2 = 0$, $H_a: \sigma^2 > 0$, also uses the same test statistic for both fixed and random treatment effects:

$$F^* = \frac{MSEE}{MSOE} \tag{26.38b}$$

Example

The data for the study of the effects of baking temperature on the crustiness of bread are contained in Table 26.9. The data are scores on a scale from 1 to 20. Figure 26.5 presents SYSTAT aligned dot plots of the data. These plots suggest the presence of temperature effects and possibly also batch effects. Note that crustiness increases steadily with the level of temperature.

The appropriate analysis of variance was obtained from a computer run and is presented in Table 26.10. To test for temperature effects:

$$H_0: \tau_1 = \tau_2 = \tau_3 = 0$$
$$H_a: \text{not all } \tau_i \text{ equal zero}$$

we use test statistic (26.38a):

$$F^* = \frac{117.72}{16.33} = 7.21$$

TABLE 26.9 Data for Single-Factor Completely Randomized Balanced Experiment with Subsampling—Bread Crustiness Example.

	Temperature					
	Low ($i = 1$)		Medium ($i = 2$)		High ($i = 3$)	
Observation Unit k	Batch 1 $j = 1$	Batch 2 $j = 2$	Batch 3 $j = 1$	Batch 4 $j = 2$	Batch 5 $j = 1$	Batch 6 $j = 2$
1	4	12	14	9	14	16
2	7	8	13	10	17	19
3	5	10	11	12	15	18

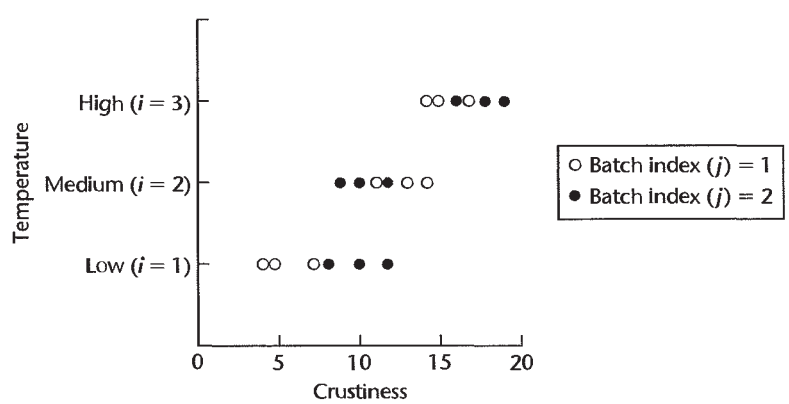


FIGURE 26.5 SYSTAT Dot Plots for Subsampling Experiment—Bread Crustiness Example.

TABLE 26.10 ANOVA—Bread Crustiness Example.

Source of Variation	*SS*	*df*	*MS*
Temperatures (*TR*)	235.44	2	117.72
Mix batches (*EE*)	49.00	3	16.33
Observation units (*OE*)	31.33	12	2.61
Total	315.78	17	

For level of significance $\alpha = .10$, we need $F(.90; 2, 3) = 5.46$. Since $F^* = 7.21 > 5.46$, we conclude H_a, that baking temperature does have an effect on the crustiness of the bread. The P-value of the test is .07.

To test for batch differences:

$$H_0: \sigma^2 = 0$$
$$H_a: \sigma^2 > 0$$

we employ test statistic (26.38b):

$$F^* = \frac{16.33}{2.61} = 6.26$$

For level of significance $\alpha = .10$, we need $F(.90; 3, 12) = 2.61$. Since $F^* = 6.26 > 2.61$, we conclude H_a, that there are batch effects on the crustiness of bread. The P-value of this test is .01. Thus, both the particular batch of flour mix and the temperature at which the bread is baked affect the crustiness of the loaf.

Estimation of Treatment Effects

When the treatment effects are fixed, there is usually interest in obtaining confidence intervals for treatment means $\mu_{i\cdot} = \mu_{\cdot\cdot} + \tau_i$ and for pairwise comparisons and contrasts of the treatment means. These can be obtained in the usual manner, using *MSEE* as the error variance since this is the quantity in the denominator of the test statistic for fixed treatment effects. The degrees of freedom are those associated with *MSEE*, namely, $(n-1)r$. For instance, the confidence limits for treatment mean $\mu_{i\cdot}$ are:

$$\bar{Y}_{i\cdot\cdot} \pm t[1-\alpha/2; (n-1)r]s\{\bar{Y}_{i\cdot\cdot}\} \tag{26.39}$$

where:

$$s^2\{\bar{Y}_{i\cdot\cdot}\} = \frac{MSEE}{nm} \tag{26.39a}$$

Similarly, confidence limits for a contrast of treatment means, $L = \sum c_i \mu_{i\cdot}$, where $\sum c_i = 0$, are obtained as follows:

$$\hat{L} \pm t[1-\alpha/2; (n-1)r]s\{\hat{L}\} \tag{26.40}$$

where:

$$\hat{L} = \sum c_i \bar{Y}_{i\cdot\cdot} \tag{26.40a}$$

$$s^2\{\hat{L}\} = \frac{MSEE}{nm} \sum c_i^2 \tag{26.40b}$$

The Bonferroni, Tukey, and Scheffé simultaneous inference procedures can be utilized in the usual manner.

Example

In the bread crustiness example, we wish to estimate the mean crustiness of bread baked at a low temperature with a 95 percent confidence coefficient. We require, using the results in Tables 26.9 and 26.10:

$$\bar{Y}_{1\cdot\cdot} = 7.67$$

$$s^2\{\bar{Y}_{1\cdot\cdot}\} = \frac{16.33}{6} = 2.722 \qquad s\{\bar{Y}_{1\cdot\cdot}\} = 1.65$$

$$t(.975; 3) = 3.182$$

Hence, the 95 percent confidence interval is:

$$2.4 = 7.67 - 3.182(1.65) \leq \mu_{1\cdot} \leq 7.67 + 3.182(1.65) = 12.9$$

It was also desired to estimate the difference in mean crustiness of bread baked at high and low temperatures with a 95 percent confidence interval. Utilizing (26.40) and the results in Tables 26.9 and 26.10, we obtain:

$$\bar{Y}_{1\cdot\cdot} = 7.67 \qquad \bar{Y}_{3\cdot\cdot} = 16.5$$

$$\hat{L} = \bar{Y}_{3\cdot\cdot} - \bar{Y}_{1\cdot\cdot} = 16.5 - 7.67 = 8.83$$

$$s^2\{\hat{L}\} = \frac{2(16.33)}{6} = 5.443 \qquad s\{\hat{L}\} = 2.33$$

Hence, the desired confidence interval is:

$$1.4 = 8.83 - 3.182(2.33) \leq \mu_{3\cdot} - \mu_{1\cdot} \leq 8.83 + 3.182(2.33) = 16.2$$

Estimation of Variances

At times, there is interest in estimating σ^2, the experimental error variance, and σ_η^2, the observation error variance. It is evident from either of the $E\{MS\}$ columns in Table 26.8 that the following are unbiased estimators:

Parameter	Unbiased Estimator	
σ^2	$s^2 = \dfrac{MSEE - MSOE}{m}$	**(26.41a)**
σ_η^2	$s_\eta^2 = MSOE$	**(26.41b)**

An approximate confidence interval for the experimental error variance σ^2 is easily obtained by the modified large sample procedure in (25.34). From Table 26.8, we have:

$$\sigma^2 = \frac{E\{MSEE\} - E\{MSOE\}}{m}$$

Thus σ^2 takes the form (25.32) with correspondences:

$$c_1 = \frac{1}{m} \qquad MS_1 = MSEE$$

$$c_2 = -\frac{1}{m} \qquad MS_2 = MSOE$$

The MLS approximate $1 - \alpha$ confidence interval for σ^2 is therefore:

$$s^2 - H_L \leq \sigma^2 \leq s^2 + H_U \tag{26.42}$$

where H_L and H_U are given by the formulas in Table 25.3, $df_1 = r(n-1)$ and $df_2 = rn(m-1)$, and s^2 is given in (26.41a).

An exact confidence interval for the observation error variance σ_η^2 can be obtained by (25.21), with *MSOE* now being the mean square and $rn(m-1)$ now being the degrees of freedom.

Example

For the bread crustiness example, we wish to estimate σ^2, the variability between batches, with a 95 percent confidence interval. From Table 26.10, we obtain the point estimate:

$$s^2 = \frac{16.33 - 2.61}{3} = 4.57$$

To obtain an approximate 95 percent confidence interval for σ^2 using (26.42), we need the following calculational results for the formulas in Table 25.3:

$$F_1 = 3.12 \quad F_2 = 1.95 \quad F_3 = 13.92 \quad F_4 = 2.73 \quad F_5 = 4.47 \quad F_6 = 14.34$$
$$G_1 = .6795 \quad G_2 = .4872 \quad G_3 = -.0397 \quad G_4 = -2.6347$$
$$H_L = 3.97 \quad H_U = 70.24$$

The desired confidence interval for σ^2 is therefore:

$$.60 = 4.57 - 3.97 \leq \sigma^2 \leq 4.57 + 70.24 = 74.81$$

and for σ, the experimental error standard deviation, the confidence interval is:

$$.77 \leq \sigma \leq 8.65$$

Comments

1. Frequently, the units for subsampling are called *observation units*, to distinguish them from the *experimental units*. For instance, in the bread crustiness example, the batches of flour mix are the experimental units and the portions selected from a batch for making loaves of bread are the observation units.

2. Observation units may be different physical entities, as in the bread crustiness example where they are portions of a batch of flour mix. Observation units also may refer to repeated observations on the entire experimental unit. An example of the latter is the earlier illustration where an employee is timed for 10 consecutive assembly operations after receiving a given type of training.

3. Note that subsampling model (26.35) contains no interaction terms. This is because the experimental error terms $\varepsilon_{j(i)}$ are nested within treatments. When one variable is nested within another, we saw earlier that interaction terms are inapplicable.

4. We have considered only the balanced case for subsampling, where an equal number of experimental units (n) are applied to each treatment and a constant number of observations (m) are made on each experimental unit. Serious complications are encountered in the unbalanced case, and no exact test for treatment effects can be made. See an advanced text, such as Reference 26.1, for a discussion. ■

26.8 Pure Subsampling in Three Stages

Sometimes an investigation does not involve a comparison of treatments, but only subsampling at several levels. Consider, for instance, a quality control engineer who wishes to investigate a certain quality characteristic of a computer assembly. These assemblies are produced in lots of 2,000. The engineer will select a random sample of r lots; from each lot n assemblies will be selected, and m observations will be made on the quality characteristic for each assembly.

Model

Assuming that all random variables are normally distributed and that equal sample sizes are employed at each stage, the model for subsampling in three stages is:

$$Y_{ijk} = \mu_{..} + \tau_i + \varepsilon_{j(i)} + \eta_{ijk} \tag{26.43}$$

where:

$\mu_{..}$ is a constant

τ_i, $\varepsilon_{j(i)}$, and η_{ijk} are independent normal random variables with expectations 0 and variances σ_τ^2, σ^2, and σ_η^2, respectively

$i = 1, \ldots, r$; $j = 1, \ldots, n$; $k = 1, \ldots, m$

For our quality control illustration, τ_i represents the lot effect, $\varepsilon_{j(i)}$ represents the assembly effect that is nested within the lot, and η_{ijk} represents the observation effect.

The observations Y_{ijk} for subsampling model (26.43) are normally distributed, with mean and variance:

$$E\{Y_{ijk}\} = \mu_{..} \tag{26.44a}$$

$$\sigma^2\{Y_{ijk}\} = \sigma_Y^2 = \sigma_\tau^2 + \sigma^2 + \sigma_\eta^2 \tag{26.44b}$$

Various correlations exist between two observations from the same lot.

Subsampling model (26.43) corresponds to subsampling model (26.35) for a single-factor study except that we assume here that the τ_i are independent $N(0, \sigma_\tau^2)$ and are independent of the $\varepsilon_{j(i)}$ and η_{ijk}. Formally, then, the only difference between models (26.35) and (26.43) is that the τ_i are fixed in one case and random in the other. Subsampling model (26.43) also corresponds to nested model (26.7) with both factor A and factor B effects random.

Analysis of Variance

The analysis of variance for pure subsampling model (26.43) uses the same sums of squares as before, namely, those in (26.37). The ANOVA table is the same as that in Table 26.8. The applicable expected mean squares are those for random τ_i effects.

Estimation of $\mu_{..}$

In the case of pure subsampling, there is often interest in estimating the overall mean $\mu_{..}$ (the process mean for the computer assembly quality characteristic in our earlier quality control example). A point estimator of $\mu_{..}$ in model (26.43) is $\bar{Y}_{...}$, and it can be shown that

its variance is:

$$\sigma^2\{\bar{Y}_{...}\} = \frac{\sigma_\tau^2}{r} + \frac{\sigma^2}{rn} + \frac{\sigma_\eta^2}{rnm} = \frac{nm\sigma_\tau^2 + m\sigma^2 + \sigma_\eta^2}{rnm} \tag{26.45}$$

An unbiased estimator of this variance is:

$$s^2\{\bar{Y}_{...}\} = \frac{MSTR}{rnm} \tag{26.46}$$

and the $1 - \alpha$ confidence limits for $\mu_{..}$ are:

$$\bar{Y}_{...} \pm t(1 - \alpha/2; r - 1)s\{\bar{Y}_{...}\} \tag{26.47}$$

26.9 Three-Factor Partially Nested Designs

Our discussion of nested designs and subsampling so far has been confined to hierarchical designs where no factors are crossed. In this section, we consider three-factor experiments where some but not all of the factors are nested. Such designs are called *partially nested, partially hierarchical,* or *cross-nested designs.* We shall utilize the following example to explain three-factor partially nested designs.

Example

The effect of cultural background on group decision making was studied by an experiment. Sixteen teams of students were formed and assigned a task. One of the response variables was the number of group interactions prior to the final group decision. Eight teams consisted of foreign students, eight of U.S. students. Half of the teams consisted of eight members, the other half of four members. Two foreign observers were used for the foreign teams, and two U.S. observers for the U.S. teams. Thus, the design may be represented as follows:

	U.S. Teams (A_1)		Foreign Teams (A_2)	
	Observer 1 (C_1)	Observer 2 (C_2)	Observer 3 (C_1)	Observer 4 (C_2)
Small team (B_1)	Replication 1 Replication 2	Replication 1 Replication 2	Replication 1 Replication 2	Replication 1 Replication 2
Large team (B_2)	Replication 1 Replication 2	Replication 1 Replication 2	Replication 1 Replication 2	Replication 1 Replication 2

Note that there are two replications (teams) in each cell.

Development of Model

Let nationality of team be factor A, size of team factor B, and observer factor C. Note that factor C is nested within factor A since the two observers for the U.S. teams were different from the two observers for the foreign teams. Also note that factors A and B are crossed, since each level of factor A appears with every level of factor B, and vice versa. Similarly, factors B and C are crossed. Factors A (nationality) and B (team size) were considered to have fixed effects, while the factor C (observer) effects were considered to be random.

In order to develop an appropriate model, we need to recognize that factor C is nested within factor A; hence the factor C effect is denoted by $\gamma_{k(i)}$. We also need to recognize that the AC and ABC interactions are to be excluded because factor C is nested within factor A. Finally, the BC interaction is nested within factor A since factor C is nested within factor A;

thus, the BC interaction is denoted by $(\beta\gamma)_{jk(i)}$. Hence, the appropriate model is:

$$Y_{ijkm} = \mu_{...} + \alpha_i + \beta_j + \gamma_{k(i)} + (\alpha\beta)_{ij} + (\beta\gamma)_{jk(i)} + \varepsilon_{ijkm} \qquad \textbf{(26.48)}$$

where:

$\mu_{...}$ is an overall constant
α_i are the fixed nationality effects
β_j are the fixed team size effects
$\gamma_{k(i)}$ are the random observer (within nationality) effects
$(\alpha\beta)_{ij}$ are the fixed nationality–team size interaction effects
$(\beta\gamma)_{jk(i)}$ are the random team size–observer interaction (within nationality) effects
ε_{ijkm} are random error terms

$\sum_i \alpha_i = 0 \qquad \sum_j \beta_j = 0 \qquad \sum_i (\alpha\beta)_{ij} = 0 \quad$ for all j
$\sum_j (\alpha\beta)_{ij} = 0 \quad$ for all $i \qquad \sum_j (\beta\gamma)_{jk(i)} = 0 \quad$ for all $k(i)$
$i = 1, \ldots, a; j = 1, \ldots, b; k = 1, \ldots, c; m = 1, \ldots, n$

Appendix D contains a simple rule for constructing ANOVA models for complex designs, such as the one here.

We assume as usual that $\gamma_{k(i)}$, $(\beta\gamma)_{jk(i)}$, and ε_{ijkm} are normally distributed with expectations zero and with constant variances σ^2_γ, $\sigma^2_{\beta\gamma}$, and σ^2, respectively, and that the three groups of random variables are pairwise independent. The interaction effects $(\beta\gamma)_{jk(i)}$ for any given observer are correlated as a result of the restrictions in model (26.48).

Analysis of Variance

Table 26.11 contains the ANOVA table for model (26.48). The sums of squares, degrees of freedom, and expected mean squares shown in this table can be developed by using the rules in Appendix D. The expected mean squares also can be obtained from some computer packages with analysis of variance capabilities. The expected mean squares column in Table 26.11 indicates directly how to form test statistics for a variety of tests.

Example

Table 26.12 contains the results of the group decision-making experiment described earlier, and Figure 26.6 presents SYSTAT aligned dot plots of the data. The dot plots suggest a strong effect of nationality on the number of group interactions before the group decision is reached. Figure 26.7 contains the MINITAB printout of the ANOVA results, including the expected mean squares and the appropriate F tests. The correspondences between the symbols used in MINITAB in its expected mean square column and the model terms in Table 26.11 are as follows: Each term in an expected mean square is represented in the MINITAB output by (1) the numeric code, in parentheses, for the variance of the model term, and (2) the preceding number which is the numerical multiple. When the model effect is fixed, the letter Q is used in the printout to show that the variance of the model term is replaced by the sum of squared effects divided by degrees of freedom. For example:

$$E\{MSA\} = (6) + 4(3) + 8\text{Q}[1] = \sigma^2 + 4\sigma^2_\gamma + 8\frac{\sum \alpha_i^2}{2-1}$$

$$E\{MSBC(A)\} = (6) + 2(5) = \sigma^2 + 2\sigma^2_{\beta\gamma}$$

TABLE 26.11 ANOVA Table for Crossed-Nested Model (26.48).

Source of Variation	*SS*	*df*	*MS*	Expected Mean Squares
A	$SSA = bcn\sum(\bar{Y}_{i\cdots} - \bar{Y}_{\cdots})^2$	$a-1$	*MSA*	$\sigma^2 + bcn\frac{\sum\alpha_i^2}{a-1} + bn\sigma_\gamma^2$
B	$SSB = acn\sum(\bar{Y}_{\cdot j\cdot\cdot} - \bar{Y}_{\cdots})^2$	$b-1$	*MSB*	$\sigma^2 + acn\frac{\sum\beta_j^2}{b-1} + n\sigma_{\beta\gamma}^2$
C(A)	$SSC(A) = bn\sum\sum(\bar{Y}_{i\cdot k\cdot} - \bar{Y}_{i\cdots})^2$	$a(c-1)$	*MSC(A)*	$\sigma^2 + bn\sigma_\gamma^2$
AB	$SSAB = cn\sum\sum(\bar{Y}_{ij\cdot\cdot} - \bar{Y}_{i\cdots} - \bar{Y}_{\cdot j\cdot\cdot} + \bar{Y}_{\cdots})^2$	$(a-1)(b-1)$	*MSAB*	$\sigma^2 + cn\frac{\sum\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)} + n\sigma_{\beta\gamma}^2$
BC(A)	$SSBC(A) = n\sum\sum\sum(\bar{Y}_{ijk\cdot} - \bar{Y}_{ij\cdot\cdot} - \bar{Y}_{i\cdot k\cdot} + \bar{Y}_{i\cdots})^2$	$a(b-1)(c-1)$	*MSBC(A)*	$\sigma^2 + n\sigma_{\beta\gamma}^2$
Error	$SSE = \sum\sum\sum\sum(Y_{ijkm} - \bar{Y}_{ijk\cdot})^2$	$abc(n-1)$	*MSE*	σ^2
Total	$SSTO = \sum\sum\sum\sum(Y_{ijkm} - \bar{Y}_{\cdots})^2$	$abcn-1$		

TABLE 26.12 Data for Crossed-Nested Three-Factor Study—Group Decision-Making Example.

	U.S. Teams ($i = 1$)		Foreign Teams ($i = 2$)	
Size of Team	Observer 1 ($k = 1$)	Observer 2 ($k = 2$)	Observer 3 ($k = 1$)	Observer 4 ($k = 2$)
4 members ($j = 1$)	16	14	7	4
	20	19	5	9
8 members ($j = 2$)	21	28	11	12
	25	19	17	15

FIGURE 26.6 SYSTAT Dot Plots for Crossed-Nested Design Experiment—Group Decision-Making Example.

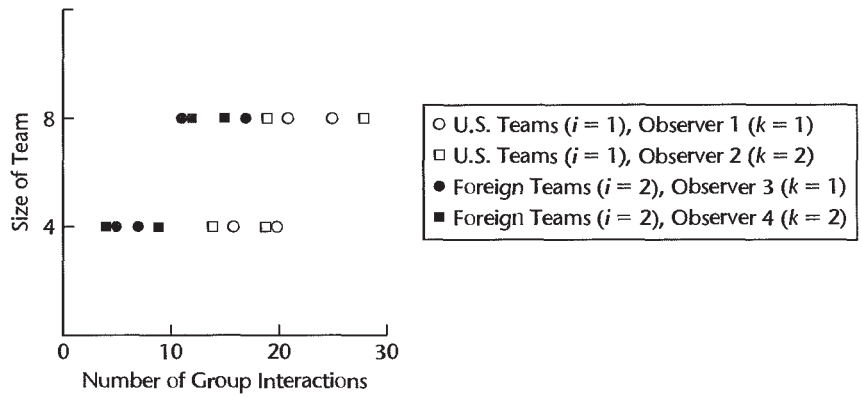


FIGURE 26.7 MINITAB Output for Crossed-Nested Design Experiment—Group Decision-Making Example.

Analysis of Variance

Source	DF	SS	MS	F	P
A	1	420.25	420.25	1681.00	0.001
B	1	182.25	182.25	145.80	0.007
C(A)	2	0.50	0.25	0.02	0.981
A*B	1	2.25	2.25	1.80	0.312
B*C(A)	2	2.50	1.25	0.09	0.911
Error	8	106.00	13.25		
Total	15	713.75			

Source	Variance component	Error term	Expected Mean Square (using restricted model)
1 A		3	(6) + 4(3) + 8Q[1]
2 B		5	(6) + 2(5) + 8Q[2]
3 C(A)	−3.250	6	(6) + 4(3)
4 A*B		5	(6) + 2(5) + 4Q[4]
5 B*C(A)	−6.000	6	(6) + 2(5)
6 Error	13.250		(6)

To test for nationality effects, the alternatives are:

$$H_0: \alpha_1 = \alpha_2 = 0$$
$$H_a: \text{not both } \alpha_i \text{ equal zero} \tag{26.49a}$$

Table 26.11 indicates that the appropriate test statistic is:

$$F^* = \frac{MSA}{MSC(A)} \tag{26.49b}$$

We have for our example, using the results in Figure 26.7:

$$F^* = \frac{420.25}{.25} = 1{,}681$$

For level of significance $\alpha = .05$, we require $F(.95; 1, 2) = 18.5$. Since $F^* = 1{,}681 > 18.5$, we conclude H_a, that nationality has an effect on the group behavior. The P-value of the test is .001. Other tests are conducted in a similar fashion. Results are summarized in Figure 26.7.

Next, we wish to estimate the difference between U.S. and foreign teams in the mean number of group interactions prior to a decision. Confidence intervals for contrasts of main factor effects are set up in the usual way when the factor effects are fixed. Hence, we require $MSC(A)$, as this is the mean square used in the denominator of the test statistic for examining nationality effects. Specifically, the confidence limits for $L = \mu_{1..} - \mu_{2..}$ are:

$$\hat{L} \pm t[1 - \alpha/2; (c-1)a]s\{\hat{L}\} \tag{26.50}$$

where:

$$s^2\{\hat{L}\} = \frac{2MSC(A)}{nbc} \tag{26.50a}$$

For our example, we obtain from Table 26.12 and Figure 26.7:

$$\bar{Y}_{1...} = 20.25 \qquad \bar{Y}_{2...} = 10.00 \qquad \hat{L} = 20.25 - 10.00 = 10.25$$
$$s^2\{\hat{L}\} = \frac{2(.25)}{8} = .063 \qquad s\{\hat{L}\} = .25$$

For confidence coefficient .95, we require $t(.975; 2) = 4.303$. The confidence limits then are $10.25 \pm 4.303(.25)$, and the desired 95 percent confidence interval is:

$$9.2 \leq \mu_{1..} - \mu_{2..} \leq 11.3$$

With confidence coefficient .95, we conclude that U.S. teams engage in 9.2 to 11.3 more interactions, on average, than foreign teams before a group decision is reached.

Comments

1. The sums of squares SSA, SSB, and $SSAB$ in Table 26.11 for the analysis of the crossed-nested experimental design are the usual sums of squares for factor A main effects, factor B main effects, and AB interactions. $SSC(A)$ simply measures the variability of the factor C level estimated means for any given level of factor A, and then aggregates these sums of squares over factor A. Similarly, $SSBC(A)$ contains the usual BC interaction sum of squares for a given level of factor A, and then aggregates these sums of squares over factor A.

2. If important AB interactions are present, analysis should usually focus on the means $\mu_{ij\cdot}$ when the factors have fixed effects, rather than on the factor level means $\mu_{i\cdot\cdot}$ and $\mu_{\cdot j\cdot}$. It can be shown that the estimated variance for comparing the two team sizes for any given nationality is:

$$s^2\{\bar{Y}_{i1\cdot\cdot} - \bar{Y}_{i2\cdot\cdot}\} = \frac{2MSBC(A)}{cn} \tag{26.51}$$

This variance has associated with it $a(b-1)(c-1)$ degrees of freedom, as is evident from Table 26.11.

No exact confidence interval exists for comparing the two nationalities for any given team size. An unbiased variance estimator that can be utilized is:

$$s^2\{\bar{Y}_{1j\cdot\cdot} - \bar{Y}_{2j\cdot\cdot}\} = \frac{2}{cn}\left[MSBC(A) + \frac{MSC(A) - MSE}{b}\right] \tag{26.52}$$

The approximate number of degrees of freedom associated with this variance is obtained from (25.28).

The reason for the different variances in (26.51) and (26.52) is that the observers are the same when the two team sizes for a given nationality are compared, while the observers differ when the two nationalities for a given team size are compared. ■

Cited Reference

26.1. Searle, S. R. *Linear Models for Unbalanced Data.* New York: John Wiley & Sons, 1987.

Problems

26.1. A student asked: "Since the mean squares in the analysis of variance table for a two-factor nested design are the same whether the factor effects are assumed to be random or fixed, what difference does it make whether we assume the factors to have fixed effects or random effects?" Comment.

26.2. A researcher declared: "I prefer analyzing a nested two-factor study as a study with crossed factors because I can isolate more sources of variation." Comment on the researcher's strategy.

26.3. Consider a three-factor study where factor C is nested within factor B, and factor B in turn is nested within factor A, and $a = b = c = 2$. Illustrate in the format of Figure 26.1 the distinction between this nested design and the corresponding crossed design.

26.4. **Bottling plant production.** A production engineer studied the effects of machine model (factor A) and operator (factor B) on the output in a bottling plant. Three bottling machines were used, each a different model. Twelve operators were employed. Four operators were assigned to a machine and worked six-hour shifts each. Data on the number of cases produced by each machine and operator were collected for a week. The data that follow represent the number of cases produced per hour for each day during the week.

Machine i:	1				2				3			
Operator j:	1	2	3	4	1	2	3	4	1	2	3	4
Day $k = 1$:	65	68	56	45	74	69	52	73	69	63	81	67
$k = 2$:	58	62	65	56	81	76	56	78	83	70	72	79
$k = 3$:	63	75	58	54	76	80	62	83	74	72	73	73
$k = 4$:	57	64	70	48	80	78	58	75	78	68	76	77
$k = 5$:	66	70	64	60	68	73	51	76	80	75	70	71

a. Obtain the residuals for nested design model (26.7) with fixed factor effects and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings about the appropriateness of model (26.7)?

b. Prepare aligned residual dot plots by machine. Do these plots support the assumption of constancy of the error variance? Discuss.

26.5. Refer to **Bottling plant production** Problem 26.4. Assume that nested design model (26.7) with fixed factor effects is appropriate.

a. Can the operator effects be distinguished from the effects of shifts in this study? Discuss.

b. Plot the data in the format of Figure 26.3. Does it appear that any factor effects are present?

c. Obtain the analysis of variance table.

d. Test whether or not the mean outputs differ for the three machine models; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Test whether or not the mean outputs differ for the operators assigned to each machine; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test? What does your conclusion imply about the mean outputs for the four operators assigned to machine 3? Explain.

f. Test for each machine separately whether or not the mean outputs for the four operators differ. For each test, use $\alpha = .01$ and state the alternatives, decision rule, and conclusion.

g. What is the family level of significance for the combined tests in parts (d), (e), and (f) using the Bonferroni inequality? Summarize the set of conclusions reached in your tests.

26.6. Refer to **Bottling plant production** Problems 26.4 and 26.5.

a. Make all pairwise comparisons among the mean outputs for the three machines. Use the Tukey procedure with a 95 percent family confidence coefficient. State your findings.

b. Make all pairwise comparisons among the mean outputs for the four operators assigned to machine 1. Use the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

c. Operator 4 assigned to machine 1 has relatively little experience compared to the other three operators. Estimate the contrast:

$$L = \frac{\mu_{11} + \mu_{12} + \mu_{13}}{3} - \mu_{14}$$

using a 99 percent confidence interval. Interpret your interval estimate.

26.7. Refer to **Bottling plant production** Problem 26.4. Assume that the four operators assigned to each machine were selected at random from a large number of operators.

a. How is nested design model (26.7) modified to fit this case?

b. Obtain a point estimate of the operator variance σ_β^2.

c. Test whether or not σ_β^2 equals zero; use $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Use the MLS procedure to obtain an approximate 90 percent confidence interval for σ_β^2. Interpret your confidence interval.

e. Test whether or not the mean outputs differ for the three machine models; use $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

f. Make all pairwise comparisons among the mean outputs for the three machines. Use the Tukey procedure with a 90 percent family confidence coefficient. State your findings.

g. Test the assumption that the $\beta_{j(i)}$ for all machines have the same variance σ_β^2. Use the Brown-Forsythe test (Section 18.2) with $\alpha = .01$. State the alternatives, decision rule, and conclusion.

26.8. Refer to **Bottling plant production** Problem 26.4. Assume that the four operators assigned to each machine were selected at random from a large number of operators and that the three machines were chosen at random from a large number of machines.

a. How is nested design model (26.7) modified to fit this case?

b. Obtain point estimates of the operator and machine variances σ_β^2 and σ_α^2, respectively.

c. Test whether or not σ_α^2 equals zero; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Use the MLS procedure to obtain an approximate 95 percent confidence interval for σ_β^2. Interpret your confidence interval.

e. The production engineer is interested in estimating the overall mean $\mu_{..}$ with a 95 percent confidence interval. Obtain the desired confidence interval and interpret your interval estimate.

*26.9. **Health awareness.** Three states (factor A) participated in a health awareness study. Each state independently devised a health awareness program. Three cities (factor B) within each state were selected for participation and five households within each city were randomly selected to evaluate the effectiveness of the program. All members of the selected households were interviewed before and after participation in the program and a composite index was formed for each household measuring the impact of the health awareness program. The data on health awareness follow (the larger the index, the greater the awareness).

State i:	**1**			**2**			**3**		
City j:	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Household $k = 1$:	42	26	34	47	56	68	19	18	16
$k = 2$:	56	38	51	58	43	51	36	40	28
$k = 3$:	35	42	60	39	65	49	24	27	45
$k = 4$:	40	35	29	62	70	71	12	31	30
$k = 5$:	28	53	44	65	59	57	33	23	21

a. Obtain the residuals for nested design model (26.7) with fixed factor effects and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings about the appropriateness of model (26.7)?

b. Prepare aligned residual dot plots by state. Do these plots support the assumption of constancy of the error variance? Discuss.

c. Plot the data in the format of Figure 26.3. Does it appear that any factor effects are present?

*26.10. Refer to **Health awareness** Problem 26.9. Assume that nested design model (26.7) with fixed factor effects is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not the mean awareness differs for the three states; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Test whether or not the mean awareness differs for the three cities within each state; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test? What does your conclusion imply about the awareness means for the three cities in state 1? Explain.

d. What is the family level of significance for the combined tests in parts (b) and (c) using the Bonferroni inequality? Summarize the set of conclusions reached in your tests.

*26.11. Refer to **Health awareness** Problem 26.9 and 26.10.

a. Estimate μ_{11} with a 95 percent confidence interval. Interpret your interval estimate.

b. Obtain separate confidence intervals for $\mu_{1\cdot}$, $\mu_{2\cdot}$, and $\mu_{3\cdot}$, each with a 99 percent confidence coefficient. Interpret your interval estimates.

c. Obtain confidence intervals for all pairwise comparisons among the state means. Use the Tukey procedure and a 90 percent family confidence coefficient. Summarize your findings.

d. It is desired to obtain a 95 percent confidence interval for $L = \mu_{11} - \mu_{32}$, since these two cities are of comparable size. Interpret your interval estimate.

*26.12. Refer to **Health awareness** Problem 26.9. Assume that the three cities in each state were chosen at random from all the cities in the state.

a. How is nested design model (26.7) modified to fit this case?

b. Obtain a point estimate of the city variance σ_β^2. Is there anything peculiar about the estimate here?

c. Test whether or not σ_β^2 equals zero; use $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test whether or not the mean awareness differs for the three states; use $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

e. Obtain confidence intervals for all pairwise comparisons between the state means. Use the Tukey procedure and a 90 percent family confidence coefficient. Summarize your findings.

f. Test the assumption that the $\beta_{j(i)}$ for all states have the same variance σ_β^2. Use the Hartley test (Section 18.2) with significance level $\alpha = .05$. State the alternatives, decision rule, and conclusion.

*26.13. Refer to **Health awareness** Problem 26.9. Assume that the three cities within each state and the three states were selected at random.

a. How is nested design model (26.7) modified to fit this case?

b. Obtain point estimates of the city and state variances σ_β^2 and σ_α^2, respectively.

c. Test whether or not σ_α^2 equals zero; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Use the MLS procedure to obtain an approximate 99 percent confidence interval for σ_α^2. Interpret your confidence interval.

e. Estimate the overall mean health awareness index $\mu_{\cdot\cdot}$ using a 99 percent confidence interval. Interpret your interval estimate.

26.14. **Internal control.** A large retailer operates three regional accounting centers (factor A). Center 1 employs three audit teams, while the other two centers employ two audit teams each. One function of each center is to review whether a certain internal control operates properly in the processing of payroll. Data on the percent of transactions where the internal control was found to be operating properly were requested for each team in each region for the previous two months. Three months' data were received in one case, and data for only one month in another. The arcsine transformation $Y' = 2 \arcsin \sqrt{p}$ was employed to stabilize the error variances. The transformed data follow.

Region i:	1			2		3	
Team j:	1	2	3	1	2	1	2
Month $k = 1$:	151.6	143.2	131.4	163.8	151.6	157.0	160.0
$k = 2$:	141.2	139.4	136.0	154.2		147.2	151.6
$k = 3$:	149.4						

a. Set up the full regression model for this case, analogous to the illustrative full model (26.31), using 1, −1, 0 indicator variables.

b. Fit this model and obtain the residuals. Plot the residuals against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings about the appropriateness of the model?

Refer to **Internal control** Problem 26.14. Assume that nested design model (26.7) with fixed factor effects, modified for unequal nestings and replications, is appropriate.

a. Test for region main effects using test statistic (7.27) and significance level $\alpha = .025$. State the alternatives, reduced model, decision rule, and conclusion. What is the P-value of the test?

b. Test for effects of audit teams within region using test statistic (7.27) and significance level $\alpha = .025$. State the alternatives, reduced model, decision rule, and conclusion.

c. Estimate $L = \mu_{1\cdot} - \mu_{2\cdot}$ (in transformed units) with a 98 percent confidence interval.

A student asked in class why all experiments do not make use of repeated observations since all measurement procedures are inexact to some degree. Comment.

Refer to **Questionnaire color** Problem 16.8. Suppose that the experiment was conducted by distributing the fliers to the assigned parking lots in two different weeks and noting the response rates for each week. The complete data on response rates follow.

Color i:	1 (Blue)					2 (Green)					3 (Orange)				
Lot j:	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Week $k = 1$:	28	26	31	27	35	34	29	25	31	29	31	25	27	29	28
$k = 2$:	32	23	29	24	37	33	27	22	34	25	35	28	25	25	31

a. Obtain the residuals for subsampling model (26.35) with fixed treatment effects and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings about the appropriateness of model (26.35)?

b. Test the assumption that the $\varepsilon_{j(i)}$ have the same variance σ^2 for all colors. Use the Brown-Forsythe test (Section 18.2) with significance level $\alpha = .01$. State the alternatives, decision rule, and conclusion.

Refer to **Questionnaire color** Problem 26.17. Assume that subsampling model (26.35) with fixed treatment effects is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not questionnaire color effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Test whether or not lot differences within colors are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Estimate the mean response rate for blue questionnaires with a 95 percent confidence interval.

e. Obtain point estimates of σ^2 and σ_η^2. Which variance appears to be larger here?

f. Use the MLS procedure to obtain an approximate 95 percent confidence interval for σ^2. Also obtain a 95 percent confidence interval for σ_η^2. Interpret your interval estimates.

Plant acid levels. Four plants of the same variety were randomly selected in an experiment to investigate the concentration of a particular acid. Three leaves per plant were randomly selected and three separate determinations of the acid concentration were obtained per leaf.

The data follow.

Plant i:	1			2			3			4		
Leaf j:	1	2	3	1	2	3	1	2	3	1	2	3
Determination												
$k = 1$:	11.2	16.5	18.3	14.1	19.0	11.9	15.3	19.5	16.5	7.3	8.9	11.3
$k = 2$:	11.6	16.8	18.7	13.8	18.5	12.4	15.9	20.1	17.2	7.8	9.4	10.9
$k = 3$:	12.0	16.1	19.0	14.2	18.2	12.0	16.0	19.3	16.9	7.0	9.3	10.5

Obtain the residuals for three-stage subsampling model (26.43) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings about the appropriateness of model (26.43)?

*26.20. Refer to **Plant acid levels** Problem 26.19. Assume that three-stage subsampling model (26.43) is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not there are variations in mean concentration levels between plants; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Test whether or not there are variations in mean concentration levels between leaves of the same plant; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Estimate the overall mean concentration in all plants of the variety; use a 95 percent confidence interval.

e. Obtain point estimates of σ_τ^2, σ^2, and σ_η^2. Which component of variance appears to be most important in the total variance σ_Y^2?

f. Use the MLS procedure to obtain an approximate 90 percent confidence interval for σ_τ^2. Does the experiment provide a precise estimate of this variance component?

26.21. **Chemical consistency.** A chemical company wished to study the consistency of the strength of one of its liquid chemical products. The product is made in batches in large vats and then is barreled. The barrels are subsequently stored for a period of time in a warehouse. To examine the consistency of the strength of the chemical, an analyst randomly selected five different batches of the product from the warehouse and then selected four barrels per batch at random. Three determinations per barrel were made. The data on strength follow.

Batch i:	1				2				...	5			
Barrel j:	1	2	3	4	1	2	3	4		1	2	3	4
Determination													
$k = 1$:	2.3	2.5	2.6	2.4	2.8	2.7	2.6	2.4		3.6	3.8	3.7	3.9
$k = 2$:	2.1	2.3	2.4	2.6	2.9	2.5	2.6	2.8		3.7	3.8	3.5	3.5
$k = 3$:	2.0	2.5	2.7	2.3	2.6	2.8	2.8	2.6	...	3.4	3.5	3.5	3.7

a. Obtain the residuals for three-stage subsampling model (26.43) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings about the appropriateness of model (26.43)?

b. Test the assumption that the $\varepsilon_{j(i)}$ have the same variance σ^2 for all batches. Use the Hartley test (Section 18.2) with significance level $\alpha = .01$. State the alternatives, decision rule, and conclusion.

26.22. Refer to **Chemical consistency** Problem 26.21. Assume that three-stage subsampling model (26.43) is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not there are variations in mean strength between batches; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Test whether or not there are variations in mean strength between barrels within batches; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Estimate the overall mean strength of the chemical using a 99 percent confidence interval.

e. Obtain point estimates of σ_τ^2, σ^2, and σ_η^2. Which component of variance appears to be most important in the total variance σ_Y^2?

f. Use the MLS procedure to obtain an approximate 95 percent confidence interval for σ_τ^2. Does the experiment provide a precise estimate of this variance component?

Exercises

26.23. Derive (26.13) by squaring (26.12) and summing over all observations.

26.24. Derive (26.16) for a balanced nested two-factor design.

26.25. Consider a balanced nested two-factor design with factor A having fixed effects and factor B (nested within factor A) having random effects.

a. Derive $\sigma^2\{\bar{Y}_{i\cdot\cdot}\}$ and $\sigma^2\{\bar{Y}_{\cdot\cdot\cdot}\}$.

b. Find an unbiased point estimator of σ_β^2.

26.26. Show that $\sigma^2\{\bar{Y}_{i\cdot\cdot}\} = (\sigma_\eta^2 + m\sigma^2)/nm$ for subsampling model (26.35) with fixed treatment effects.

26.27. Derive variance (26.45) for three-stage subsampling model (26.43). Using the expected mean squares in Table 26.8, show that the estimated variance (26.46) is an unbiased estimator of variance (26.45).

26.28. Use (26.52) and the fact that this estimated variance is unbiased to find $\sigma^2\{\bar{Y}_{1j\cdot\cdot} - \bar{Y}_{2j\cdot\cdot}\}$ for ANOVA model (26.48). What is the approximate number of degrees of freedom associated with the estimated variance?

Projects

26.29. Refer to the **Drug effect experiment** data set in Appendix C.12. Consider only Part I of the study and dosage level 4; i.e., include only observations for which variable 2 equals 1 and variable 5 equals 4. Assume that initial lever press rate (factor A) has fixed effects and that rats are a second factor (factor D) with random effects.

a. State the appropriate model for this nested two-factor study.

b. Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings about the appropriateness of your model?

26.30. Refer to the **Drug effect experiment** data set in Appendix C.12 and Project 26.29. Assume that nested design model (26.7), with $\beta_{j(i)}$ and ε_{ijk} random, is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not the mean lever press rate differs for the three initial rate groups; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Test whether or not the mean lever press rate differs for the rats within the initial rate groups; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test? What does your conclusion imply about the four rats in the slow initial rate group?

d. Make all pairwise comparisons between the mean lever press rates for the three initial rate groups. Use the Tukey procedure with a 90 percent family confidence coefficient.

e. Obtain an approximate 90 percent confidence interval for the between-rats variance, using the MLS procedure. Interpret your interval estimate.

26.31. Refer to the **Drug effect experiment** data set in Appendix C.12. Consider only Part II of the study and dosage level 3; i.e., include only observations for which variable 2 equals 2 and variable 5 equals 3. Assume that the initial lever press rate groups are the treatments with fixed effects, and that the rats are the experimental units with two observations for each experimental unit.

a. State the appropriate model for this single-factor study with subsampling.

b. Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings about the appropriateness of your model?

c. Test the assumption that the $\varepsilon_{j(i)}$ have the same variance σ^2 for all lever press rates. Use the Brown-Forsythe test (Section 18.2) with $\alpha = .01$. State the alternatives, decision rule, and conclusion.

26.32. Refer to the **Drug effect experiment** data set in Appendix C.12 and Project 26.31. Assume that single-factor subsampling model (26.35) with fixed treatment effects is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not the mean lever press rate differs for the three initial rate groups; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Test whether or not differences in the mean lever press rate between rats are present; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Make all pairwise comparisons between the mean lever press rates for the three initial rate groups. Use the Tukey procedure with a 95 percent family confidence coefficient. Summarize your findings.

e. Obtain interval estimates for σ^2 and σ_η^2, with confidence coefficient .90 for each. Interpret your confidence intervals. Which variance component appears to be larger?

Chapter 27

Repeated Measures and Related Designs

In this chapter we take up repeated measures designs—designs that are widely used in the behavioral and life sciences. We begin by considering some basic elements of repeated measures designs. We then take up single-factor repeated measures designs, after which we consider two-factor experiments with repeated measures on both one factor and on two factors. We conclude this chapter with an introduction to split-plot designs, which include two-factor repeated measures designs with repeated measures on one factor.

27.1 Elements of Repeated Measures Designs

Description of Designs

Repeated measures designs utilize the same subject (person, store, plant, test market, etc.) for each of the treatments under study. The subject therefore serves as a block, and the experimental units within a block may be viewed as the different occasions when a treatment is applied to the subject. A repeated measures study may involve several treatments or only a single treatment that is evaluated at different points in time. Subjects used in repeated measures studies in the behavioral and life sciences include persons, households, observers, and experimental animals. At other times the subjects in repeated measures designs are stores, test markets, cities, and plants. We shall refer to all of these study units used in repeated measures designs as *subjects*.

Three examples of repeated measures designs follow.

1. Fifteen test markets are to be used to study each of two different advertising campaigns. In each test market, the order of the two campaigns will be randomized, with a sufficient time lapse between the two campaigns so that the effects of the initial campaign will not carry over into the second campaign. The subjects in this study are the test markets.

2. Two hundred persons who have persistent migraine headaches are each to be given two different drugs and a placebo, for two weeks each, with the order of the drugs randomized for each person. The subjects in the study are the persons with migraine headaches.

3. In a weight loss study, 100 overweight persons are to be given the same diet and their weights measured at the end of each week for 12 weeks to assess the weight loss over

time. Here the subjects are the overweight persons, who are observed repeatedly to provide information about the effects of a single treatment over time.

Each of these studies involves a *repeated measures design* because the same subject is measured repeatedly. This key characteristic distinguishes this type of design from the designs considered earlier.

Advantages and Disadvantages

A principal advantage of repeated measures designs is that they provide good precision for comparing treatments because all sources of variability between subjects are excluded from the experimental error. Only variation within subjects enters the experimental error, since any two treatments can be compared directly for each subject. Thus, one may view the subjects as serving as their own controls. Another advantage of a repeated measures design is that it economizes on subjects. This is particularly important when only a few subjects (e.g., stores, plants, test markets) can be utilized for the experiment. Also, when interest is in the effects of a treatment over time, as when the shape of the learning curve for a new process operation is to be studied, it is usually desirable to observe the same subject at different points in time rather than observing different subjects at the specified points in time.

Repeated measures designs have a serious potential disadvantage, however, namely, that there may be several types of interference. One type of interference is an *order effect*, which is connected with the position in the treatment order. For instance, in evaluating five different advertisements, subjects may tend to give higher (or lower) ratings for advertisements shown toward the end of the sequence than at the beginning. Another type of interference is connected with the preceding treatment or treatments. For instance, in evaluating five different soup recipes, a bland recipe may get a higher (or lower) rating when preceded by a highly spiced recipe than when preceded by a blander recipe. This type of interference is called a *carryover effect*.

Various steps can be taken to minimize the danger of interference effects. Randomization of the treatment orders for each subject independently will make it more reasonable to analyze the data as if the error terms are independent. Allowing sufficient time between treatments is often an effective means of reducing carryover effects. It may be desirable at times to balance the order of treatment presentations and sometimes even the number of times each treatment is preceded by any other treatment. Latin square designs and crossover designs (discussed in Chapter 28) are helpful to this end.

How to Randomize

The randomization of the order of the treatments assigned to a subject is straightforward. For each subject, a random permutation is used to define the treatment order, and independent permutations are selected for the different subjects.

Comment

Designs with repeated measures, discussed here, need to be distinguished from designs with repeated observations, discussed in Section 26.7. In repeated measures designs. several or all of the treatments are applied to the same subject. Designs with repeated observations, on the other hand, are designs where several observations on the response variable are made for a given treatment applied to an experimental unit. It is possible to develop a repeated measures design with repeated observations, as when a given subject is exposed to each of the treatments under study and a number of observations are made at the end of each treatment application. ■

27.2 Single-Factor Experiments with Repeated Measures on All Treatments

We first consider repeated measures designs where the treatments are based on a single factor, as in the examples in Section 27.1. Almost always, the subjects in repeated measures designs (persons, stores, test markets, experimental animals) are viewed as a random sample from a population. Hence, *in all of the models for repeated measures designs to be presented in this chapter, the effects of subjects will be viewed as random.*

Figure 27.1 contains the layout for a single-factor experiment with repeated measures on all treatments. Here, there are five subjects and four treatments, with the order of treatments independently randomized for each subject. Notice that this layout corresponds to the one in Figure 21.1 for a randomized complete block design. Indeed, as we shall see next, the models for single-factor repeated measures designs are formally the same as the ones for randomized block designs, with blocks now considered to be subjects.

Model

When treatment effects are fixed, a model often appropriate for a single-factor repeated measures design is the following additive model:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \varepsilon_{ij} \tag{27.1}$$

where:

$\mu_{..}$ is a constant
ρ_i are independent $N(0, \sigma_\rho^2)$
τ_j are constants subject to $\sum \tau_j = 0$
ε_{ij} are independent $N(0, \sigma^2)$
ρ_i and ε_{ij} are independent
$i = 1, \ldots, s;\ j = 1, \ldots, r$

FIGURE 27.1 Layout for Single-Factor Repeated Measures Design ($s = 5, r = 4$).

	Treatment Order			
	1	2	3	4
Subject 1	T_4	T_3	T_2	T_1
2	T_3	T_4	T_1	T_2
3	T_4	T_3	T_1	T_2
4	T_2	T_1	T_4	T_3
5	T_1	T_2	T_4	T_3

Note that repeated measures model (27.1) is identical to randomized block model (25.67) with random block effects, except that $n_b = s$.

Hence, we know from Section 25.5 that repeated measures model (27.1) assumes the following about the observations Y_{ij}:

$$E\{Y_{ij}\} = \mu_{..} + \tau_j \tag{27.2a}$$

$$\sigma^2\{Y_{ij}\} = \sigma_Y^2 = \sigma_\rho^2 + \sigma^2 \tag{27.2b}$$

$$\sigma\{Y_{ij}, Y_{ij'}\} = \sigma_\rho^2 = \omega\sigma_Y^2 \qquad j \neq j' \tag{27.2c}$$

$$\sigma\{Y_{ij}, Y_{i'j'}\} = 0 \qquad i \neq i' \tag{27.2d}$$

where ω is the coefficient of correlation between any two observations for the same subject:

$$\omega = \frac{\sigma_\rho^2}{\sigma_Y^2} \tag{27.2e}$$

Thus, repeated measures model (27.1) assumes that in advance of the random trials, any two treatment observations Y_{ij} and $Y_{ij'}$ for a given subject are correlated in the same fashion for all subjects. This key assumption implies, as we saw in (25.71), that the variance-covariance matrix of the observations Y_{ij} for any given subject has compound symmetry. Any two observations from different subjects in advance of the random trials are independent according to model (27.1).

Equally important, we know from Chapter 25 that repeated measures model (27.1) assumes that, once the subjects have been selected, any two observations for a given subject are independent. Thus, model (27.1) assumes that there are no interference effects in the repeated measures study, such as order effects or carryover effects from one treatment to the next.

Comment

If interaction effects between subjects and treatments are present, interaction model (25.74) can be employed. As we noted in Chapter 25, both the additive and interaction models lead to the same procedures for making inferences about the treatment effects. ■

Analysis of Variance and Tests

Since repeated measures model (27.1) is the same as randomized complete block model (25.67), the analysis of variance and the test for treatment effects will be the same as before.

Analysis of Variance. The ANOVA sums of squares for repeated measures model (27.1) are the same as in (21.6), but the names of two of the sums of squares are usually changed for repeated measures applications. The sum of squares for blocks in (21.6a) will now be called the *sum of squares for subjects,* and the interaction sum of squares between blocks and treatments in (21.6c) will now be called the *interaction sum of squares between treatments and subjects*. These two sums of squares will be denoted, respectively, by *SSS* and *SSTR.S*. Thus, the analysis of variance decomposition for single-factor repeated measures model (27.1) is:

$$SSTO = SSS + SSTR + SSTR.S \tag{27.3}$$

TABLE 27.1 ANOVA Table for Single-Factor Repeated Measures Design—ANOVA Model (27.1) with Subject Effects Random and Treatment Effects Fixed.

Source of Variation	*SS*	*df*	*MS*	$E\{MS\}$
Subjects	*SSS*	$s-1$	*MSS*	$\sigma^2 + r\sigma_p^2$
Treatments	*SSTR*	$r-1$	*MSTR*	$\sigma^2 + s\frac{\sum \tau_j^2}{r-1}$
Error	*SSTR.S*	$(r-1)(s-1)$	*MSTR.S*	σ^2
Total	*SSTO*	$sr-1$		

where:

$$SSTO = \sum_i \sum_j (Y_{ij} - \bar{Y}_{\cdot\cdot})^2 \tag{27.3a}$$

$$SSS = r \sum_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 \tag{27.3b}$$

$$SSTR = s \sum_j (\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2 \tag{27.3c}$$

$$SSTR.S = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot})^2 \tag{27.3d}$$

Note that no error sum of squares is present because there are no replications here.

Table 27.1 contains the analysis of variance table for repeated measures model (27.1). It is the same as the ANOVA table in Table 25.8 for additive randomized block model (25.67), except for the change in notation. Note again that in the absence of interactions between treatments and subjects, the interaction mean square *MSTR.S* is an unbiased estimator of the error variance σ^2.

Comment

In repeated measures studies, *SSTR* and *SSTR.S* are sometimes combined into a *within-subjects sum of squares SSW:*

$$SSW = SSTR + SSTR.S \tag{27.4}$$

which can be shown to equal:

$$SSW = \sum_i \sum_j (Y_{ij} - \bar{Y}_{i\cdot})^2 \tag{27.4a}$$

Hence, the ANOVA decomposition in (27.3) can also be expressed as follows:

$$SSTO = \underbrace{SSS}_{\substack{\text{Between-}\\\text{subjects}\\\text{variability}}} + \underbrace{SSW}_{\substack{\text{Within-}\\\text{subjects}\\\text{variability}}} \tag{27.5}$$

■

Test for Treatment Effects. As the $E\{MS\}$ column in Table 27.1 indicates, the appropriate statistic for the test on treatment effects:

$$H_0\text{: all } \tau_j = 0$$
$$H_a\text{: not all } \tau_j \text{ equal zero} \tag{27.6a}$$

is:

$$F^* = \frac{MSTR}{MSTR.S} \tag{27.6b}$$

When H_0 holds, F^* follows the F distribution, and the decision rule for controlling the Type I error at α is:

$$\text{If } F^* \leq F[1-\alpha; r-1, (r-1)(s-1)], \text{ conclude } H_0$$
$$\text{If } F^* > F[1-\alpha; r-1, (r-1)(s-1)], \text{ conclude } H_a \tag{27.6c}$$

Example

In a wine-judging competition, four Chardonnay wines of the same vintage were judged by six experienced judges. Each judge tasted the wines in a blind fashion, i.e., without knowing their identities. The order of the wine presentation was randomized independently for each judge. To reduce carryover and other interference effects, the judges did not drink the wines and rinsed their mouths thoroughly between tastings. Each wine was scored on a 40-point scale; the higher the score, the greater is the excellence of the wine. The data for this competition are presented in Table 27.2. A plot of the wine scores for each judge is shown in Figure 27.2. We see that there are some distinct differences in ratings between judges but that the ratings for wines 3 and 4 are consistently best and for wine 1 generally worst. We also see that the rating curves for the judges do not appear to exhibit substantial departures from being parallel. Hence, an additive model appears to be appropriate.

The six judges are considered to be a random sample from the population of possible judges, while the four wines tasted are of interest in themselves. Hence, single-factor repeated measures model (27.1) was expected to be appropriate, with the effects of subjects (judges) considered random and the effects of treatments (wines) considered fixed. As

TABLE 27.2 Data—Wine-Judging Example (ratings on a scale of 0 to 40).

Judge	Wine (j)				
i	1	2	3	4	$\bar{Y}_{i\cdot}$
1	20	24	28	28	25
2	15	18	23	24	20
3	18	19	24	23	21
4	26	26	30	30	28
5	22	24	28	26	25
6	19	21	27	25	23
$\bar{Y}_{\cdot j}$	20.00	22.00	26.67	26.00	$23.67 = \bar{Y}_{\cdot\cdot}$

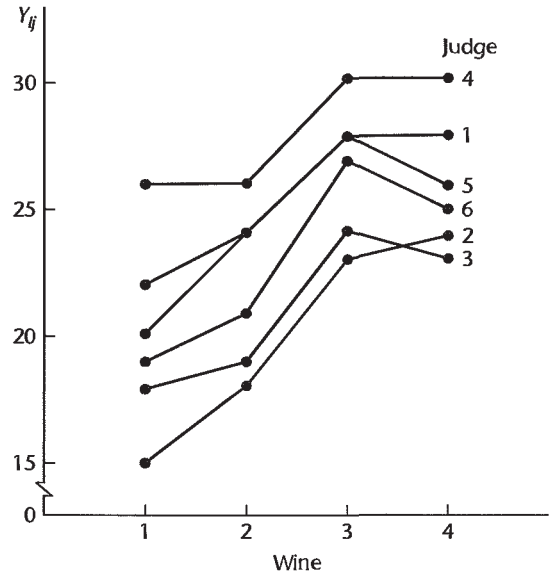


FIGURE 27.2 Plot of Wine Scores for Each Judge—Wine-Judging Example.

FIGURE 27.3 MINITAB ANOVA Table for Single-Factor Repeated Measures Design—Wine-Judging Example.

Factor	Type	Levels	Values					
Judge	random	6	1	2	3	4	5	6
Wine	fixed	4	1	2	3	4		

Analysis of Variance for Rating

Source	DF	SS	MS	F	P
Judge	5	173.333	34.667	32.50	0.000
Wine	3	184.000	61.333	57.50	0.000
Error	15	16.000	1.067		
Total	23	373.333			

we shall see later, additional diagnostic analysis supports the appropriateness of ANOVA model (27.1).

Figure 27.3 contains MINITAB ANOVA output for the wine-judging data in Table 27.2. To test for treatment effects:

$$H_0\colon \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0$$

$$H_a\colon \text{not all } \tau_j \text{ equal zero}$$

we use the results of Table 27.3:

$$F^* = \frac{MSTR}{MSTR.S} = \frac{61.333}{1.067} = 57.5$$

For level of significance $\alpha = .01$, we require $F(.99; 3, 15) = 5.42$. Since $F^* = 57.5 > 5.42$, we conclude H_a, that the mean wine ratings for the four wines differ. The P-value for this test is 0+.

TABLE 27.3 Estimated Within-Subjects Variance-Covariance Matrix between Treatment Observations—Wine-Judging Example.

		j'			
		1	2	3	4
j	1	14.000	11.000	9.200	8.200
	2		10.000	8.200	7.600
	3			7.067	6.200
	4				6.800

Comments

1. As we noted in Chapter 25 (in Comment 2 on p. 1065), a conservative test for treatment effects should be used if the assumptions of compound symmetry in repeated measures model (27.1) are not met (i.e., if either the variances of the observations for different treatments for a given subject are not the same for all subjects or if the correlations between any two treatment observations for a given subject are not the same for all treatment pairs and for all subjects). In repeated measures studies, the compound symmetry assumption will be violated, for instance, if repeated responses over time are more highly correlated for observations closer together than for observations further apart in time.

2. When the treatment effects are random, test statistic (27.6b) and decision rule (27.6c) are still appropriate for testing treatment effects.

3. The efficiency of the repeated measures design in the wine-judging example, relative to a completely randomized design where each judge is used to assess a single wine, can be measured by means of (21.14). Using the results in Figure 27.3 with $n_b = s$, we obtain:

$$\hat{E} = \frac{(s-1)MSS + s(r-1)MSTR.S}{(sr-1)MSTR.S} = \frac{5(34.667) + 6(3)(1.067)}{23(1.067)} = 7.85$$

Thus, almost eight times as many replications per treatment would have been required with a completely randomized design in which each judge rates a single wine as in the repeated measures design to achieve the same precision for any estimated contrast.

4. When a single-factor repeated measures design involves $r = 2$ treatments, the F^* statistic in (27.6b) is equivalent to the two-sided t test for paired observations based on test statistic (A.69).

5. Occasionally, a formal test for subject effects is desired:

$$H_0: \sigma_\rho^2 = 0$$
$$H_a: \sigma_\rho^2 > 0$$

Table 27.1 indicates that the appropriate test statistic for repeated measures model (27.1) is $F^* = MSS/MSTR.S$. ■

Evaluation of Appropriateness of Repeated Measures Model

Since repeated measures model (27.1) is equivalent to randomized block model (25.67), the earlier discussion on diagnostics for randomized block models is entirely applicable here. In particular, a plot of the responses Y_{ij} by subject, as in Figure 27.2, can be examined for indications of serious lack of parallelism, which would suggest that additive model (27.1) may not be appropriate.

Residual sequence plots by subject can be helpful for studying constancy of the error variance and presence of interference effects. The residuals for repeated measures models (27.1) are the same as in (21.5):

$$e_{ij} = Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot} \tag{27.7}$$

A normal probability plot of the estimated residuals in (27.7) can be helpful for evaluating whether the residuals are normally distributed.

In addition to these graphic diagnostics, the estimated within-subjects variance-covariance and correlation matrices for the treatment observations Y_{ij} can be examined for appropriateness of the repeated measures model. A typical entry in the variance-covariance matrix is the estimated within-subjects covariance between observations for treatments j and j':

$$\frac{\sum_{i=1}^{s}(Y_{ij} - \bar{Y}_{\cdot j})(Y_{ij'} - \bar{Y}_{\cdot j'})}{s-1} \tag{27.8}$$

The estimated within-subjects variance-covariance matrix should show variances of the same order of magnitude, and all of the covariances should be of similar magnitude. Of course, estimated variances and covariances tend to be subject to large sampling errors unless the sample sizes are very large. Hence, moderate differences in variances and covariances should be viewed as likely to be the result of sampling errors.

The estimated correlation matrix should show approximately similar coefficients of correlation between pairs of treatment observations within a subject.

Finally, the Tukey test described in Section 20.2 can be conducted to examine the appropriateness of the additive model. This test will need to be interpreted here as conditional on the subjects actually used in the repeated measures study.

Example

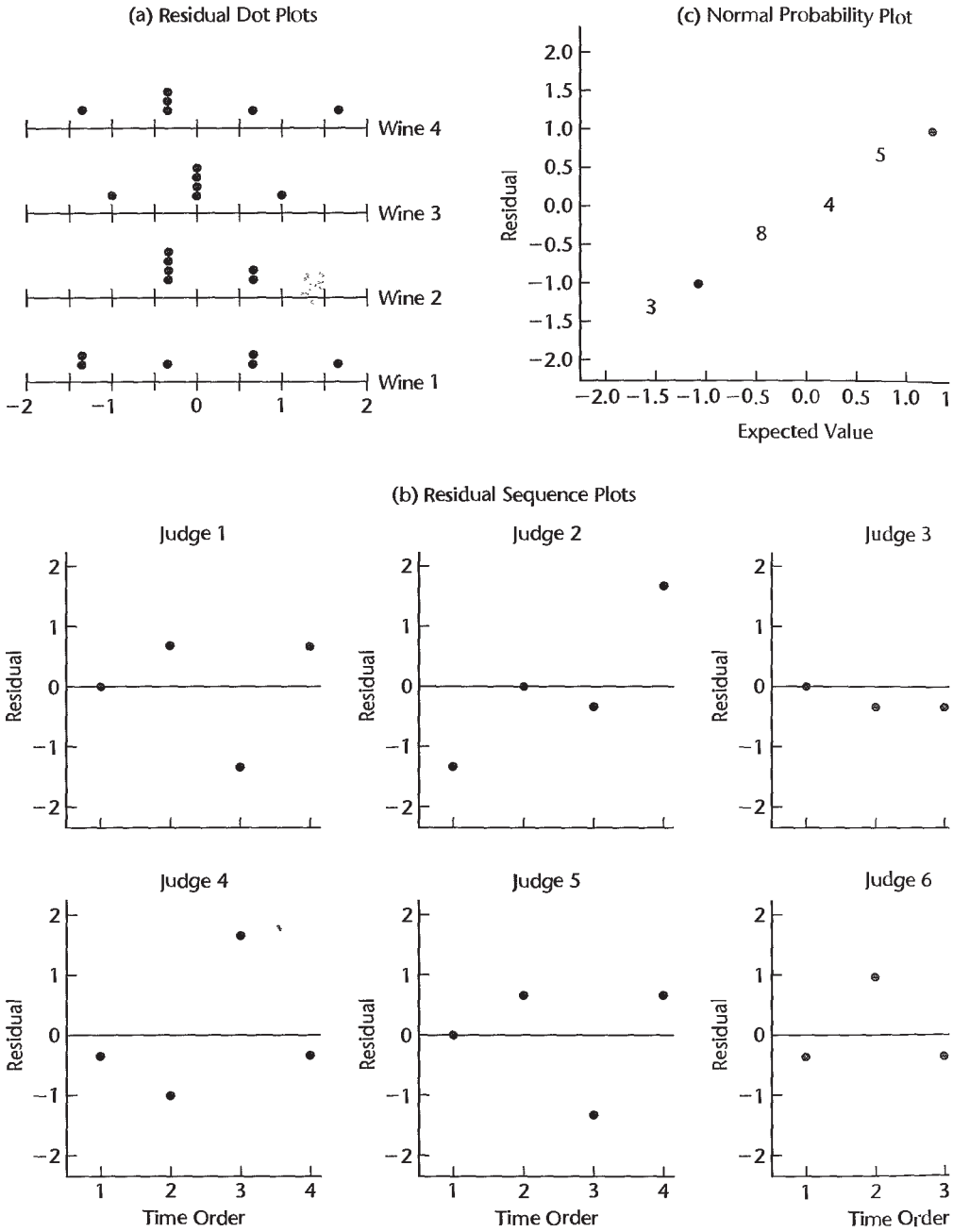
For the wine-judging example, the residuals were obtained from (27.7), and are presented in Figure 27.4a in SAS/GRAPH aligned dot plots by wine. These plots support the assumption of constant error variance. Figure 27.4b presents residual sequence plots for each judge, where the residuals are plotted in the order in which the wines were tasted by the judge. These plots do not indicate any correlations of the error terms within a judge, and thus suggest that no interference effects are present. Finally, a normal probability plot of the residuals is presented in Figure 27.4c. This plot shows evidence of the effects of the rounded nature of the data, but does not suggest any major departure from normality. The correlation between the ordered residuals and their expected values under normality is .993, which also suggests that lack of normality is not a problem here.

Table 27.3 presents the estimated within-subjects variance-covariance matrix for the treatment observations. The differences found there could easily arise from sampling errors.

As we noted earlier, the plot of the responses by subject in Figure 27.2 also supports the appropriateness of model (27.1), since the plots for the judges are reasonably parallel. Thus, there is no indication of interactions between subjects and treatments.

On the basis of these and other diagnostics, it was concluded that repeated measures model (27.1) is reasonably appropriate for the data in the wine-judging example.

FIGURE 27.4 SAS/GRAPH Diagnostic Residual Plots—Wine-Judging Example.



Analysis of Treatment Effects

The analysis of treatment effects for single-factor repeated measures model (27.1) proceeds in exactly the same fashion as described in Section 21.5 for randomized block designs with fixed treatment effects. The multiples in (21.9) for setting up confidence intervals are applicable here as they stand. The mean square used in estimating the variance of the estimated contrast is still the interaction mean square, which is now denoted by *MSTR.S*. We shall illustrate the estimation procedures by an example.

Example

In the wine-judging example, it was desired to compare all treatment means $\mu_{\cdot j}$ pairwise, with a 95 percent family confidence coefficient. Here $\mu_{\cdot j}$ is the mean rating of wine j averaged over judges. The Tukey procedure was utilized for this purpose. Using (17.30) with *MSE* replaced by *MSTR.S* and the estimated pairwise difference denoted by $\hat{L}$, we obtain using the results in Figure 27.3:

$$s^2\{\hat{L}\} = MSTR.S\left(\frac{1}{s} + \frac{1}{s}\right) = 1.067\left(\frac{2}{6}\right) = .3557$$

Using (21.9b), we find for a 95 percent family confidence coefficient:

$$T = \frac{1}{\sqrt{2}}q(.95; 4, 15) = \frac{1}{\sqrt{2}}(4.08) = 2.885$$

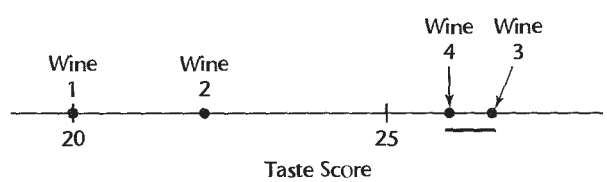
Hence:

$$Ts\{\hat{L}\} = 2.885\sqrt{.3557} = 1.72$$

Thus we obtain for the pairwise comparisons (see Table 27.2 for the $\bar{Y}_{\cdot j}$):

$$\begin{aligned}
-2.39 &= (26.00 - 26.67) - 1.72 \leq \mu_{\cdot 4} - \mu_{\cdot 3} \leq (26.00 - 26.67) + 1.72 = 1.05 \\
2.28 &= (26.00 - 22.00) - 1.72 \leq \mu_{\cdot 4} - \mu_{\cdot 2} \leq (26.00 - 22.00) + 1.72 = 5.72 \\
4.28 &= (26.00 - 20.00) - 1.72 \leq \mu_{\cdot 4} - \mu_{\cdot 1} \leq (26.00 - 20.00) + 1.72 = 7.72 \\
2.95 &= (26.67 - 22.00) - 1.72 \leq \mu_{\cdot 3} - \mu_{\cdot 2} \leq (26.67 - 22.00) + 1.72 = 6.39 \\
4.95 &= (26.67 - 20.00) - 1.72 \leq \mu_{\cdot 3} - \mu_{\cdot 1} \leq (26.67 - 20.00) + 1.72 = 8.39 \\
.28 &= (22.00 - 20.00) - 1.72 \leq \mu_{\cdot 2} - \mu_{\cdot 1} \leq (22.00 - 20.00) + 1.72 = 3.72
\end{aligned}$$

We display these results graphically as follows:



We conclude from these pairwise comparisons that wines 3 and 4 are judged best, and do not differ significantly from each other. Wines 1 and 2 are judged to be inferior to wines 3 and 4, with wine 1 receiving a mean rating significantly lower than that for wine 2. The family confidence coefficient of .95 applies to the entire set of comparisons.

TABLE 27.4 Ranked Data for Coffee Sweeteners in a Repeated Measures Design—Coffee Sweeteners Example.

Subject i	Sweetener (j) A	B	C	D	E
1	5	1	2	4	3
2	4	2	1	5	3
3	3	2	1	4	5
4	5	2	3	4	1
5	4	1	2	3	5
6	4	1	3	5	2
$\bar{R}_{\cdot j}$	4.17	1.50	2.00	4.17	3.17

Comment

When the treatments are time order positions, as when process rework is observed for a new manufacturing process at periodic intervals, the nature of the time effect may be analyzed by developing an appropriate regression model. ■

Ranked Data

In repeated measures studies, the observations are frequently ranks, as when a number of tasters are each asked to rank recipes or when several university admissions officers are each asked to rank applicants for admission. When the data in a repeated measures study are ranks, the nonparametric rank F test described in Comment 3 on page 900 may be used for testing whether the treatment means are equal. No new principles are involved, so we shall proceed directly to an example.

Example

Six subjects were each asked to rank five coffee sweeteners according to their taste preferences, with rank 5 assigned to the most preferred sweetener. The data are presented in Table 27.4 and suggest that a sweetener effect may be present. For example, no judge ranked sweetener B higher than 2 (not preferred).

Test statistic (21.7b) for the ranked data here is:

$$F_R^* = \frac{9.00}{1.20} = 7.5$$

For level of significance $\alpha = .05$, we need $F(.95; 4, 20) = 2.87$. Since $F_R^* = 7.5 > 2.87$, we conclude that the five sweeteners are not equally liked. The P-value of the test is .0007.

Multiple Pairwise Testing Procedure

Just as in the case of the rank F test for single-factor studies (Section 18.7), we can use a large-sample testing analog of the Bonferroni pairwise comparison procedure to obtain information about the comparative magnitudes of the treatment means for repeated measures designs when the rank F test (or the Friedman test) indicates that the treatment means differ. Testing limits for all $g = r(r-1)/2$ pairwise comparisons using the mean ranks $\bar{R}_{\cdot j}$ are set up as follows for family level of significance α:

$$\bar{R}_{\cdot i} - \bar{R}_{\cdot i'} \pm B\left[\frac{r(r+1)}{6s}\right]^{1/2} \tag{27.9}$$

where:

$$B = z(1 - \alpha/2g) \tag{27.9a}$$

$$g = \frac{r(r-1)}{2} \tag{27.9b}$$

If the testing limits include zero, we conclude that the corresponding treatment means $\mu_{\cdot j}$ and $\mu_{\cdot j'}$ do not differ. If the testing limits do not include zero, we conclude that the two corresponding treatment means differ. We can then set up groups of treatments whose means do not differ according to this simultaneous testing procedure.

Example

We now wish to make all pairwise tests by means of (27.9) with family level of significance $\alpha = .20$ for the coffee sweeteners example. For $r = 5$, we have $g = 5(4)/2 = 10$ and obtain:

$$B = z[1 - .20/2(10)] = z(.99) = 2.326$$

Thus, the right term in (27.9) for $s = 6$ and $r = 5$ is:

$$B\left[\frac{r(r+1)}{6s}\right]^{1/2} = 2.326\left[\frac{5(6)}{6(6)}\right]^{1/2} = 2.12$$

We note from Table 27.4 that the pairs of mean ranks whose difference does not exceed 2.12 are (B, C), (B, E), (C, E), (A, E), (D, E), and (A, D). Hence, we can set up two groups, within which the treatment means do not differ:

Group 1		Group 2	
Sweetener B	$\bar{R}_{\cdot 2} = 1.50$	Sweetener E	$\bar{R}_{\cdot 5} = 3.17$
Sweetener C	$\bar{R}_{\cdot 3} = 2.00$	Sweetener A	$\bar{R}_{\cdot 1} = 4.17$
Sweetener E	$\bar{R}_{\cdot 5} = 3.17$	Sweetener D	$\bar{R}_{\cdot 4} = 4.17$

Thus, we conclude with family level of significance of .20 that sweeteners A and D are preferred to sweeteners B and C, and that it is not clear whether sweetener E belongs in the preferred group or in the other group.

Comments

1. The rank F test can also be used for repeated measures designs where the observations are not ranked, in case the distribution of the error terms departs far from normality. Ranks of the observations Y_{lj} are then assigned within each subject, and the rank F test is carried out in the usual manner.

2. The test statistic F_R^* is related to Kendall's coefficient of concordance W in the following way:

$$W = \frac{F_R^*}{F_R^* + n - 1} \tag{27.10}$$

The coefficient of concordance W is a measure of the agreement of the rankings of the s subjects. It equals 1 if there is perfect agreement, and equals 0 if there is no agreement, that is, if all treatments receive the same mean ranking. For the coffee sweeteners example in Table 27.4, the coefficient of concordance W is:

$$W = \frac{7.5}{7.5 + 6 - 1} = .60$$

This measure indicates that a fair amount of agreement exists between the subjects. ■

27.3 Two-Factor Experiments with Repeated Measures on One Factor

Description of Design

In many two-factor studies, repeated measures can only be made on one of the two factors. Consider, for instance, an experimenter who wished to study the effects of two types of incentives (factor A) on a person's ability to solve problems. The researcher also wanted to study two types of problems (factor B)—abstract and concrete problems. Each experimental subject could be asked to do each type of problem, but could not be exposed to more than one type of incentive stimulus because of potential interference effects. Thus, the design the experimenter utilized may be represented schematically as shown in Figure 27.5.

In a two-factor experiment with repeated measures on one factor, two randomizations generally need to be employed. First, the level of the nonrepeated factor (A, in Figure 27.5) needs to be randomly assigned to the subjects. Second, the order of the levels of the repeated factor (B, in Figure 27.5) needs to be randomized independently for all subjects.

Since s subjects are randomly assigned incentive stimulus A_1 and s subjects are randomly assigned incentive stimulus A_2, as far as factor A is concerned the experiment is a completely randomized one. On the other hand, as far as factor B (type of problem) is concerned, each subject is a block. Thus, for factor B, the experiment is a randomized complete block design, with block effects random. We call this experimental design a *two-factor experiment with repeated measures on factor B*.

In the experiment depicted in Figure 27.5, comparisons between factor A level means involve differences between groups of subjects as well as differences associated with the two factor A levels. On the other hand, comparisons between factor B level means at the same level of factor A are based on the same subject, and hence only involve differences associated with the two factor B levels. Thus, for these latter comparisons, each subject serves as its own control. The main effects of factor A are therefore said to be confounded

FIGURE 27.5 Layout for Two-Factor Design with Random Assignments of Factor A Level to Subjects and Repeated Measures on Factor B.

		Treatment Order	
Incentive Stimulus	Subject	1	2
A_1	1	A_1B_1	A_1B_2
	⋮	⋮	
	s	A_1B_1	A_1B_2
A_2	$s+1$	A_2B_2	A_2B_1
	⋮	⋮	
	$2s$	A_2B_1	A_2B_2

with differences between groups of subjects, whereas the main effects of factor B are free of such confounding. It is for this reason that tests on factor B main effects will generally be more sensitive than tests on the main effects for factor A.

Comments

1. A two-factor experiment with repeated measures on one factor may be viewed as an incomplete block design. With reference to the repeated measures design in Figure 27.5, there are four treatments (A_1B_1, A_1B_2, A_2B_1, and A_2B_2) and one-half of the blocks (subjects) contain treatments A_1B_1 and A_1B_2 while the other half of the blocks contain treatments A_2B_1 and A_2B_2.

2. When the factor on which repeated measures are taken is time, randomization of the levels of the repeated factor is impossible. Consider, for instance, a study of two different advertising campaigns in which the effect on sales is to be measured in 10 test markets during four consecutive months. Here, the only randomization required is for assigning the advertising campaigns to the test markets. Similarly, when the nonrepeated factor is a characteristic of the subject, such as age of subject, no randomization is involved for that factor. ■

Model

The development of a model for a two-factor experiment with repeated measures on one factor is only a little more complex than for earlier cases. As before, we shall develop the model for random subject effects and fixed factor A and factor B effects. Let, as usual, α_j and β_k denote the factor A and factor B main effects, respectively, $(\alpha\beta)_{jk}$ the AB interaction effect, and ρ the subject (block) main effect. We do need to recognize, however, that the subject effect in this design is nested within factor A. Therefore, we will denote this effect by $\rho_{i(j)}$. As before, we assume that there are no interactions between treatments and subjects, although this condition is not essential here. A model that incorporates the above specifications is as follows for a balanced study, where the number of subjects receiving each level of factor A is the same:

$$Y_{ijk} = \mu_{...} + \rho_{i(j)} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \tag{27.11}$$

where:

$\mu_{...}$ is a constant
$\rho_{i(j)}$ are independent $N(0, \sigma_\rho^2)$
α_j are constants subject to $\sum \alpha_j = 0$
β_k are constants subject to $\sum \beta_k = 0$
$(\alpha\beta)_{jk}$ are constants subject to $\sum_j (\alpha\beta)_{jk} = 0$ for all k and $\sum_k (\alpha\beta)_{jk} = 0$ for all j
ε_{ijk} are independent $N(0, \sigma^2)$
$\rho_{i(j)}$ and ε_{ijk} are independent
$i = 1, \ldots, s; j = 1, \ldots, a; k = 1, \ldots, b$

The observations Y_{ijk} for repeated measures model (27.11) have the following properties:

$$E\{Y_{ijk}\} = \mu_{...} + \alpha_j + \beta_k + (\alpha\beta)_{jk} \tag{27.12a}$$

$$\sigma^2\{Y_{ijk}\} = \sigma_Y^2 = \sigma_\rho^2 + \sigma^2 \tag{27.12b}$$

$$\sigma\{Y_{ijk}, Y_{ijk'}\} = \sigma_\rho^2 \qquad k \neq k' \tag{27.12c}$$

$$\sigma\{Y_{ijk}, Y_{i'j'k'}\} = 0 \qquad i \neq i' \text{ and/or } j \neq j' \tag{27.12d}$$

Note that the observations Y_{ijk} have constant variance. In addition, in advance of the random trials any two observations for different levels of factor B for the same subject have constant covariance, for all subjects, while observations for different subjects are independent. Also, all observations are assumed to be normally distributed.

Once the subjects have been selected, repeated measures model (27.11) assumes that any two observations for the same subject are independent, that is, that there are no interference effects.

Analysis of Variance and Tests

Analysis of Variance. The ANOVA sums of squares for repeated measures model (27.11) can be obtained by means of the rules in Appendix D. The sum of squares that is used for estimating the error variance turns out to be the interaction sum of squares $SSB.S(A)$. The ANOVA sums of squares are shown in Table 27.5. Also shown there are the degrees of freedom for each sum of squares.

Tests for Factor Effects. The expected mean squares for the analysis of variance in Table 27.5 are given in Table 27.6. These expected mean squares can be obtained by means of the rules in Appendix D.

It is clear from the expected mean squares in Table 27.6 that the test for AB interaction effects:

$$\begin{aligned} &H_0\text{: all } (\alpha\beta)_{jk} = 0 \\ &H_a\text{: not all } (\alpha\beta)_{jk} \text{ equal zero} \end{aligned} \tag{27.13a}$$

uses the test statistic:

$$F^* = \frac{MSAB}{MSB.S(A)} \tag{27.13b}$$

TABLE 27.5 Analysis of Variance for Two-Factor Experiment with Repeated Measures on Factor B—Model (27.11).

Source of Variation	SS	df
Factor A	$SSA = bs\sum_j(\bar{Y}_{j\cdot} - \bar{Y}_{\cdot\cdot})^2$	$a-1$
Factor B	$SSB = as\sum_k(\bar{Y}_{\cdot k} - \bar{Y}_{\cdot\cdot})^2$	$b-1$
AB interactions	$SSAB = s\sum_j\sum_k(\bar{Y}_{jk} - \bar{Y}_{j\cdot} - \bar{Y}_{\cdot k} + \bar{Y}_{\cdot\cdot})^2$	$(a-1)(b-1)$
Subjects (within factor A)	$SSS(A) = b\sum_i\sum_j(\bar{Y}_{ij\cdot} - \bar{Y}_{j\cdot})^2$	$a(s-1)$
Error	$SSB.S(A) = \sum_i\sum_j\sum_k(Y_{ijk} - \bar{Y}_{jk} - \bar{Y}_{ij\cdot} + \bar{Y}_{j\cdot})^2$	$a(s-1)(b-1)$
Total	$SSTO = \sum_i\sum_j\sum_k(Y_{ijk} - \bar{Y}_{\cdot\cdot})^2$	$abs-1$

TABLE 27.6 Expected Mean Squares for o-Factor xperiment with Repeated Measures on Factor *B*—Model (27.11) (*A*, *B* fixed, subjects random).

Source of Variation	*MS*	*E*{*MS*}
Factor *A*	*MSA*	$\sigma^2 + b\sigma_\rho^2 + bs\frac{\sum \alpha_j^2}{a-1}$
Factor *B*	*MSB*	$\sigma^2 + as\frac{\sum \beta_k^2}{b-1}$
AB interactions	*MSAB*	$\sigma^2 + s\frac{\sum\sum (\alpha\beta)_{jk}^2}{(a-1)(b-1)}$
Subjects (within factor *A*)	*MSS*(*A*)	$\sigma^2 + b\sigma_\rho^2$
Error	*MSB.S*(*A*)	σ^2

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; (a-1)(b-1), a(s-1)(b-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; (a-1)(b-1), a(s-1)(b-1)], \text{ conclude } H_a \end{aligned} \tag{27.13c}$$

The test for factor A main effects:

$$\begin{aligned} &H_0\text{: all } \alpha_j = 0 \\ &H_a\text{: not all } \alpha_j \text{ equal zero} \end{aligned} \tag{27.14a}$$

uses the test statistic:

$$F^* = \frac{MSA}{MSS(A)} \tag{27.14b}$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; a-1, a(s-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; a-1, a(s-1)], \text{ conclude } H_a \end{aligned} \tag{27.14c}$$

Finally, the test for factor B main effects:

$$\begin{aligned} &H_0\text{: all } \beta_k = 0 \\ &H_a\text{: not all } \beta_k \text{ equal zero} \end{aligned} \tag{27.15a}$$

uses the test statistic:

$$F^* = \frac{MSB}{MSB.S(A)} \tag{27.15b}$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; b-1, a(s-1)(b-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; b-1, a(s-1)(b-1)], \text{ conclude } H_a \end{aligned} \tag{27.15c}$$

Comments

1. When the assumption of compound symmetry in repeated measures model (27.11) is not met, the conservative test discussed in Comment 2 on page 1065 should be employed.

2. When the study is not balanced (i.e., when the number of subjects within each level of factor A is not the same), the tests described here are no longer appropriate. Instead, the methods for unbalanced mixed and random effects models discussed in Section 25.7 can be employed. ■

Evaluation of Appropriateness of Repeated Measures Model

Our earlier discussion on evaluating the appropriateness of a repeated measures model applies here also. The residuals for repeated measures model (27.11) are:

$$e_{ijk} = Y_{ijk} - \bar{Y}_{.jk} - \bar{Y}_{ij.} + \bar{Y}_{i.} \tag{27.16}$$

A special feature of repeated measures model (27.11) also warrants attention. This model requires that the variance between subjects, σ_ρ^2, be constant for all levels of factor A. This assumption can be examined by dot plots of the estimated subject effects $\bar{Y}_{ij.} - \bar{Y}_{i.}$ for each level of factor A.

We can also conduct a formal test of the equality of the between-subjects variances by noting that the variation between subjects within factor A, $SSS(A)$, can be decomposed into components for each factor A level:

$$SSS(A) = SSS(A_1) + SSS(A_2) + \cdots + SSS(A_a) \tag{27.17}$$

where:

$$SSS(A_j) = b\sum_i (\bar{Y}_{ij.} - \bar{Y}_{i.})^2 \tag{27.17a}$$

Each component sum of squares has $n - 1$ degrees of freedom associated with it. We can therefore test the equality of the between-subjects variances by means of the Hartley test statistic (18.8) or the Brown-Forsythe test statistic (18.12). For the latter test, d_{ij} in (18.11) is defined as the absolute difference between the estimated mean, $\bar{Y}_{ij.}$, and the median of the estimated means $\bar{Y}_{1j.}, \ldots, \bar{Y}_{aj.}$.

Similarly, the error variation, $SSB.S(A)$, can be decomposed into components for each factor A level:

$$SSB.S(A) = SSB.S(A_1) + SSB.S(A_2) + \cdots + SSB.S(A_a) \tag{27.18}$$

where:

$$SSB.S(A_j) = \sum_i \sum_k (Y_{ijk} - \bar{Y}_{.jk} - \bar{Y}_{ij.} + \bar{Y}_{.j.})^2 \tag{27.18a}$$

Each component has $(s - 1)(b - 1)$ degrees of freedom associated with it. The Hartley or Brown-Forsythe tests can be conducted here also, this time to test for the equality of the error variance σ^2 for the different factor A levels.

The Hartley test assumes normality and is sensitive to this assumption. Hence, the appropriateness of the normality assumption should be established first before the Hartley test is employed. Unlike the Hartley test, the Brown-Forsythe test is robust and relatively insensitive to departures from normality.

Analysis of Factor Effects: Without Interaction

When the two factors do not interact or the interactions are not important, the main effects may be analyzed in a straightforward fashion. The relevant mean square to be used in the estimated variance of an estimated contrast of factor A level means for repeated measures model (27.11) is $MSS(A)$ because this mean square is the denominator of the appropriate F^* statistic for testing factor A main effects. Similarly, the mean square for estimating contrasts of factor B level means is $MSB.S(A)$.

The multiples for the estimated standard deviation of an estimated contrast of factor A or factor B level means are as follows:

Main A Effect	Main B Effect	
Single comparison		
$t[1-\alpha/2; a(s-1)]$	$t[1-\alpha/2; a(s-1)(b-1)]$	**(27.19a)**
Tukey procedure (for pairwise comparisons)		
$T = \frac{1}{\sqrt{2}}q[1-\alpha; a, a(s-1)]$	$T = \frac{1}{\sqrt{2}}q[1-\alpha; b, a(s-1)(b-1)]$	**(27.19b)**
Scheffé procedure		
$S^2 = (a-1)F[1-\alpha; a-1, a(s-1)]$	$S^2 = (b-1)F[1-\alpha; b-1, a(s-1)(b-1)]$	**(27.19c)**
Bonferroni procedure		
$B = t[1-\alpha/2g; a(s-1)]$	$B = t[1-\alpha/2g; a(s-1)(b-1)]$	**(27.19d)**

Note from Table 27.6 that the analysis of factor B effects can be carried out more precisely than that for factor A effects. The reason is that comparisons among factor A levels utilize $MSS(A)$, which involves the variability among the subjects as well as the experimental error, while comparisons among factor B levels utilize $MSB.S(A)$, which involves only experimental error.

Example 1

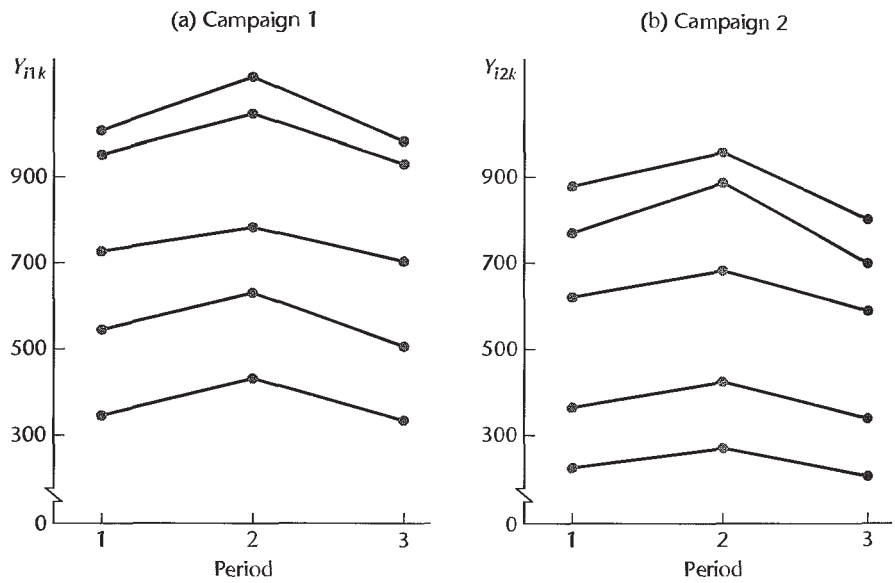
A national retail chain wanted to study the effects of two advertising campaigns (factor A) on the volume of sales of athletic shoes over time (factor B). Ten similar test markets (subjects, S) were chosen at random to participate in this study. The two advertising campaigns (A_1 and A_2) were similar in all respects except that a different national sports personality was used in each. Sales data were collected for three two-week periods (B_1: two weeks prior to campaign; B_2: two weeks during which campaign occurred; B_3: two weeks after campaign was concluded). The experiment was conducted during a six-week period when sales of athletic shoes are usually quite stable.

The data on sales (coded) are presented in Table 27.7, and are plotted in Figure 27.6 by test market for each advertising campaign. There is no evidence in Figure 27.6 of any interactions between the test markets and the treatments. In general, sales tended to increase during each advertising campaign, and then tended to decline to previous or lower levels than just before the campaign.

TABLE 27.7
Data—Athletic Shoes Sales Example.

Advertising Campaign	Test Market	Time Period $k=1$	$k=2$	$k=3$
	$i=1$	958	1,047	933
	$i=2$	1,005	1,122	986
$j=1$	$i=3$	351	436	339
	$i=4$	549	632	512
	$i=5$	730	784	707
	$i=1$	780	897	718
	$i=2$	229	275	202
$j=2$	$i=3$	883	964	817
	$i=4$	624	695	599
	$i=5$	375	436	351

FIGURE 27.6
Plots of Sales Data by Test Market and Campaign—Athletic Shoes Sales Example.



From Figure 27.6 and other diagnostic analyses (not shown), it was concluded that repeated measures model (27.11) is appropriate here. Figure 27.7 contains the MINITAB output for the fit of this model.

First we wish to test for campaign-time interaction effects:

$$H_0\text{: all } (\alpha\beta)_{jk} = 0$$

$$H_a\text{: not all } (\alpha\beta)_{jk} \text{ equal zero}$$

We use the results from Figure 27.7 in test statistic (27.13b):

$$F^* = \frac{MSAB}{MSB.S(A)} = \frac{196}{358} = .55$$

FIGURE 27.7 MINITAB Output for ANOVA—Athletic Shoes Sales Example.

Factor	Type	Levels	Values				
A	fixed	2	1	2			
S(A)	random	5	1	2	3	4	5
B	fixed	3	1	2	3		

Analysis of Variance for Y

Source	DF	SS	MS	F	P
A	1	168151	168151	0.73	0.417
S(A)	8	1833681	229210	640.31	0.000
B	2	67073	33537	93.69	0.000
A*B	2	391	196	0.55	0.589
Error	16	5727	358		
Total	29	2075023	71553		

Source	Variance Component	Error Term	Expected Mean Square (using restricted model)
1 A		2	(5) + 3(2) + 15Q[1]
2 S(A)	76284.0	5	(5) + 3(2)
3 B		5	(5) + 10Q[3]
4 A*B		5	(5) + 5Q[4]
5 Error	358.0		(5)

MEANS

A	N	Y
1	15	739.40
2	15	589.67

B	N	Y
1	10	648.40
2	10	728.80
3	10	616.40

For level of significance $\alpha = .05$, we require $F(.95; 2, 16) = 3.63$. Since $F^* = .55 \leq 3.63$, we conclude H_0, that no significant interaction effects are present. The P-value for the test is .59.

Next we wish to test for advertising campaign main effects:

$$H_0\text{: all } \alpha_j = 0$$
$$H_a\text{: not all } \alpha_j \text{ equal zero}$$

We use the results from Figure 27.7 in test statistic (27.14b):

$$F^* = \frac{MSA}{MSS(A)} = \frac{168{,}151}{229{,}210} = .73$$

For level of significance $\alpha = .05$, we require $F(.95; 1, 8) = 5.32$. Since $F^* = .73 \leq 5.32$, we conclude H_0, that no advertising campaign main effects exist. The P-value for the test is .42. Thus, either of the two national sports personalities is equally effective in the advertising campaign.

Finally, we wish to test for time period effects:

$$H_0\text{: all } \beta_k = 0$$
$$H_a\text{: not all } \beta_k \text{ equal zero}$$

Using the results from Figure 27.7 in test statistic (27.15b), we obtain:

$$F^* = \frac{MSB}{MSB.S(A)} = \frac{33{,}537}{358} = 93.7$$

For level of significance $\alpha = .05$, we require $F(.95; 2, 16) = 3.63$. Since $F^* = 93.7 > 3.63$, we conclude H_a, that period main effects exist. The P-value for the test is 0+.

To examine the nature of the time period effects, we shall conduct pairwise comparisons of mean sales for the three time periods:

$$L = \mu_{\cdot\cdot k} - \mu_{\cdot\cdot k'}$$

The Tukey procedure will be employed, with a 99 percent family confidence coefficient. We require:

$$T = \frac{1}{\sqrt{2}} q(.99; 3, 16) = \frac{1}{\sqrt{2}}(4.78) = 3.38$$
$$s^2\{\hat{L}\} = \frac{2MSB.S(A)}{as} = \frac{2(358)}{2(5)} = 71.60$$

Hence, $Ts\{\hat{L}\} = 3.38\sqrt{71.60} = 28.6$.

The point estimates of the changes in mean sales, based on the estimated factor B level means $\bar{Y}_{\cdot\cdot k}$ in Figure 27.7, are:

$$\hat{L}_1 = \bar{Y}_{\cdot\cdot 2} - \bar{Y}_{\cdot\cdot 1} = 728.8 - 648.4 = 80.4$$
$$\hat{L}_2 = \bar{Y}_{\cdot\cdot 3} - \bar{Y}_{\cdot\cdot 1} = 616.4 - 648.4 = -32.0$$
$$\hat{L}_3 = \bar{Y}_{\cdot\cdot 3} - \bar{Y}_{\cdot\cdot 2} = 616.4 - 728.8 = -112.4$$

and the desired confidence intervals therefore are:

$$52 \le \mu_{\cdot\cdot 2} - \mu_{\cdot\cdot 1} \le 109$$
$$-61 \le \mu_{\cdot\cdot 3} - \mu_{\cdot\cdot 1} \le -3$$
$$-141 \le \mu_{\cdot\cdot 3} - \mu_{\cdot\cdot 2} \le -84$$

We conclude with family confidence coefficient .99 that the two advertising campaigns lead to an immediate increase in mean sales of between 52 and 109 (8 to 17 percent), but that mean sales in the following period fall below those for the period preceding the campaign by somewhere between 3 and 61 (.5 to 9 percent).

Analysis of Factor Effects: With Interaction

When interactions exist between the two factors, the analysis of factor effects becomes considerably more complex. As we saw in Chapter 19, page 848, when interaction effects are important, attention usually focuses on simple effects. To compare simple main effects of the repeated measure factor B, the appropriate error term for these pairwise comparisons remains $MSB.S(A)$, the same as when there is no interaction. However, the appropriate

error term used for the pairwise comparisons of the simple main effects for factor A needs to be modified from that used without interaction in comparing main effects of factor A. For each level of factor B considered individually, the analysis reduces to a single-factor experiment in which there are no repeated measures. Hence, the mean square within treatments is the appropriate error term to make pairwise comparisons among the treatment effects within each level of factor B. This mean square is a weighted average of $MSB.S(A)$ and $MSS(A)$ where the weights are the corresponding degrees of freedom:

$$MS(\text{Within Treatments}) = \frac{a(b-1)(s-1)MSB.S(A) + a(s-1)MSS(A)}{ab(s-1)}$$

Note that MS(Within Treatments) is a linear combination of mean squares whose expectations are not necessarily the same. Stated differently, MS(Within Treatments) represents a pooling of what will often be heterogeneous sources of variability.

To employ this error term as a basis for pairwise comparisons among the simple main effects, we employ the Satterthwaite procedure. The correspondences to (25.26) for $\hat{L} = MS(\text{Within Treatments})$ are:

$$MS_1 = MSB.S(A) \qquad MS_2 = MSS(A) \qquad c_1 = \frac{a(b-1)(s-1)}{ab(s-1)} \qquad c_2 = \frac{a(s-1)}{ab(s-1)}$$

Substitution of these values into (25.28) leads to the Satterthwaite adjusted degrees of freedom:

$$df_{adj} = \frac{[SSB.S(A) + SSS(A)]^2}{\dfrac{[SSB.S(A)]^2}{a(b-1)(s-1)} + \dfrac{[SSS(A)]^2}{a(s-1)}} \tag{27.20}$$

We will now illustrate the analysis of factor effects in the presence of interactions with an example.

Example 2

During exercise, blood flow increases in some parts of the body in response to metabolic demand. Using radioactive microspheres, an experiment was conducted to determine in which of five parts of the body (factor B) this occurs. Microspheres distribute in tissue as a function of blood flow; i.e., the greater the blood flow to a part of the body, the more microspheres (and radioactivity) it will contain. The experiment was designed to compare blood flow in five different parts of the body (factor B) between the resting control condition (factor A_1) and during exercise (factor A_2). Tissues were examined in the following parts of the body: bone, brain, skin, muscle, and heart. The experiment was conducted by injecting a total of eight rats (subjects) intravenously with radioactive microspheres. After the microspheres were injected, four rats were exercised on a treadmill for 15 minutes (factor A_2) and the other four rats were placed on the treadmill, but the treadmill was not turned on (factor A_1). At the end of the 15-minute period, the rats were sacrificed and tissues in the five parts were harvested and the radioactivity in the tissues was measured. The data for this blood flow experiment are presented in Table 27.8 and plotted in Figure 27.8 by body part for each exercise condition.

On the basis of Figure 27.8 and other diagnostic analyses (not shown), it was decided that repeated measure model (27.11) is appropriate here. Table 27.9 contains the analysis of variance table based on repeated measures model (27.11).

TABLE 27.8 Data—Blood Flow during Exercise Example.*

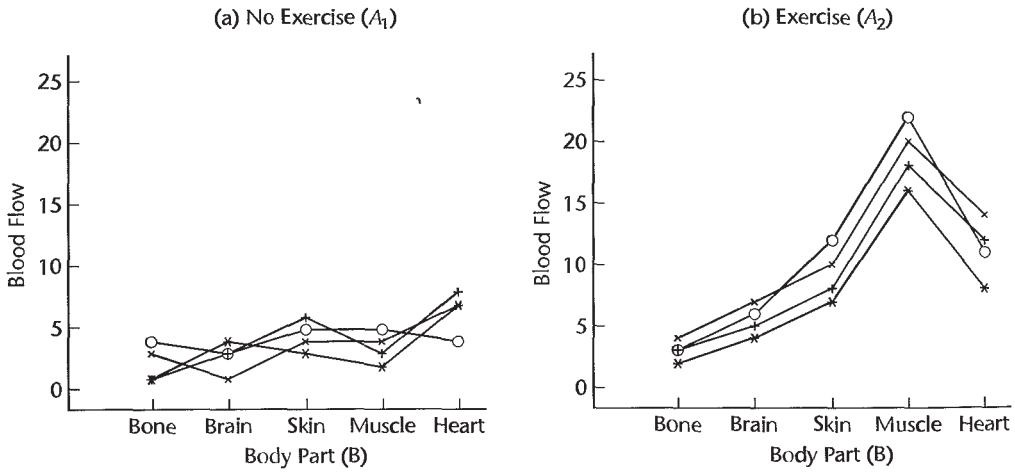
Exercise Condition		Body Part $k=1$ (Bone)	$k=2$ (Brain)	$k=3$ (Skin)	$k=4$ (Muscle)	$k=5$ (Heart)
(No Exercise)	$i=1$	4	3	5	5	4
	$i=2$	1	3	6	3	8
$j=1$	$i=3$	3	1	4	4	7
	$i=4$	1	4	3	2	7
(Exercise)	$i=1$	3	6	12	22	11
	$i=2$	3	5	8	18	12
$j=2$	$i=3$	4	7	10	20	14
	$i=4$	2	4	7	16	8

*Adapted from F.J. Gordon, *Analysis of Variance: Designs, Computations, and Multiple Comparisons*. Department of Pharmacology, Emory University School of Medicine, 2003.

TABLE 27.9 Analysis of Variance Table—Blood Flow during Exercise Example.

Source of Variation	*SS*	*df*	*MS*	*F**	*P*-value
A	324.9000	1	324.9000	44.104	.0006
S(*A*)	44.2000	6	7.3667		
B	389.5000	4	97.3750	49.936	.0000
AB	262.1000	4	65.5250	33.603	.0000
B.*S*(*A*)	46.8000	24	1.9500		
Total	1067.5000	39			

FIGURE 27.8 Plot of Exercise Condition by Body Part for Each Rat—Blood Flow during Exercise Example.



First we wish to test for exercise by body part interaction effects:

$$H_0\text{: all } (\alpha\beta)_{jk} = 0$$
$$H_a\text{: not all } (\alpha\beta)_{jk} \text{ equal zero}$$

We use the results from Table 27.9 as the test statistic (27.18a):

$$F^* = \frac{MSAB}{MSB.S(A)} = \frac{65.5250}{1.9500} = 33.603$$

For level of significance $\alpha = .05$, we require $F(.95; 4, 24) = 2.776$. Since $F^* = 33.6 > 2.776$, we conclude H_a, suggesting that interaction effects are present. The P-value for the test is 0+.

Next, because of the presence of a strong interaction effect, we wish to compare simple main effects of the repeated measures factor B (body part). We shall conduct pairwise comparisons of mean blood flows among body parts separately within the exercise and no exercise conditions; namely,

No Exercise	Exercise
$D_1 = \mu_{.11} - \mu_{.12}$	$D_{11} = \mu_{.21} - \mu_{.22}$
$D_2 = \mu_{.11} - \mu_{.13}$	$D_{12} = \mu_{.21} - \mu_{.23}$
$D_3 = \mu_{.11} - \mu_{.14}$	$D_{13} = \mu_{.21} - \mu_{.24}$
$D_4 = \mu_{.11} - \mu_{.15}$	$D_{14} = \mu_{.21} - \mu_{.25}$
$D_5 = \mu_{.12} - \mu_{.13}$	$D_{15} = \mu_{.22} - \mu_{.23}$
$D_6 = \mu_{.12} - \mu_{.14}$	$D_{16} = \mu_{.22} - \mu_{.24}$
$D_7 = \mu_{.12} - \mu_{.15}$	$D_{17} = \mu_{.22} - \mu_{.25}$
$D_8 = \mu_{.13} - \mu_{.14}$	$D_{18} = \mu_{.23} - \mu_{.24}$
$D_9 = \mu_{.13} - \mu_{.15}$	$D_{19} = \mu_{.23} - \mu_{.25}$
$D_{10} = \mu_{.14} - \mu_{.15}$	$D_{20} = \mu_{.24} - \mu_{.25}$

The Tukey procedure will be employed, with a 90 percent confidence coefficient, for each exercise condition. Then to combine these two Tukey procedures, a Bonferroni adjustment will be made for each exercise condition. Thus, we require

$$T = \frac{1}{\sqrt{2}} q(.95; 5, 24) = \frac{4.17}{\sqrt{2}} = 2.95$$
$$s^2\{\hat{D}\} = \frac{2MSB.S(A)}{s} = \frac{2(1.95)}{4} = .975$$

where .95 is used in the T argument instead of .90 to incorporate the Bonferroni adjustment for the two conditions. Hence, $Ts\{\hat{D}\} = 2.95\sqrt{.975} = 2.91$. Table 27.10 lists the cell means by exercise group and body part.

Any means within an exercise group that differ by more than 2.91 units are concluded to be significantly different from one another at the .10 level of significance. Therefore, for the no exercise group, heart is significantly different from bone, brain, and muscle. For the exercise group: heart is significantly different from bone, brain, and muscle; muscle is significantly different from bone, brain, skin, and heart; and skin is significantly different from bone, brain, and muscle.

TABLE 27.10 Treatment Means by Exercise Group and Body Part—Blood Flow during Exercise Example.

	$k = 1$ (Bone)	$k = 2$ (Brain)	$k = 3$ (Skin)	$k = 4$ (Muscle)	$k = 5$ (Heart)
$j = 1$ (No exercise)	2.25	2.75	4.50	3.50	6.50
$j = 2$ (Exercise)	3.00	5.50	9.25	19.00	11.25

To examine simple main effects of the nonrepeated measure factor A (exercise) for each level of B (body part), we shall conduct the five pairwise comparisons of mean blood flows between the two exercise groups within each body part; namely,

$$D_1 = \mu_{.11} - \mu_{.21}$$
$$D_2 = \mu_{.12} - \mu_{.22}$$
$$D_3 = \mu_{.13} - \mu_{.23}$$
$$D_4 = \mu_{.14} - \mu_{.24}$$
$$D_5 = \mu_{.15} - \mu_{.25}$$

The Tukey procedure will be employed using a 95 percent confidence coefficient for each body part with a Bonferroni adjustment for the five body parts. The within-treatment sum of squares is

$$SS(\text{Within Treatments}) = SSB.S(A) + SSS(A) = 46.8000 + 44.2000 = 91.0000$$

The approximate Satterthwaite adjusted degrees of freedom from (27.20) are:

$$df_{adj} = \frac{[46.8000 + 44.2000]^2}{\dfrac{(46.8000)^2}{2(4)(3)} + \dfrac{(44.2000)^2}{2(3)}} = \frac{8281.0000}{416.8667} = 19.86$$

Being conservative, we use $df_{adj} = 19$ associated with $MS(\text{Within Treatments})$, where

$$MS(\text{Within Treatments}) = \frac{91.0000}{30} = 3.033$$

Thus, we require

$$T = \frac{1}{\sqrt{2}} q(.99; 2, 19) = \frac{4.05}{\sqrt{2}} = 2.86$$

$$s^2\{\hat{D}\} = \frac{2MS(\text{Within Treatments})}{s} = \frac{2(3.033)}{4} = 1.52$$

Hence, $Ts\{\hat{D}\} = 2.86\sqrt{1.52} = 3.53$. Any means within body parts that differ by more than 3.53 units are significantly different from one another at the .10 level of significance. Therefore, we conclude that average blood flow for skin, muscle, and heart differ significantly between exercise groups.

FIGURE 27.9 Layout for Blocked Repeated Measures Design with Random Assignments of Factor A Level to Subjects and Repeated Measures on Factor B.

		Treatment Order 1	Treatment Order 2
Block 1	Subject 1	A_2B_1	A_2B_2
	Subject 2	A_1B_2	A_1B_1
Block 2	Subject 3	A_1B_2	A_1B_1
	Subject 4	A_2B_2	A_2B_1
⋮	⋮		⋮
Block n_b	Subject $2n_b - 1$	A_1B_1	A_1B_2
	Subject $2n_b$	A_2B_2	A_2B_1

Blocking of Subjects in Repeated Measures Designs

As already noted, comparisons among factor B effects can usually be carried out with greater precision than those for factor A effects because the latter involve between-subject variability as well as experimental error. To improve the precision of factor A comparisons, it is often helpful to block the subjects by some appropriate characteristic(s) so that the subjects within a block are homogeneous. Figure 27.9 illustrates the blocking of subjects in connection with the repeated measures design of Figure 27.5. Altogether, n_b blocks are used, each consisting of two similar subjects. One subject in each block is assigned at random to factor level A_1, the other is assigned to factor level A_2. In the second stage of randomization, each subject is randomly assigned the order of the two levels of factor B, namely, type of problem. Thus, the only difference between the repeated measures designs in Figures 27.9 and 27.5 is the blocking of the subjects for purposes of studying factor A effects more precisely. Note that for this layout, the number of subjects is $s = 2n_b$.

When there is a choice between which of the two factors should be the one on which repeated measures are taken (factor B), it should be the one for which more precise estimates are required. The reason is that even with blocking, the variability between subjects within a block will usually be greater than the variability within a subject.

27.4 Two-Factor Experiments with Repeated Measures on Both Factors

In Section 27.2 we considered single-factor repeated measures studies. The model for these designs can be extended when the treatments follow a factorial structure. For example, consider a study where four treatments are employed that represent two levels of each of two factors. Figure 27.10 depicts the layout for such a design when four subjects are utilized in the study. Note that the order of the treatments is randomized within each subject. When the treatments represent a factorial structure, we can explore as usual interaction effects as well as the main effects for the two factors. The design in Figure 27.10 is said to represent

FIGURE 27.10 Layout for Two-Factor Repeated Measures Design with Repeated Measures on Both Factors ($s = 4, a = 2, b = 2$).

	Treatment Order			
	1	2	3	4
Subject 1	A_1B_2	A_2B_2	A_1B_1	A_2B_1
2	A_2B_1	A_1B_2	A_2B_2	A_1B_1
3	A_2B_2	A_1B_1	A_2B_1	A_1B_2
4	A_1B_1	A_2B_1	A_1B_2	A_2B_2

repeated measures on both factors because each subject receives all treatments defined by the factorial structure.

Model

When both factor effects are fixed, the subjects constitute a random sample, and there are repeated measures on both factors, a model frequently appropriate is given by:

$$Y_{ijk} = \mu_{...} + \rho_i + \alpha_j + \beta_k + (\alpha\beta)_{jk} + (\rho\alpha)_{ij} + (\rho\beta)_{ik} + \varepsilon_{ijk} \tag{27.21}$$

where:

$\mu_{...}$ is a constant

ρ_i are independent $N(0, \sigma_\rho^2)$

α_j are constants subject to $\sum \alpha_j = 0$

β_k are constants subject to $\sum \beta_k = 0$

$(\alpha\beta)_{jk}$ are constants subject to $\sum_j (\alpha\beta)_{jk} = 0$ for all k and $\sum_k (\alpha\beta)_{jk} = 0$ for all j

$(\rho\beta)_{ik}$ are $N\left(0, \frac{b-1}{b}\sigma_{\rho\beta}^2\right)$ subject to the restrictions $\sum_k (\rho\beta)_{ik} = 0$ for all i

$\sigma\{(\rho\beta)_{ik}, (\rho\beta)_{ik'}\} = -\frac{1}{b}\sigma_{\rho\beta}^2$ for $k \neq k'$

$(\rho\alpha)_{ij}$ are $N\left(0, \frac{a-1}{a}\sigma_{\rho\alpha}^2\right)$ subject to the restrictions $\sum_j (\rho\alpha)_{ij} = 0$ for all i

$\sigma\{(\rho\alpha)_{ij}, (\rho\alpha)_{ij'}\} = -\frac{1}{a}\sigma_{\rho\alpha}^2$ for $j \neq j'$

ρ_i, $(\rho\alpha)_{ij}$ and $(\rho\beta)_{ik}$ are pairwise independent

ε_{ijk} are independent $N(0, \sigma^2)$ and independent of ρ_i, $(\rho\alpha)_{ij}$ and $(\rho\beta)_{ik}$

$i = 1, \ldots, s; j = 1, \ldots, a; k = 1, \ldots, b$

Note that two of the interaction terms in the model are random since the factor ρ_i is a random effect and that all sums of effects over the fixed factor levels are zero.

The observations Y_{ijk} for repeated measures model (27.21) have the following properties:

$$E\{Y_{ijk}\} = \mu_{...} + \alpha_j + \beta_k + (\alpha\beta)_{jk} \tag{27.22a}$$

$$\sigma^2\{Y_{ijk}\} = \sigma_Y^2 = \sigma_\rho^2 + \frac{a-1}{a}\sigma_{\rho\alpha}^2 + \frac{b-1}{b}\sigma_{\rho\beta}^2 + \sigma^2 \tag{27.22b}$$

Model (27.21) is an extension of the single-factor repeated measures model (27.1), where the treatment effect τ_j is now decomposed into factor A and factor B main effects and an AB interaction effect. However, separate first-order treatment-by-subject interaction terms are assumed to exist.

Once the subjects have been selected, repeated measures model (27.21), like the earlier repeated measures model (27.1), assumes that all of the treatment observations for a given subject are independent—that is, that there are no interference effects.

Analysis of Variance and Tests

Analysis of Variance. The ANOVA sums of squares for model (27.21) and the expected mean squares can be obtained readily by following the rules in Appendix D. The sum of squares for estimating the error variance terms reflects the interactions between treatments and subjects. Table 27.11 presents the ANOVA decomposition, degrees of freedom, and expected mean squares for two-factor repeated measures model (27.21).

Tests for Factor Effects. It is clear from the expected mean squares column in Table 27.11a that the test for AB interaction effects:

$$H_0\text{: all } (\alpha\beta)_{jk} = 0 \qquad H_a\text{: not all } (\alpha\beta)_{jk} \text{ equal zero} \tag{27.23a}$$

uses the test statistic:

$$F^* = \frac{MSAB}{MSABS} \tag{27.23b}$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; (a-1)(b-1), (a-1)(b-1)(s-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; (a-1)(b-1), (a-1)(b-1)(s-1)], \text{ conclude } H_a \end{aligned} \tag{27.23c}$$

The test for factor A main effects:

$$H_0\text{: all } \alpha_j = 0 \qquad H_a\text{: not all } \alpha_j \text{ equal zero} \tag{27.24a}$$

uses the test statistic:

$$F^* = \frac{MSA}{MSAS} \tag{27.24b}$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \leq F[1-\alpha; a-1, (a-1)(s-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; a-1, (a-1)(s-1)], \text{ conclude } H_a \end{aligned} \tag{27.24c}$$

Similarly, the test for factor B main effects:

$$H_0\text{: all } \beta_k = 0 \qquad H_a\text{: not all } \beta_k \text{ equal zero} \tag{27.25a}$$

uses the test statistic:

$$F^* = \frac{MSB}{MSBS} \tag{27.25b}$$

TABLE 27.11 ANOVA Table and Sums of Squares for Two-Factor Repeated Measures Design with Repeated Measures on Both Factors—Subjects Random, Factors *A* and *B* Fixed.

(a) ANOVA Table

Source of Variation	*SS*	*df*	*MS*	$E\{MS\}$
Subjects(S)	*SSS*	$s-1$	*MSS*	$\sigma^2 + ab\sigma^2_\rho$
Factor *A*	*SSA*	$a-1$	*MSA*	$\sigma^2 + b\sigma^2_{\rho\alpha} + bs\frac{\sum \alpha_j^2}{a-1}$
Factor *B*	*SSB*	$b-1$	*MSB*	$\sigma^2 + a\sigma^2_{\rho\beta} + as\frac{\sum \beta_k^2}{b-1}$
AB interactions	*SSAB*	$(a-1)(b-1)$	*MSAB*	$\sigma^2 + s\frac{\sum\sum(\alpha\beta)^2_{jk}}{(a-1)(b-1)}$
AS interactions	*SSAS*	$(a-1)(s-1)$	*MSAS*	$\sigma^2 + b\sigma^2_{\rho\alpha}$
BS interactions	*SSBS*	$(b-1)(s-1)$	*MSBS*	$\sigma^2 + a\sigma^2_{\rho\beta}$
Error	*SSABS*	$(a-1)(b-1)(s-1)$	*MSABS*	σ^2
Total	*SSTO*	$abs-1$		

(b) Sums of Squares

$$SSS = ab\sum_i(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSA = sb\sum_j(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSB = sa\sum_k(\bar{Y}_{\cdot\cdot k} - \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSAB = s\sum_j\sum_k(\bar{Y}_{\cdot jk} - \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot k} + \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSAS = b\sum_i\sum_j(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSBS = a\sum_i\sum_k(\bar{Y}_{i\cdot k} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot k} + \bar{Y}_{\cdot\cdot\cdot})^2$$

$$SSABS = \sum_i\sum_j\sum_k(Y_{ijk} - \bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot k} - \bar{Y}_{\cdot jk} + \bar{Y}_{i\cdot\cdot} + \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot k} - \bar{Y}_{\cdot\cdot\cdot})^2$$

and the decision rule for controlling the Type I error at α is:

$$\begin{aligned} &\text{If } F^* \le F[1-\alpha; b-1, (b-1)(s-1)], \text{ conclude } H_0 \\ &\text{If } F^* > F[1-\alpha; b-1, (b-1)(s-1)], \text{ conclude } H_a \end{aligned} \tag{27.25c}$$

Comments

1. When the effects of either factor *A* or factor *B* are random, the expected mean squares can be found by employing the rules in Appendix D. In turn, these expected mean squares will identify the appropriate test statistics.

2. Conservative *F* tests described in Section 25.5 should be used when the assumption of compound symmetry in repeated measures model (27.21) is not met.

3. Repeated measures model (27.21) assumes that treatments and subjects interact. If treatments and subjects do not interact, it can be shown that the treatment by subject interaction sum of squares is made up of three components:

$$SSTR.S = SSAS + SSBS + SSABS$$

Thus, it is possible to pool the first-order interactions in the model (the factor A by subject interactions and the factor B by subject interactions) with the second-order interactions (the factor A by factor B by subject interactions). When the repeated measures model does not allow for interactions between treatments and subjects, the analysis of factor effects becomes somewhat easier. However, in many cases, $MSABS$ tends to be considerably smaller than either $MSAS$ or $MSBS$, justifying the use of separate error terms. ■

Evaluation of Appropriateness of Repeated Measures Model

Our earlier discussion on the evaluation of the appropriateness of repeated measures model (27.1) applies here as well. In particular, residual sequence plots by subject should be constructed to examine whether interference effects are present and whether the error variance is constant. Plots of the observations by subject should be utilized to see whether the assumption of no treatment by subject interactions is appropriate.

Analysis of Factor Effects

If factors A and B do not interact or interact only in an unimportant fashion, the analysis of factor A and factor B main effects proceeds as usual. For the analysis of either factor A or factor B main effects, either $MSAS$ or $MSBS$, respectively, will be used in the estimated variance of the estimated contrast since this mean square is the denominator of the F^* test statistic for testing factor A or factor B main effects.

The multiples for the estimated standard deviation of an estimated contrast of factor A or factor B level means are as follows:

Main A Effect	Main B Effect	
Single comparison		
$t[1-\alpha/2;(a-1)(s-1)]$	$t[1-\alpha/2;(b-1)(s-1)]$	**(27.26a)**
Tukey procedure (for pairwise comparisons)		
$T=\frac{1}{\sqrt{2}}q[1-\alpha;a,(a-1)(s-1)]$	$T=\frac{1}{\sqrt{2}}q[1-\alpha;b,(b-1)(s-1)]$	**(27.26b)**
Scheffé procedure		
$S^2=(a-1)F[1-\alpha;a-1,(a-1)(s-1)]$	$S^2=(b-1)F[1-\alpha;b-1,(b-1)(s-1)]$	**(27.26c)**
Bonferroni procedure		
$B=t[1-\alpha/2g;(a-1)(s-1)]$	$B=t[1-\alpha/2g;(b-1)(s-1)]$	**(27.26d)**

If strong interactions between factors A and B exist that cannot be made unimportant by some simple transformation, the analysis of the factor effects should be performed in terms of the treatment means $\mu_{\cdot jk}$, which are averaged over subjects. This analysis is similar to that in Section 27.3 for a two-factor study with interaction. The pooled mean square

MSTR.S will be used in estimating the variance of any estimated contrast of the treatment means. The degrees of freedom associated with *MSTR.S* will need to be estimated using the Satterthwaite procedure discussed before in Chapter 25, page 1043.

Example

A clinician studied the effects of two drugs used either alone or together on the blood flow in human subjects. Twelve healthy middle-aged males participated in the study and they are viewed as a random sample from a relevant population of middle-aged males. The four treatments used in the study are defined as follows:

A_1B_1	placebo (neither drug)
A_1B_2	drug B alone
A_2B_1	drug A alone
A_2B_2	both drugs A and B

The 12 subjects received each of the four treatments in independently randomized orders. The response variable is the increase in blood flow from before to shortly after the administration of the treatment. The treatments were administered on successive days. This wash-out period prevented any carryover effects because the effect of each drug is short-lived. The experiment was conducted in a double-blind fashion so that neither the physician nor the subject knew which treatment was administered when the change in blood flow was measured.

Table 27.12 contains the data for this study. A negative entry denotes a decrease in blood flow. Figure 27.11 contains the MINITAB output for the fit of repeated measures model (27.21). Included in the output are the expected mean squares for the specified ANOVA model. As explained in Chapter 25, each term in an expected mean square is represented in the MINITAB output by (1) the numeric code, in parentheses, for the variance of the model term and (2) the preceding number, which is the numerical multiple. When the model term is fixed, the letter Q is used in the printout to show that the variance is replaced by the sum of squared effects divided by degrees of freedom. For example, the expected value of *MSA* as shown in Figure 27.11 is:

$$(7) + 2(5) + 24Q[2] = \sigma^2 + 2\sigma^2_{\rho\alpha} + 24\frac{\sum \alpha_j^2}{2-1}$$

which corresponds, of course, to the factor *A* expected mean square shown in Table 27.11a.

TABLE 27.12 Data—Blood Flow Example.

Subject i	Treatment A_1B_1	A_1B_2	A_2B_1	A_2B_2
1	2	10	9	25
2	−1	8	6	21
3	0	11	8	24
...	...	...	...	...
10	−2	10	10	28
11	2	8	10	25
12	−1	8	6	23

FIGURE 27.11 MINITAB Output for ANOVA—Blood Flow Example.

(a) MINITAB Output

Analysis of Variance for Flow

Source	DF	SS	MS	F	P
Subject	11	258.50	23.50	20.68	0.000
A	1	1587.00	1587.00	775.87	0.000
B	1	2028.00	2028.00	524.89	0.000
A*B	1	147.00	147.00	129.36	0.000
Subject*A	11	22.50	2.05	1.80	0.172
Subject*B	11	42.50	3.86	3.40	0.027
Error	11	12.50	1.14		
Total	47	4098.00			

Source	Variance Component	Error Term	Expected Mean Square for Each Term (using restricted model)
1 Subject	5.5909	7	(7) + 4(1)
2 A		5	(7) + 2(5) + 24Q[2]
3 B		6	(7) + 2(6) + 24Q[3]
4 A*B		7	(7) + 12Q[4]
5 Subject*A	0.4545	7	(7) + 2(5)
6 Subject*B	1.3636	7	(7) + 2(6)
7 Error	1.1364		(7)

(b) SAS Output

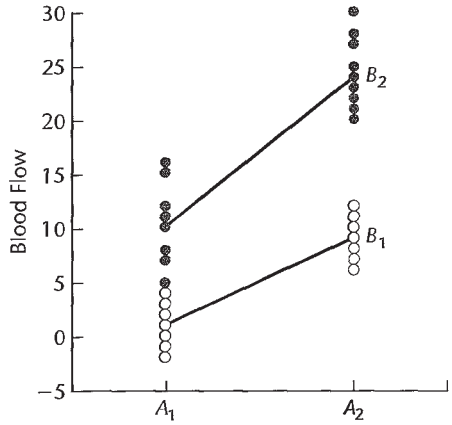
Source	DF	Type III SS	Mean Square	F Value	Pr > F
a	1	1587.000000	1587.000000	775.87	<.0001
Error(a)	11	22.500000	2.045455		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
b	1	2028.000000	2028.000000	524.89	<.0001
Error(b)	11	42.500000	3.863636		

Source	DF	Type III SS	Mean Square	F Value	Pr > F
a*b	1	147.0000000	147.0000000	129.36	<.0001
Error(a*b)	11	12.5000000	1.1363636		

	N	Mean	Std Dev	Minimum	Maximum
a1b1	12	0.5000000	2.1105794	−2.0000000	4.0000000
a1b2	12	10.0000000	3.1908961	5.0000000	16.0000000
a2b1	12	8.5000000	2.0225996	6.0000000	12.0000000
a2b2	12	25.0000000	3.4377583	20.0000000	31.0000000

FIGURE 27.12 Interaction Plot with Responses Superimposed—Blood Flow Example.



Various diagnostics were utilized to see if repeated measures model (27.21) is appropriate for the data in Table 27.12. The results (not shown here) supported the appropriateness of this model. The clinician expected the two drugs to interact in increasing the blood flow. To test for interaction effects:

$$H_0\text{: all } (\alpha\beta)_{jk} = 0$$
$$H_a\text{: not all } (\alpha\beta)_{jk} \text{ equal zero}$$

we use test statistic (27.23b) and the results from Figure 27.11:

$$F^* = \frac{MSAB}{MSABS} = \frac{147.000}{1.1364} = 129.36$$

For level of significance $\alpha = .01$, we require $F(.99; 1, 11) = 9.65$. Since $F^* = 129.36 > 9.65$, we conclude H_a, that interaction effects exist. The P-value for this test is 0+.

Figure 27.12 contains an interaction plot of the estimated treatment means, with the responses superimposed. Substantial interaction effects are evident. To study the nature of the interaction effects, the clinician wished to compare the joint use of the two drugs with the use of each drug alone, drug A with drug B, and each drug with no drug. Thus, the following pairwise comparisons are to be made:

$$L_1 = \mu_{\cdot 22} - \mu_{\cdot 21} \qquad L_4 = \mu_{\cdot 21} - \mu_{\cdot 11}$$
$$L_2 = \mu_{\cdot 22} - \mu_{\cdot 12} \qquad L_5 = \mu_{\cdot 12} - \mu_{\cdot 11}$$
$$L_3 = \mu_{\cdot 21} - \mu_{\cdot 12}$$

Point estimates of these pairwise comparisons are ($\bar{Y}_{jk}$ values are in Figure 27.11b):

$$\hat{L}_1 = 25.0 - 8.5 = 16.5 \qquad \hat{L}_4 = 8.5 - .5 = 8.0$$
$$\hat{L}_2 = 25.0 - 10.0 = 15.0 \qquad \hat{L}_5 = 10.0 - .5 = 9.5$$
$$\hat{L}_3 = 8.5 - 10.0 = -1.5$$

The estimated variance of each estimate $\hat{L}$ is given in (17.22), with the relevant mean square here being *MSABS*. Hence, we have:

$$s^2\{\hat{L}\} = MSABS\left(\frac{1}{s} + \frac{1}{s}\right) = 1.1364\left(\frac{2}{12}\right) = .1894$$

and $s\{\hat{L}\} = .435$. Using the Bonferroni procedure with a 95 percent family confidence coefficient, we require $B = t[1 - (.05)/2(5); 11] = t(.995; 11) = 3.106$. Hence, $t(.995; 11)s\{\hat{L}\} = 3.106(.435) = 1.35$ and the desired confidence intervals with a 95 percent family confidence coefficient are:

$$15.15 \leq \mu_{\cdot 22} - \mu_{\cdot 21} \leq 17.85 \qquad 6.65 \leq \mu_{\cdot 21} - \mu_{\cdot 11} \leq 9.35$$
$$13.65 \leq \mu_{\cdot 22} - \mu_{\cdot 12} \leq 16.35 \qquad 8.15 \leq \mu_{\cdot 12} - \mu_{\cdot 11} \leq 10.85$$
$$-2.85 \leq \mu_{\cdot 21} - \mu_{\cdot 12} \leq -.15$$

It is clear from these results that either drug A alone or drug B alone leads to an increase in blood flow, and that the combination of the two drugs leads to a substantial additional increase in blood flow as compared to when either drug is used alone. Finally, a significant difference exists in the mean effects of the two drugs used alone.

Comments

1. Repeated measures designs are discussed in more detail in References 27.1 and 27.2.

2. In economics and econometrics, repeated measurement data over time are commonly referred to as *panel data*. The process of combining cross-sectional data and data over time to form a panel is called pooling. See References 27.3 and 27.4 for a discussion of these models and their analyses.

3. Another area of application for repeated measurement data is referred to as growth curve model analyses. Here separate regression models are fit to each subject over time. See Reference 27.5 for a discussion of these models and their analyses. ■

27.5 Regression Approach to Repeated Measures Designs

When the repeated measures study is balanced and the treatment effects are fixed, the analysis of variance model can be expressed in the form of a regression model with indicator variables for purposes of obtaining the various sums of squares and conducting tests for treatment effects. Repeated measures models (27.1) and (27.21) can be stated in the form of a regression model as explained in Section 23.4 for randomized block designs. Repeated measures model (27.11), which also involves nested effects, can be expressed in the form of a regression model by including suitable indicator variables as explained in Section 26.6 on page 1105.

When the repeated measures study is not balanced, as, for instance, when there are missing observations, the tests based on the expected mean squares in Tables 27.1, 27.6, and 27.11 are no longer appropriate. Methods for analyzing unbalanced mixed and random effects models are discussed in Section 25.7.

27.6 Split-Plot Designs

Split-plot designs are frequently used in field, laboratory, industrial, and social science experiments. The repeated measures design in Figure 27.5 for a study with repeated measures on one factor is a type of split-plot design. We shall discuss split-plot designs only for two-factor studies, but these designs can be extended to apply when three or more factors are under investigation.

Split-plot designs were originally developed for agricultural experiments. Consider an investigation to study the effects of two irrigation methods (factor A) and two fertilizers (factor B) on yield of a crop, using four available fields as experimental units. In a completely randomized design, four treatments (A_1B_1, A_1B_2, A_2B_1, A_2B_2) would then be assigned at random to the four fields. Since there are four treatments and just four experimental units, there will be no degrees of freedom for estimation of error, as shown in the following abbreviated ANOVA table, listing source of variation and degrees of freedom only:

Source of Variation	Degrees of Freedom
Factor A (irrigation methods)	1
Factor B (fertilizer types)	1
AB interactions	1
Error	0
Total	3

If the fields could be subdivided into smaller experimental units, replicates of each factor-level combination could be obtained and the error variance could then be estimated. Unfortunately, in this investigation it is not possible to apply different irrigation methods (factor A) in areas smaller than a field, although different fertilizer types (factor B) could be applied in relatively small areas. A split-plot design can accommodate this situation.

In a split-plot design, each of the two irrigation methods is randomly assigned to two of the four fields, which are usually called *whole plots*. In turn, each whole plot is then subdivided into two or more smaller areas called *split plots*, and the two fertilizers are then randomly assigned to the split plots within each whole plot. The key feature of split-plot designs is the use of two (or more) distinct levels of randomization. At the first level of randomization, the whole-plot treatments are randomly assigned to whole plots; at the second level, the split-plot treatments are randomly assigned to split plots.

The layout for the agricultural experiment example is shown in Figure 27.13. Note that this layout is conceptually identical to the layout for the two-factor repeated measures design in Figure 27.5. The fields in Figure 27.13 correspond to the subjects in Figure 27.5, and the split plots correspond to the occasions on which treatments can be applied to a subject. Consequently, the split-plot model here is the same as in (27.11):

$$Y_{ijk} = \mu_{\cdots} + \rho_{i(j)} + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \varepsilon_{ijk} \tag{27.27}$$

For the split-plot agricultural experiment example, α_j denotes the main effect of the jth irrigation method (jth whole-plot treatment) and β_k denotes the main effect of the kth

FIGURE 27.13 Layout for Two-Factor Split-Plot Experiment—Agricultural Experiment Example (factor A is whole-plot treatment and factor B is split-plot treatment).

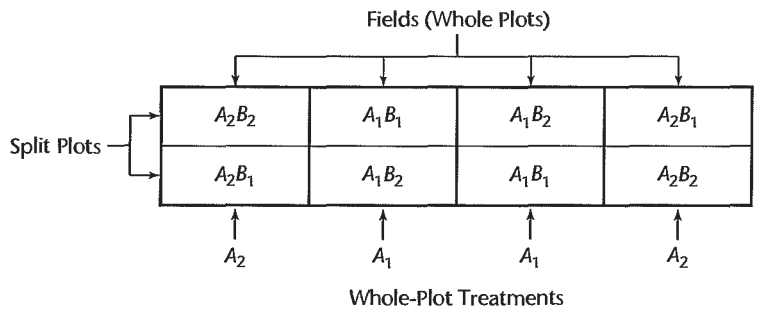


TABLE 27.13 ANOVA Table for Two-Factor Split-Plot Experiment.

Source of Variation	*SS*	*df*	*MS*
Whole plots			
Factor A	SSA	$a-1$	MSA
Whole-plot error	$SSW(A)$	$a(s-1)$	$MSW(A)$
Split plots			
Factor B	SSB	$b-1$	MSB
AB interactions	$SSAB$	$(a-1)(b-1)$	$MSAB$
Split-plot error	$SSB.W(A)$	$a(s-1)(b-1)$	$MSB.W(A)$
Total	$SSTO$	$abs-1$	

fertilizer type (kth split-plot treatment). Also, $\rho_{i(j)}$ denotes the effect of the ith whole plot, nested within the jth level of factor A (irrigation method).

Some computer packages produce special ANOVA tables that list the whole-plot effects and split-plot effects separately. Table 27.13 illustrates such a table. These tables serve as a reminder that the denominator of the F test for the whole-plot treatments is given by the error mean square for whole plots and that the denominator of the F test for the split-plot treatments and for the interactions between the whole-plot and split-plot treatments is given by the split-plot error mean square, as shown in Table 27.13. Note that this table is simply a rearrangement of the ANOVA table in Table 27.5 for a two-factor study with repeated measures on one factor. $SSS(A)$ is now denoted by $SSW(A)$ and $SSB.S(A)$ is now denoted by $SSB.W(A)$. The expected mean squares are the same as in Table 27.6.

Comments

1. Whenever subjects can receive all treatments in a two-factor study without interference effects, a repeated measures design with repeated measures on both factors might be preferable, because the factor effects for both factors may be estimated more precisely than in a split-plot design.

2. Split-plot designs are useful in industrial experiments when one factor requires larger experimental units than another. Consider, for instance, a study of the effects of two additives (factor A) and two different containers (factor B) for prolonging the shelf life of a milk product. Here, it is easier to make larger batches of the milk product with a given additive, whereas the different containers can be used with smaller batches.

3. Split-plot designs may be viewed as a type of incomplete block design where the whole plots are considered to be the blocks, with each whole plot being given only some of the full set of treatments. Incomplete block designs are discussed in Chapter 28.

4. A wide variety of split-plot designs has been developed. For instance, split-plot designs can involve more than two stages of randomization. In a split-split-plot experiment, three stages of randomization are generally involved. Whole plots are divided into split plots and split plots are further divided into split split plots. Three treatments are then assigned to the various levels of experimental units, using three distinct stages of randomization. References 27.2 and 27.6 provide further information about these designs. ■

Cited References

27.1. Winer. B. J., D. R. Brown, and K. M. Michels. *Statistical Principles in Experimental Design.* 3rd ed. New York: McGraw-Hill Book Co., 1991.

27.2. Koch, G. G., J. D. Elashoff, and I. A. Amara. "Repeated Measurements—Design and Analysis," in *Encyclopedia of Statistical Sciences,* vol. 8, eds. S. Kotz and N. L. Johnson. New York: John Wiley & Sons, 1988, pp. 46–73.

27.3. Pindyck, R. S., and D. L. Rubinfeld. *Econometric Models and Economic Forecasts.* 4th ed. Boston: Irwin/McGraw-Hill, 1998.

27.4. Hsiao. C.. *Analysis of Panel Data.* Cambridge: Cambridge University Press, 1986.

27.5. Graybill, F. A., *Theory and Application of the Linear Model.* Boston: Duxbury Press, 1976.

27.6. Dean, A., and D. Voss. *Design and Analysis of Experiments.* New York: Springer-Verlag, 1999.

Problems

27.1. A serious potential problem with repeated measures designs is associated with carryover effects. Describe some steps that can be taken to minimize this problem.

27.2. In designing a two-factor repeated measures study with repeated measures on one factor, does it matter which of the two factors is included as the repeated measures factor? Explain fully.

27.3. **Blood pressure.** The relationship between the dose of a drug that increases blood pressure and the actual amount of increase in mean diastolic blood pressure was investigated in a laboratory experiment. Twelve rabbits received in random order six different dose levels of the drug, with a suitable interval between each drug administration. The increase in blood pressure was used as the response variable. The data on blood pressure increase follow.

Rabbit	Dose (j)						Rabbit	Dose (j)					
i	.1	.3	.5	1.0	1.5	3.0	i	.1	.3	.5	1.0	1.5	3.0
1	21	21	23	35	36	48	7	9	12	17	22	33	40
2	19	24	27	36	36	46	8	20	20	30	30	38	41
3	12	25	27	26	33	40	9	18	18	27	31	42	49
4	9	17	18	27	34	39	10	8	12	11	24	26	31
5	7	10	19	25	31	38	11	18	22	25	32	38	38
6	18	26	26	29	39	44	12	17	23	26	28	34	35

a. Obtain the residuals for repeated measures model (27.1) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.1)?

b. Prepare aligned residual dot plots by dose level. Do these plots support the assumption of constancy of the error variance? Discuss.

c. Plot the observations Y_{ij} for each rabbit in the format of Figure 27.2. Does the assumption of no interactions between subjects (rabbits) and treatments appear to be reasonable here?

d. Conduct the Tukey test for additivity, conditional on the rabbits actually selected; use $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

27.4. Refer to **Blood pressure** Problem 27.3. Assume that repeated measures model (27.1) is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not the mean increase in blood pressure differs for the various dose levels; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Analyze the effects of the six dose levels by comparing the means for successive dose levels using the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings and summarize them by a suitable line plot.

d. According to the estimated efficiency measure (21.14), how effective was the repeated measures design here as compared to a completely randomized design?

27.5. Refer to **Blood pressure** Problems 27.3 and 27.4.

a. Develop a regression model in which the subject effects are represented by 1, −1, 0 indicator variables and the dose effect is represented by linear, quadratic, and cubic terms in $x = X - \bar{X}$, where X is the dose level. For instance, the x value for the first dose level ($X = .1$) is $x = .1 - 1.07 = -.97$.

b. Fit the regression model to the data.

c. Obtain the residuals and plot them against the fitted values. Does the model utilized appear to provide a reasonable fit?

d. Test whether or not the cubic effect is required in the model; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

27.6. **Grapefruit sales.** A supermarket chain studied the relationship between grapefruit sales and the price at which grapefruits are offered. Three price levels were studied: (1) the chief competitor's price, (2) a price slightly higher than the chief competitor's price, and (3) a price moderately higher than the chief competitor's price. Eight stores of comparable size were randomly selected for the study. Sales data were collected for three one-week periods, with the order of the three price levels randomly assigned for each store. The experiment was conducted during a time period when sales of grapefruits are usually quite stable, and no carryover effects were anticipated for this product. Data on store sales of grapefruits during the study period follow (data coded).

Store	Price level (j)		
i	1	2	3
1	62.1	61.3	60.8
2	58.2	57.9	55.1
...	...	...	...
7	46.8	43.2	41.5
8	51.2	49.8	47.9

a. Obtain the residuals for repeated measures model (27.1) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.1)?

b. Prepare aligned residual dot plots by price level. Do these plots support the assumption of constancy of the error variance? Discuss.

c. Plot the observations Y_{ij} for each store in the format of Figure 27.2. Does the assumption of no interactions between subjects (stores) and treatments appear to be reasonable here?

d. Conduct the Tukey test for additivity, conditional on the stores actually selected; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

*27.7. Refer to **Grapefruit sales** Problem 27.6. Assume that repeated measures model (27.1) is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not the mean sales of grapefruits differ for the three price levels; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Analyze the effects of the three price levels by estimating all pairwise comparisons of the price level means. Use the most efficient multiple comparison procedure with a 95 percent family confidence coefficient. State your findings and summarize them by a suitable line plot.

d. According to the estimated efficiency measure (21.14), how effective was the repeated measures design compared to a completely randomized design?

27.8. Refer to **Blood pressure** Problem 27.3. A consultant is concerned about the validity of the model assumptions and suggests that the study should be analyzed by means of the nonparametric rank F test. Rank the data within each rabbit and perform the rank F test; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. Comment on the consultant's concern here.

*27.9. Refer to **Grapefruit sales** Problem 27.6. It has been suggested that the nonparametric rank F test should be used here. Rank the data within each store and perform the rank F test; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. Is your conclusion the same as that obtained in Problem 27.7b?

27.10. **Truth in advertising.** A consumer research organization showed five different advertisements to 10 subjects and asked each to rank them in order of truthfulness. A rank of 1 denotes the most truthful. The results were:

Subject	Advertisement (j)					Subject	Advertisement (j)				
i	A	B	C	D	E	i	A	B	C	D	E
1	3	1	2	5	4	6	4	2	1	3	5
2	4	2	1	3	5	7	4	1	2	3	5
3	4	2	3	1	5	8	5	1	3	2	4
4	3	1	2	5	4	9	4	2	3	1	5
5	4	1	2	5	3	10	5	1	2	3	4

a. Do the subjects perceive the five advertisements as having equal truthfulness? Conduct the nonparametric rank F test using level of significance $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

b. Use the multiple pairwise testing procedure (27.9) to group the five different advertisements according to mean perceived truthfulness; employ family significance level $\alpha = .10$. Summarize your findings.

c. Obtain the coefficient of concordance (27.10) and interpret this measure.

27.11. **Incentive stimulus.** Refer to the example in Section 27.3 about the effects of two types of incentives (factor A) on a person's ability to solve two types of problems (factor B);

the repeated measures design is illustrated in Figure 27.5. Twelve persons were randomly selected and assigned in equal numbers to the two incentive groups. The order of the two types of problems was then randomized independently for each person. The problem-solving ability scores follow (the higher the score, the greater the ability to solve problems).

Incentive Stimulus	Subject	Problem Type: Abstract ($k = 1$)	Problem Type: Concrete ($k = 2$)
$j = 1$	$i = 1$	10	18
	$i = 2$	14	19
	$i = 3$	17	18
	$i = 4$	8	12
	$i = 5$	12	14
	$i = 6$	15	20
$j = 2$	$i = 1$	16	35
	$i = 2$	19	32
	$i = 3$	22	37
	$i = 4$	20	33
	$i = 5$	24	39
	$i = 6$	21	32

a. Obtain the residuals for repeated measures model (27.11) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.11)?

b. Plot the problem-solving ability scores by incentive stimulus and problem type, in the format of Figure 27.6. What do you conclude about the appropriateness of model (27.11)? Discuss.

12. Refer to **Incentive stimulus** Problem 27.11. Assume that repeated measures model (27.11) is appropriate.

a. Obtain the analysis of variance table.

b. Plot the data and the estimated treatment means in the format of Figure 27.12. Does it appear that interaction effects are present? That main effects are present?

c. Test whether or not the two factors interact; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. The following comparisons between problem types are of interest:

$$L_1 = \mu_{\cdot 11} - \mu_{\cdot 12} \qquad L_2 = \mu_{\cdot 21} - \mu_{\cdot 22}$$

Estimate these comparisons by means of confidence intervals. Use the Tukey procedure with a 90 percent family confidence coefficient for each problem type. Then combine these two Tukey procedures with a Bonferroni adjustment for each problem type. State your findings.

e. The following comparisons between incentive stimuli are of interest:

$$L_3 = \mu_{\cdot 11} - \mu_{\cdot 21} \qquad L_4 = \mu_{\cdot 12} - \mu_{\cdot 22}$$

Estimate these comparisons by means of confidence intervals. Use the Tukey procedure with a 90 percent family confidence coefficient for each incentive stimulus. Then combine these two Tukey procedures with a Bonferroni adjustment for each incentive stimulus. State your findings.

*27.13. **Store displays.** A repeated measures study was conducted to examine the effects of two different store displays for a household product (factor A) on sales in four successive time periods (factor B). Eight stores were randomly selected, and four were assigned at random to each display. The sales data (coded) follow.

Type of Display	Store	Time Period $k=1$	$k=2$	$k=3$	$k=4$
	$i=1$	956	953	938	1,049
$j=1$	$i=2$	1,008	1,032	1,025	1,123
	$i=3$	350	352	338	438
	$i=4$	412	449	385	532
	$i=1$	769	766	739	859
$j=2$	$i=2$	880	875	860	915
	$i=3$	176	185	168	280
	$i=4$	209	223	217	301

a. Obtain the residuals for repeated measures model (27.11) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.11)?

b. Plot the sales data by type of display and time period, in the format of Figure 27.6. What do you conclude about the appropriateness of model (27.11)? Discuss.

*27.14. Refer to **Store displays** Problem 27.13. The experimenter wished to explore further the appropriateness of repeated measures model (27.11).

a. Conduct a formal test of the constancy of the between-subjects variances. Use (27.17) and perform the Hartley test, with $\alpha = .01$. State the alternatives, decision rule, and conclusion.

b. Decompose the error variation $SSB.S(A)$ into components using (27.18), and perform the Hartley test for the constancy of the error variance σ^2 for the different factor A levels; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

*27.15. Refer to **Store displays** Problem 27.13. Assume that repeated measures model (27.11) is appropriate.

a. Obtain the analysis of variance table.

b. Plot the data and the estimated treatment means in the format of Figure 27.12. Does it appear that interaction effects are present? That main effects are present?

c. Test whether or not the two factors interact; use $\alpha = .025$. State the alternatives, decision rule, and conclusion. What is the P-value for the test?

d. Test separately whether or not display and time main effects are present; use $\alpha = .025$ for each test. State the alternatives, decision rule, and conclusion for each test. What is the P-value for each test?

e. To study the nature of the factor A and factor B main effects, estimate the following pairwise comparisons:

$$L_1 = \mu_{\cdot 1\cdot} - \mu_{\cdot 2\cdot} \qquad L_3 = \mu_{\cdot\cdot 2} - \mu_{\cdot\cdot 3}$$
$$L_2 = \mu_{\cdot\cdot 1} - \mu_{\cdot\cdot 2} \qquad L_4 = \mu_{\cdot\cdot 3} - \mu_{\cdot\cdot 4}$$

Use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

27.16. **Calculator efficiency.** To test the efficiency of its new programmable calculator, a computer company selected at random six engineers who were proficient in the use of both this calculator and an earlier model and asked them to work out two problems on both calculators. One of the problems was statistical in nature, the other was an engineering problem. The order of the four calculations was randomized independently for each engineer. The length of time (in minutes) required to solve each problem was observed. The results follow (type of problem is factor A and calculator model is factor B):

		$j = 1$ Statistical Problem		$j = 2$ Engineering Problem	
Engineer i		$k = 1$ New Model	$k = 2$ Earlier Model	$k = 1$ New Model	$k = 2$ Earlier Model
1	Jones	3.1	7.5	2.5	5.1
2	Williams	3.8	8.1	2.8	5.3
3	Adams	3.0	7.6	2.0	4.9
4	Dixon	3.4	7.8	2.7	5.5
5	Erickson	3.3	6.9	2.5	5.4
6	Maynes	3.6	7.8	2.4	4.8

a. Obtain the residuals for repeated measures model (27.21) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.21)?

b. Prepare aligned residual dot plots by treatment ignoring the factorial nature of the treatments. Do these plots support the assumption of constancy of the error variance? Discuss.

27.17. Refer to **Calculator efficiency** Problem 27.16. Assume that repeated measures model (27.21) is appropriate.

a. Obtain the analysis of variance table.

b. Plot the data and the estimated treatment means in the format of Figure 27.12. Does it appear that treatment interaction effects are present?

c. Test whether or not the two treatment factors interact; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. It is desired to study the nature of the interaction effects by considering the three comparisons:

$$L_1 = \mu_{\cdot 12} - \mu_{\cdot 11} \qquad L_3 = L_2 - L_1$$
$$L_2 = \mu_{\cdot 22} - \mu_{\cdot 21}$$

Obtain confidence intervals for these comparisons; use the Bonferroni procedure with a 95 percent family confidence coefficient. State your findings.

*27.18. **Migraine headaches.** Two experimental pain killer drugs for relief of migraine headaches were studied at a major medical center. Ten persistent migraine sufferers were randomly selected for a pilot study and received in random order each of the four treatment combinations, with a suitable interval between drug administrations. The decrease in pain intensity was used as the response variable. The four treatments used in the study are defined as follows: A_1B_1 = low dose of both drugs; A_1B_2 = low dose of drug A, high dose of drug B; A_2B_1 = high dose of drug A, low dose of drug B; A_2B_2 = high dose of both drugs. The data

on reduction in pain intensity follow (the higher the score, the greater the reduction in pain).

Person	A_1 ($j = 1$)		A_2 ($j = 2$)	
i	B_1 ($k = 1$)	B_2 ($k = 2$)	B_1 ($k = 1$)	B_2 ($k = 2$)
1	1.6	3.4	2.7	4.3
2	2.3	5.1	4.2	6.5
3	4.2	5.3	4.6	6.0
.		...	...	...
8	6.0	7.2	6.3	7.3
9	1.2	1.4	1.3	1.7
10	2.7	3.0	3.0	3.1

a. Obtain the residuals for repeated measures model (27.21) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.21)?

b. Prepare aligned residual dot plots by treatment ignoring the factorial nature of the treatments. Do these plots support the assumption of constancy of the error variance? Discuss.

*27.19. Refer to **Migraine headaches** Problem 27.18. Assume that repeated measures model (27.21) is appropriate.

a. Obtain the analysis of variance table.

b. Plot the data and the estimated treatment means in the format of Figure 27.12. Does it appear that treatment interaction effects are present? That main effects are present?

c. Test whether or not the two treatment factors interact; use $\alpha = .005$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Test separately whether or not factor A and factor B main effects are present; use $\alpha = .05$ for each test. State the alternatives, decision rule, and conclusion for each test. What is the P-value for each test?

e. Estimate the following comparisons by means of confidence intervals:

$$L_1 = \mu_{\cdot 21} - \mu_{\cdot 11} \qquad L_3 = \mu_{\cdot 21} - \mu_{\cdot 12}$$
$$L_2 = \mu_{\cdot 12} - \mu_{\cdot 11} \qquad L_4 = \mu_{\cdot 22} - \mu_{\cdot 11}$$

Use the Bonferroni procedure and family confidence coefficient .95. Summarize your findings.

27.20. **Wheat yield.** Refer to the split-plot agricultural experiment of Section 27.6, for which the layout is shown in Figure 27.13. The results of this experiment to investigate the effects of two irrigation methods (factor A) and two fertilizers (factor B) on wheat yield follow for the 10 fields used in the study.

Irrigation Method j:	1					2				
Field i:	1	2	3	4	5	1	2	3	4	5
Fertilizer $k = 1$:	43	40	31	27	36	63	52	45	47	54
$k = 2$:	48	43	36	30	39	70	53	48	51	57

a. Obtain the residuals for split-plot model (27.27) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of model (27.27)?

b. Plot the wheat yield data by irrigation method and type of fertilizer in the format of Figure 27.6. What do you conclude about the appropriateness of model (27.27)? Discuss.

27.21. Refer to **Wheat yield** Problem 27.20. Assume that split-plot model (27.27) is appropriate.

a. Obtain the analysis of variance table.

b. Plot the data and the estimated treatment means in the format of Figure 27.12. Does it appear that interaction effects are present? That main effects are present?

c. Test whether or not the two factors interact; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value for the test?

d. Test separately whether or not factor A and factor B main effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion for each test. What is the P-value for each test?

e. To study the nature of the factor A and factor B main effects, estimate the following pairwise comparisons:

$$L_1 = \mu_{\cdot 1 \cdot} - \mu_{\cdot 2 \cdot} \qquad L_2 = \mu_{\cdot\cdot 1} - \mu_{\cdot\cdot 2}$$

Use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

Exercise

27.22. Derive the total sum of squares breakdown in (27.5).

Projects

27.23. Refer to **Blood pressure** Problem 27.3. Obtain the estimated within-subjects variance-covariance matrix using (27.8). Are the estimated variances and covariances of the same orders of magnitude? Is the compound symmetry assumption reasonable here?

27.24. Refer to **Grapefruit sales** Problem 27.6. Obtain the estimated within-subjects variance-covariance matrix using (27.8). Are the variances and covariances roughly of the same order of magnitude? Is the compound symmetry assumption reasonably satisfied here?

27.25. Refer to the **Drug effect experiment** data set in Appendix C.12. Consider only Part I of the study and observation unit 1 for each drug dosage level; i.e., include only observations for which variable 2 equals 1 and variable 6 equals 1. Treat the 12 rats as subjects and ignore the classification of the rats into the three initial lever press rate groups. Assume that the subjects (rats) have random effects and that the treatments (dosage levels) have fixed effects.

a. State the additive repeated measures model for this study.

b. Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of the model employed?

c. Plot the responses for each rat in the format of Figure 27.2. Does the assumption of no interactions between subjects (rats) and treatments appear to be appropriate?

27.26. Refer to the **Drug effect experiment** data set in Appendix C.12 and Project 27.25.

a. Obtain the analysis of variance table.

b. Test whether or not the drug dosage level affects the mean lever press rate; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Analyze the effects of the four dosage levels by comparing the mean responses for each pair of successive dosage levels; use the Bonferroni procedure with a 90 percent family confidence coefficient. State your findings.

d. Fit a regression model in which the subject effects are represented by $1, -1, 0$ indicator variables and the dosage effect is represented by linear and quadratic terms in $x = X - \bar{X}$, where X is the dosage level. Assume that there are no interactions between subjects and treatments.

e. Obtain the residuals and plot them against the fitted values. Does the regression model appear to provide a good fit? Discuss.

f. Test whether or not the quadratic term can be dropped from the regression model; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

27.27. Refer to the **Drug effect experiment** data set in Appendix C.12. Consider the combined study. Assume that subjects (rats) and observation units have random effects, and that factor A (initial lever press rate), factor B (dosage level), and factor C (reinforcement schedule) have fixed effects. Also assume that there are no interactions between subjects and treatments.

a. Use rules (D.1) and (D.6) in Appendix D to develop the model for this experiment.

b. Fit the model in part (a), obtain the residuals, and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What do you conclude about the appropriateness of your model?

27.28. Refer to the **Drug effect experiment** data set in Appendix C.12 and Project 27.27. Assume that the model in Project 27.27a is appropriate.

a. Use an appropriate statistical package to obtain the analysis of variance table and the expected mean squares.

b. Test whether or not ABC interactions are present; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. For each reinforcement schedule, plot the estimated treatment means against dosage level with different curves for the three initial lever press rate groups, in the format of Figure 24.5. Examine your plots for the nature of the interaction effects and report your findings.

27.29. Consider a repeated measures design study with $s = 3$ and $r = 3$, where each subject ranks all treatments (with no ties allowed).

a. Develop the exact sampling distribution of F_R^* when H_0 holds. [*Hint:* All ranking permutations for a subject are equally likely under H_0 and all subjects are assumed to act independently.]

b. How does the 90th percentile of the exact sampling distribution obtained in part (a) compare with $F(.90; 2, 4)$? What is the implication of this?

Chapter 28

Balanced Incomplete Block, Latin Square, and Related Designs

In this chapter we introduce balanced incomplete block and latin square designs. Incomplete block designs are block designs where the number of experimental units in each block is less than the number of treatment combinations. This is in contrast with randomized complete block designs, where each block contains a complete replicate of the experiment. A latin square design is a particular form of incomplete block design, where two blocking variables are employed to reduce experimental errors while requiring only a small number of experimental trials.

28.1 Balanced Incomplete Block Designs

In Chapter 15, we described the use of an incomplete block design in the context of a food product taste-testing experiment. In that example, the food manufacturer wished to assess consumer acceptance of five breakfast cereal formulations. The formulations differed in terms of the amount of sweetener to be used in the formulation. Products were to be rated on a 10-point hedonic scale, and 12 consumers were available to rate the products. We noted that consumers differ considerably in their sensory perception of food products, and so it would be desirable to have each consumer rate all of the products. A randomized complete block (and repeated measures) design would result if each consumer were to rate all five of the formulations. However, consumers are generally unable to evaluate effectively more than three food products in a single session. With this restriction, the three tastings by any given consumer represent a single, incomplete block.

In situations such as that just described for the taste-testing example, an effective experimental arrangement can often be achieved using a *balanced incomplete block design,* or BIBD. An incomplete block design is *balanced* if every treatment appears with every other treatment in the same block the same number of times. For example, a candidate BIBD for the food product taste-testing experiment is shown in Figure 28.1. Note that there are $n_b = 10$ blocks, and that every treatment occurs together with every other treatment

FIGURE 28.1 Balanced Incomplete Block Design for Five Treatments and Block Size Three—Food Product Taste-Testing Example.

Consumer (Block)	Product Formulation				
	1	2	3	4	5
1	X	X	X		
2	X	X		X	
3	X	X			X
4	X		X	X	
5	X		X		X
6	X			X	X
7		X	X	X	
8		X	X		X
9		X		X	X
10			X	X	X

exactly three times. For example, formulations 1 and 2 appear together in blocks 1, 2, and 3. Formulations 1 and 3 appear together in blocks 1, 4, and 5—and so on. We shall use r_b to denote the number of treatments in each block (or block size), n_p to denote the number of times that pairs of treatments occur together in the same block, and n to denote the number of replicates of each treatment. Use of this design for the food product taste-testing example would mean that only 10 of the 12 available consumers could be used as subjects, because no balanced incomplete block design exists for $r = 5$ treatments, block size $r_b = 3$, and number of blocks $n_b = 12$.

If there is no restriction on the number of blocks, a BIBD can be constructed for any incomplete block size r_b ($2 \leq r_b < r$) by listing all of the possible subsets of size r_b from the set of r treatments. The number of such subsets is:

$$n_b = \frac{r!}{r_b!(r - r_b)!} \tag{28.1}$$

For example, the food product taste-testing example BIBD was constructed in this fashion. In this example, $r = 5$, $r_b = 3$, and the number of required blocks from (28.1) is $n_b = 5/[3!(5 - 3)!] = 10$. A limitation of this simple approach is that the number of blocks required can be quite large, and there may be alternative BIBDs with the same number of treatments and block size requiring fewer blocks. For example, with $r = 8$ and $r_b = 4$, the number of blocks required is $n_b = 8!/(4!4!) = 70$, but an alternative BIBD exists for $r = 8$ and $r_b = 4$ that requires just $n_b = 7$ blocks.

A useful set of BIBDs is provided in Appendix B.15 for the combinations of treatments, block sizes, and numbers of blocks shown in Table 28.1. For example, the BIBD for the food product taste-testing example shown in Figure 28.1 corresponds to design number 4 in Table 28.1. For this design, we have:

$$r = 5 \qquad r_b = 3 \qquad n_b = 10 \qquad n = 6 \qquad n_p = 3$$

A more extensive listing of BIBDs is provided in Reference 28.1.

ABLE 28.1
alanced
complete
lock Designs
Provided in
ppendix B.15.

Design Number	Number of Treatments r	Block Size r_b	Number of Blocks n_b	Number of Replicates n	Treatment Pairings n_p
1	4	2	6	3	1
2		3	4	3	2
3	5	2	10	4	1
4		3	10	6	3
5		4	5	4	3
6	6	2	15	5	1
7		3	10	5	2
8		3	20	10	4
9		4	15	10	6
10		5	6	5	4
11	7	2	21	6	1
12		3	7	3	1
13		4	7	4	2
14		6	7	6	5
15	8	2	28	7	1
16		4	14	7	3
17		7	8	7	6
18	9	3	12	4	1

Advantages and Disadvantages of BIBDs

Advantages of balanced incomplete block designs include:

1. A BIBD layout enables an investigator to run an experiment when the size of the available blocks of experimental units is smaller than the number of treatments. This is particularly helpful when a large number of treatments are under study.

2. Estimates of treatment effects have equal precision, and, as we shall see, the expressions for the variances of the estimated cell means and of contrasts of treatment means or effects are relatively simple. This simplifies the analysis and can facilitate sample size planning.

3. The presence of balance permits the use of the Scheffé and Tukey procedures for the analysis of treatment effects. These procedures cannot be used if an incomplete block design is not balanced.

Disadvantages of balanced incomplete block designs include:

1. As we have noted, balanced incomplete block designs exist only for certain combinations of numbers of treatments, block sizes, and numbers of blocks. Investigators may be compelled to adjust one or more of these parameters—i.e., by eliminating treatments, available blocks, or available experimental units—so that the available BIBD can be implemented. This may lead to a design that is balanced and relatively easy to analyze, but does not achieve fully the objectives of the study.

2. The assumption that there are no interactions between the blocking variable and the treatments is restrictive.

FIGURE 28.2 MINITAB Regression Results—Food Product Taste-Testing Example.

(a) Model (28.3)

Predictor	Coef	SE Coef	T	P
Constant	6.1667	0.1639	37.63	0.000
z1	0.1222	0.5130	0.24	0.815
z2	−0.5667	0.5130	−1.10	0.286
z3	1.2556	0.5130	2.45	0.026
z4	−1.7222	0.5130	−3.36	0.004
z5	1.4333	0.5130	2.79	0.013
z6	−0.9222	0.5130	−1.80	0.091
z7	0.3667	0.5130	0.71	0.485
z8	−0.8111	0.5130	−1.58	0.133
z9	−1.1667	0.5130	−2.27	0.037
x1	−1.6000	0.3590	−4.46	0.000
x2	1.1333	0.3590	3.16	0.006
x3	1.6000	0.3590	4.46	0.000
x4	0.6667	0.3590	1.86	0.082

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	13	97.2778	7.4829	9.29	0.000
Residual Error	16	12.8889	0.8056		
Total	29	110.1667			

(b) Regression Results for Model (28.4)

Predictor	Coef	SE Coef	T	P
Constant	6.1667	0.3249	18.98	0.000
z1	0.5000	0.9747	0.51	0.614
z2	−0.5000	0.9747	−0.51	0.614
z3	0.5000	0.9747	0.51	0.614
z4	−1.5000	0.9747	−1.54	0.139
z5	0.8333	0.9747	0.85	0.403
z6	−1.8333	0.9747	−1.88	0.075
z7	1.5000	0.9747	1.54	0.139
z8	−0.5000	0.9747	−0.51	0.614
z9	−1.1667	0.9747	−1.20	0.245

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	46.833	5.204	1.64	0.170
Residual Error	20	63.333	3.167		
Total	29	110.167			

(c) Regression Results for Model (28.5)

Predictor	Coef	SE Coef	T	P
Constant	6.1667	0.2662	23.16	0.000
x1	−1.6667	0.5325	−3.13	0.004
x2	1.0000	0.5325	1.88	0.072
x3	1.8333	0.5325	3.44	0.002
x4	0.3333	0.5325	0.63	0.537

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	57.000	14.250	6.70	0.001
Residual Error	25	53.167	2.127		
Total	29	110.167			

3. The analysis of a balanced incomplete block design is more complex than the analysis of a randomized complete block design. As we will see in Section 28.2, treatment and block effects are not orthogonal in BIBDs, and so the analysis is carried out using the regression approach.

We now turn to the statistical analysis of BIBDs, including the development of tests for treatment and block effects, and the analysis of factor-level effects. The analysis of a balanced incomplete block design is similar to the analysis of a randomized complete block design with missing cells, which was discussed earlier in Chapter 23.

Comment

When no BIBD exists for the desired number of treatments, number of blocks, and block size, some statisticians recommend the use of designs that are *nearly balanced.* Computer-based methods for constructing nearly-balanced incomplete block designs, available in statistical software packages such as JMP, are discussed in Reference 28.2. Related designs, called *partially balanced incomplete block designs,* have also been developed, a number of which are listed in Reference 28.1. The use of unbalanced incomplete block designs leads to a more complex analysis. For example, as already noted, the Scheffé and Tukey multiple comparisons procedures cannot be used with these designs for the analysis of treatment means. ■

28.2 Analysis of Balanced Incomplete Block Designs

BIBD Model

The model for a balanced incomplete block design is the same as that for a randomized complete block design. Thus either model (21.1) for fixed block effects, or model (25.67) for random block effects may be employed. The analysis of variance is the same for these two models, and all tests and estimates of treatment effects are conducted as for fixed block effects. For this reason we shall present only the fixed block effects case. Model (21.1) is:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \varepsilon_{ij} \qquad \textbf{(28.2)}$$

where:

$\mu_{..}$ is a constant
ρ_i are constants for the block (row) effects, subject to the restriction $\sum \rho_i = 0$
τ_j are constants for the treatment effects, subject to the restriction $\sum \tau_j = 0$
ε_{ij} are independent $N(0, \sigma^2)$
$i = 1, \ldots, n_b; j = 1, \ldots, r$

Note that model (28.2) assumes that no block-treatment interactions are present.

In Section 23.4, we discussed the analysis of randomized complete block designs when one or several observations are missing. This discussion is relevant to the analysis of BIBDs, because there are $r - r_b$ missing cells in each block. We noted there that missing cells destroy the orthogonality of the complete block design and make the usual ANOVA calculations inappropriate. However, the regression approach, as described on page 967, is still appropriate for additive model (28.2). Since no new principles are involved, we turn now to the use of the regression approach for the food product taste-testing example.

Regression Approach to Analysis of Balanced Incomplete Block Designs

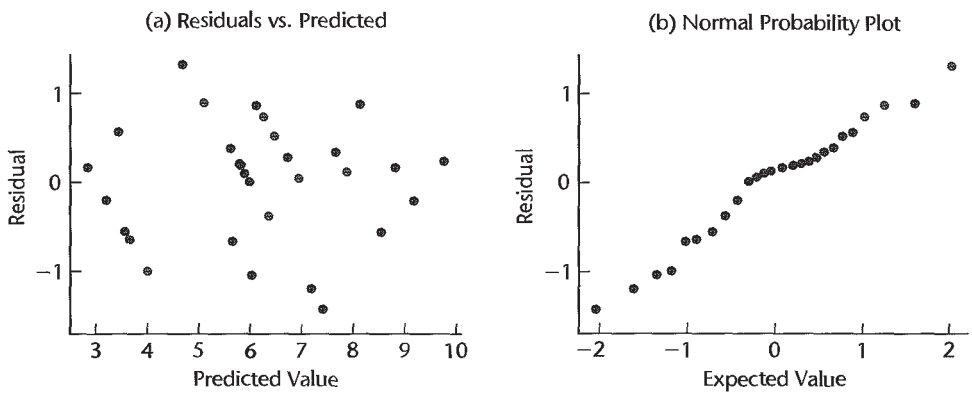
For the food product taste-testing example, the regression model equivalent to block design model (28.2) is as follows, where we use Xs to denote the 1, 0, −1, indicator variable predictors corresponding to treatment effects τ_1 through τ_4 and Zs to denote analogous predictors corresponding to block effects ρ_1 through ρ_9:

$$Y_{ij} = \mu_{..} + \rho_1 Z_{ij1} + \cdots + \rho_9 Z_{ij9} + \tau_1 X_{ij1} + \cdots + \tau_4 X_{ij4} + \varepsilon_{ij} \qquad \text{Full model} \qquad \textbf{(28.3)}$$

TABLE 28.2 Responses and Predictors—Food Product Taste-Testing Example.

i	j	(1) Y_{ij}	(2) Z_{ij1}	(3) Z_{ij2}	(4) Z_{ij3}	(5) Z_{ij4}	(6) Z_{ij5}	(7) Z_{ij6}	(8) Z_{ij7}	(9) Z_{ij8}	(10) Z_{ij9}	(11) X_{ij1}	(12) X_{ij2}	(13) X_{ij3}	(14) X_{ij4}
1	1	6	1	0	0	0	0	0	0	0	0	1	0	0	0
1	2	6	1	0	0	0	0	0	0	0	0	0	1	0	0
1	3	8	1	0	0	0	0	0	0	0	0	0	0	1	0
2	1	3	0	1	0	0	0	0	0	0	0	1	0	0	0
2	2	7	0	1	0	0	0	0	0	0	0	0	1	0	0
2	4	7	0	1	0	0	0	0	0	0	0	0	0	0	1
3	1	6	0	0	1	0	0	0	0	0	0	1	0	0	0
3	2	8	0	0	1	0	0	0	0	0	0	0	1	0	0
3	5	6	0	0	1	0	0	0	0	0	0	−1	−1	−1	−1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10	3	10	−1	−1	−1	−1	−1	−1	−1	−1	−1	0	0	1	0
10	4	9	−1	−1	−1	−1	−1	−1	−1	−1	−1	0	0	0	1
10	5	6	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1

FIGURE 28.3 Residual Plots—Food Product Taste-Testing Example.



where:

$$X_{ijk} = \begin{cases} 1 & \text{if response from product } k \text{ (i.e., if } j = k\text{), for } k = 1, 2, 3, 4 \\ -1 & \text{if response from product 5} \\ 0 & \text{otherwise} \end{cases}$$

$$Z_{ijk} = \begin{cases} 1 & \text{if response from subject } k \text{ (i.e., if } i = k\text{), for } k = 1, \ldots, 9 \\ -1 & \text{if response from subject 10} \\ 0 & \text{otherwise} \end{cases}$$

A portion of the data for the food product taste-testing BIBD is shown in Table 28.2. The response vector Y is displayed in column 1, Z_{ij1} through Z_{ij9} are shown in columns 2 through 10, and X_{ij1} through X_{ij4} are shown in columns 11 through 14. MINITAB regression output for the initial fit of model (28.3) is shown in Figure 28.2a. These results were obtained by regressing column 1 in Table 28.2 on columns 2 through 14. Residuals obtained from this fit are plotted against predicted values in Figure 28.3a and a normal probability plot of these residuals is shown in Figure 28.3b. No violations in assumptions are suggested

by the residual plots. The correlation between the residuals and the expected values under normality in Figure 28.3b is .988, which supports the assumption of approximate normality of the residuals.

Testing for the presence of treatment effects and block effects is carried out in the usual manner by first fitting full model (28.3) and then fitting each of the following reduced models:

Test for Treatment Effects

$$Y_{ij} = \mu_{..} + \rho_1 Z_{ij1} + \cdots + \rho_9 Z_{ij9} + \varepsilon_{ij} \qquad \text{Reduced model} \tag{28.4}$$

Test for Block Effects

$$Y_{ij} = \mu_{..} + \tau_1 X_{ij1} + \cdots + \tau_4 X_{ij4} + \varepsilon_{ij} \qquad \text{Reduced model} \tag{28.5}$$

Regression results for these two reduced models are shown in Figures 28.2b and 28.2c, respectively. We first consider the test for treatment effects.

The alternatives in the test for treatment effects implied by full model (28.3) and reduced model (28.4) are:

$$\begin{aligned} &H_0\colon \tau_1 = \tau_2 = \tau_3 = \tau_4 = 0 \\ &H_a\colon \text{not all } \tau_i = 0 \end{aligned} \tag{28.6}$$

Using general linear test statistic (2.70) and results from Figures 28.2a and 28.2b, we have:

$$\begin{aligned} F^* &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div MSE(F) \\ &= \frac{63.333 - 12.8889}{20 - 16} \div .8056 \\ &= 15.65 \end{aligned}$$

For $\alpha = .05$, we require $F(.95; 4, 16) = 3.01$. Since $15.65 > 3.01$, we conclude H_a that treatment effects are present. The P-value of the test is 0+.

In similar fashion, a test for block effects is obtained using full model (28.3) and reduced model (28.5). In this case, the alternatives are:

$$\begin{aligned} &H_0\colon \rho_1 = \rho_2 = \cdots = \rho_9 = 0 \\ &H_a\colon \text{not all } \rho_i = 0 \end{aligned} \tag{28.7}$$

and test statistic (2.70) is, using results from Figures 28.2a and 28.2c:

$$\begin{aligned} F^* &= \frac{SSE(R) - SSE(F)}{df_R - df_F} \div MSE(F) \\ &= \frac{53.167 - 12.8889}{25 - 16} \div .8056 \\ &= 5.56 \end{aligned}$$

For $\alpha = .05$, we require $F(.95; 9, 16) = 2.54$. Since $5.56 > 2.54$, we conclude H_a, that block effects are present. The P-value of the test is .0015.

At this point we have demonstrated that there are significant differences among the treatment means and that the use of blocking was effective. We now turn to the analysis of treatment effects for balanced incomplete block designs.

Analysis of Treatment Effects

Once the presence of treatment effects has been established using the regression approach, the analysis of these effects proceeds as described in Section 21.5 for randomized complete block designs, with the following modifications:

1. The least squares estimate of the jth treatment mean $\mu_{\cdot j}$ is given by:

$$\hat{\mu}_{\cdot j} = \hat{\mu}_{\cdot\cdot} + \hat{\tau}_j \tag{28.8}$$

   where $\hat{\mu}_{\cdot\cdot}$ and $\hat{\tau}_j$ are the least squares estimates of the regression coefficients $\mu_{\cdot\cdot}$ and τ_j in (28.3). Note that the least squares estimate of the ith treatment mean is *not* given here by $\bar{Y}_{\cdot j}$.
2. It can be shown that the variance of a contrast of estimated treatment means (or effects) is:

$$\sigma^2\{\hat{L}\} = \sigma^2\left\{\sum_{j=1}^{r} c_j\hat{\mu}_{\cdot j}\right\} = \sigma^2\frac{r_b}{rn_p}\sum_{j=1}^{r} c_j^2 \tag{28.9}$$

3. The estimated variance of a contrast of treatment means or effects is obtained by substituting the estimated variance $MSE(F)$ for full model (28.2) for σ^2 in (28.9):

$$s^2\{\hat{L}\} = MSE(F)\frac{r_b}{rn_p}\sum_{j=1}^{r} c_j^2 \tag{28.10}$$

4. The error degrees of freedom are now $n_T - (n_b - 1) - (r - 1) - 1 = n_b r_b - n_b - r + 1$.

The multiples for the estimated standard deviation of an estimated treatment mean or treatment contrast are then as follows:

Tukey procedure (for pairwise comparisons) $$T = \frac{1}{\sqrt{2}}q[1-\alpha; r, n_b r_b - n_b - r + 1] \tag{28.11a}$$

Scheffé procedure $$S^2 = (r-1)F[1-\alpha; r-1, n_b r_b - n_b - r + 1] \tag{28.11b}$$

Bonferroni procedure $$B = t[1-\alpha/2g; n_b r_b - n_b - r + 1] \tag{28.11c}$$

We illustrate the use of the Tukey procedure for the food product taste-testing example.

Example

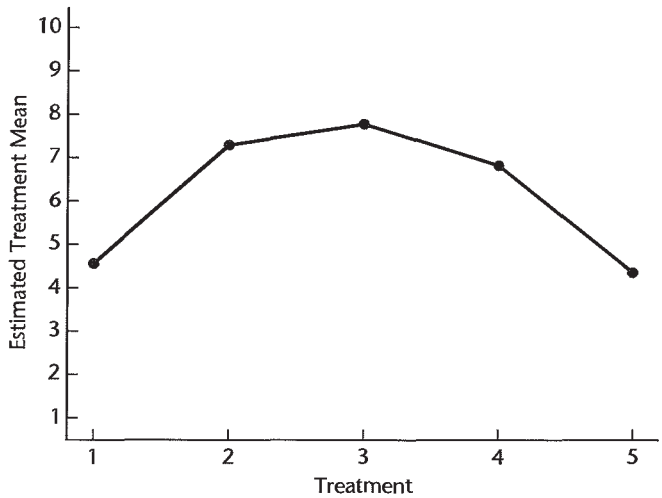
The least squares estimates of the five treatment means listed below were obtained using (28.8) and the regression results in Figure 28.2a:

j:	1	2	3	4	5
$\hat{\mu}_{\cdot j}$:	4.57	7.30	7.77	6.83	4.37

For example, the first estimated cell mean is $\hat{\mu}_{\cdot\cdot} + \hat{\tau}_1 = 6.17 + (-1.60) = 4.57$. These estimated treatment means are plotted against treatment number (j) in Figure 28.4. Note that the treatments 2, 3, and 4 lead to the largest estimated mean responses, and that treatments 1 and 5 appear to be substantially smaller. The investigators utilized the Tukey procedure to obtain all pairwise comparisons, employing a 95 percent family confidence coefficient.

For the food product taste-testing example, we have $r = 5$, $n_b = 10$, $n_p = 3$, and, from Figure 28.2a, $MSE(F) = .8056$. The estimated variance of the estimated difference between

FIGURE 28.4 Estimated Treatment Means Plot—Food Product Taste-Testing Example.



cell means 1 and 2, $\hat{D} = \hat{\mu}_{\cdot 1} - \hat{\mu}_{\cdot 2}$, using (28.10) is:

$$s^2\{\hat{D}\} = MSE(F)\frac{r_b}{rn_p}\sum_{j=1}^{r} c_j^2$$

$$= .8056\frac{3}{5(3)}(1^2 + (-1)^2 + 0^2 + 0^2 + 0^2) = .3222$$

Using (28.11a), we find for a 95 percent family confidence coefficient:

$$T = \frac{1}{\sqrt{2}}q(.95; 5, 16) = \frac{1}{\sqrt{2}}(4.33) = 3.06$$

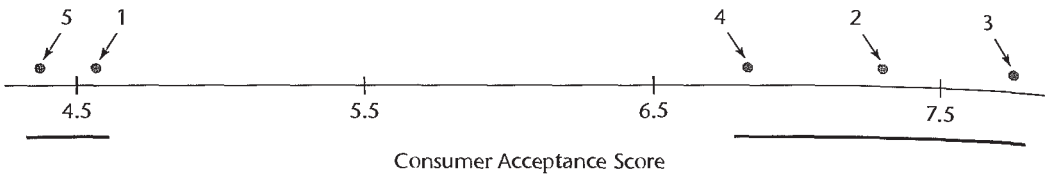
Hence:

$$Ts\{\hat{D}\} = 3.06\sqrt{.3222} = 1.74.$$

We now obtain all pairwise comparisons using (17.30) with $\hat{\mu}_{\cdot 1} = 4.57$, $\hat{\mu}_{\cdot 2} = 7.30$, $\hat{\mu}_{\cdot 3} = 7.77$, $\hat{\mu}_{\cdot 4} = 6.83$, and $\hat{\mu}_{\cdot 5} = 4.37$:

$$-4.47 = (4.57 - 7.30) - 1.74 \leq \mu_{\cdot 1} - \mu_{\cdot 2} \leq (4.57 - 7.30) + 1.74 = -0.99$$
$$-4.94 = (4.57 - 7.77) - 1.74 \leq \mu_{\cdot 1} - \mu_{\cdot 3} \leq (4.57 - 7.77) + 1.74 = -1.46$$
$$-4.00 = (4.57 - 6.83) - 1.74 \leq \mu_{\cdot 1} - \mu_{\cdot 4} \leq (4.57 - 6.83) + 1.74 = -0.52$$
$$-1.54 = (4.57 - 4.37) - 1.74 \leq \mu_{\cdot 1} - \mu_{\cdot 5} \leq (4.57 - 4.37) + 1.74 = 1.94$$
$$-2.21 = (7.30 - 7.77) - 1.74 \leq \mu_{\cdot 2} - \mu_{\cdot 3} \leq (7.30 - 7.77) + 1.74 = 1.27$$
$$-1.27 = (7.30 - 6.83) - 1.74 \leq \mu_{\cdot 2} - \mu_{\cdot 4} \leq (7.30 - 6.83) + 1.74 = 2.21$$
$$1.19 = (7.30 - 4.37) - 1.74 \leq \mu_{\cdot 2} - \mu_{\cdot 5} \leq (7.30 - 4.37) + 1.74 = 4.67$$
$$-0.80 = (7.77 - 6.83) - 1.74 \leq \mu_{\cdot 3} - \mu_{\cdot 4} \leq (7.77 - 6.83) + 1.74 = 2.68$$
$$1.66 = (7.77 - 4.37) - 1.74 \leq \mu_{\cdot 3} - \mu_{\cdot 5} \leq (7.77 - 4.37) + 1.74 = 5.14$$
$$0.72 = (6.83 - 4.37) - 1.74 \leq \mu_{\cdot 4} - \mu_{\cdot 5} \leq (6.83 - 4.37) + 1.74 = 4.20$$

We conclude that the five treatment means cluster into two distinct groups. The three largest estimated means corresponding to treatments 2, 3, and 4 are significantly different from treatment means 1 and 5, but not significantly different from each other, and the two smallest estimated treatment means (for treatments 1 and 5) are not significantly different from each other. A line plot of the estimated treatment means summarizes the results:



Planning of Sample Sizes with Estimation Approach

The essence of this approach is to specify the major comparisons of interest and to determine the expected widths of the confidence intervals for various sample sizes, given an advance planning value of the standard deviation. For a given number of treatments r and block size r_b, we need to determine the number of blocks n_b required to achieve confidence intervals of a specified width. We then determine if a BIBD exists for number of treatments r and block size r_b that has approximately the required number of blocks. In doing so, we will utilize the following two relations that hold for any balanced incomplete block design:

$$rn = r_b n_b$$

$$n_p(r-1) = n(r_b - 1)$$

From these relations we have: $n_b = rn/r_b$ and $n_p = n(r_b - 1)/(r-1)$.

We illustrate the estimation approach to the planning of sample sizes based on Tukey's pairwise comparison procedure and the taste-testing example.

Example

Suppose that Tukey's method for all pairwise comparisons will be used to analyze the BIBD for the food product taste-testing example with $r = 5$ and $r_b = 3$. Assume that σ will be no larger than 1.0 and the widths of the simultaneous 95 percent confidence intervals are not to exceed 2.0. In a BIBD the widths of all such intervals are the same, since the Tukey multiple T is the same for all pairs and since, from (28.10), $s^2\{\hat{D}\} = 2MSE(F)r_b/(rn_p)$. Using the fact that $n_b = rn/r_b = 5n/3$, the error degrees of freedom are:

$$df_e = n_b r_b - r - n_b + 1$$

$$= 5n - 5 - \frac{5n}{3} + 1 = \frac{10n}{3} - 4$$

The Tukey multiple comparison confidence limits for all pairwise comparisons $D_j = \mu_{\cdot j} - \mu_{\cdot j'}$ are:

$$\hat{D}_j \pm T\sigma\{\hat{D}_j\}$$

where $\sigma^2\{\hat{D}_j\} = 2\sigma^2(3)/(5n_p)$ from (28.9) and $T = (1/\sqrt{2})q[.95; 5, 10n/3 - 4]$. Furthermore, since $n_p = n(r_b - 1)/(r-1) = n(3-1)/(5-1) = n/2$, we obtain:

$$\sigma^2\{\hat{D}_j\} = \frac{6\sigma^2 4}{5(2)n} = \frac{12\sigma^2}{5n}$$

Therefore, the confidence interval halfwidth is:

$$T\sigma\{\hat{D}_j\} = \frac{1}{\sqrt{2}}q[.95; 5, 10n/3 - 4]\sqrt{\frac{12\sigma^2}{5n}}$$

With $\sigma^2 = 1$, the only unknown is n. We need to determine n so that $T\sigma\{\hat{D}_j\} \leq 1.0$ or $n \geq 1.20q^2[.95; 5, 10n/3 - 4])$. Using Table B.9, we find by trial and error that n must be greater than or equal to 24. For $r = 5$ and $r_b = 3$, note that the number of replicates for design 4 in Table 28.1 is $n = 6$. Therefore, the required number of replicates is achieved by repeating this particular BIBD four times, for which $n = 24$ and $n_b = 40$.

Comment

It is also possible to use the power approach or to use the selection of the "best" treatment approach to plan sample sizes. See Reference 28.3 for a discussion of sample size planning using the power approach. ■

28.3 Latin Square Designs

We saw in Section 21.6 that two blocking variables can be used simultaneously in randomized complete block designs to eliminate from experimental error the variation associated with each of the blocking variables. For instance, the blocking variables might be age and income of subject, with a block containing subjects in a given age and income group.

However, the full use of two blocking variables in a complete block design often requires too many experimental units. For instance, if the age and income variables in the illustration have six classes each, 36 blocks would be required. If six treatments were to be studied, 216 subjects would be needed for the experiment. Cost considerations may not permit the use of 216 experimental units, yet precision and range of validity considerations may require the simultaneous use of two blocking variables, each with six classes, in order to reduce the experimental error variance sufficiently and to have a reasonable variety of experimental subjects. In this type of situation, a *latin square design* may be helpful.

Basic Ideas

Taking incomplete block designs to the extreme in our example, given the employment of 36 blocks, the number of experimental units is minimized if only one treatment is run in each block. This extreme case, where each block contains only one treatment, is the type of situation for which a latin square design is appropriate. Table 28.3 provides an illustration of the difference between complete and incomplete block designs for the example considered. Column 1 shows the complete block design for this case, while columns 2 and 3 illustrate incomplete block designs, with three treatments and one treatment in each block, respectively.

There is another reason, besides economy, why a latin square design with only one treatment per block is used, namely, that blocks sometimes cannot contain more than one treatment. Consider the repeated measures design discussed in Section 27.2 where each subject receives every treatment. The repeated measures model in (27.1) assumes that no interference effects due to order position are present. If, indeed, such effects are possible, it may be desirable to use the order position as another blocking variable. Thus, "subject"

TABLE 28.3 Complete and Incomplete Block Designs.

Block Description	(1) Complete Block Design	(2) Incomplete Block Design (three treatments per block)	(3) Incomplete Block Design (one treatment per block)
Age under 25, income under $10,000	$T_1, T_2, T_3, T_4, T_5, T_6$	T_1, T_3, T_5	T_2
Age under 25, income $10,000–$19,999	$T_1, T_2, T_3, T_4, T_5, T_6$	T_2, T_4, T_6	T_5
...	...	...	...
Age 25–34, income under $10,000	$T_1, T_2, T_3, T_4, T_5, T_6$	T_2, T_4, T_5	T_3
...	...	...	...
Age 35–44, income under $10,000	$T_1, T_2, T_3, T_4, T_5, T_6$	T_3, T_4, T_6	T_2
etc.	etc.	etc.	etc.

would be one blocking variable and "order position of treatment" a second blocking variable Blocks would then be defined as follows for a study involving six treatments:

Block 1:	Subject 1, position 1
Block 2:	Subject 1, position 2
...	..
Block 6:	Subject 1, position 6
Block 7:	Subject 2, position 1
etc.	etc.

Notice that the blocks so defined can contain only one treatment, since the order position refers to the place of a single treatment in the sequence of treatments for a subject.

Description of Latin Square Designs

Let A, B, C represent three treatments; it is conventional with latin square designs to use Latin letters for the treatments. Suppose that day of week (Monday, Tuesday, Wednesday) and operator (1, 2, 3) are to be used as blocking variables. A latin square design might then be shown as follows:

	Operator		
Day	1	2	3
Monday	B	A	C
Tuesday	A	C	B
Wednesday	C	B	A

Operator 1 would run treatment B on Monday, treatment A on Tuesday, and treatment C on Wednesday, and so on for the other operators. Note that each operator runs each treatment, and that all treatments are run on each day.

A latin square design thus has the following features:

1. There are r treatments.
2. There are two blocking variables, each containing r classes.
3. Each row and each column in the design square contains all treatments; that is, each class of each blocking variable constitutes a replication.

Advantages and Disadvantages of Latin Square Designs

Advantages of a latin square design include:

1. The use of two blocking variables often permits greater reductions in the variability of experimental errors than can be obtained with either blocking variable alone.
2. Treatment effects can be studied from a small-scale experiment. This is particularly helpful in preliminary or pilot studies.
3. It is often helpful in repeated measures experiments to take into account the order position effect of treatments by means of a latin square design.

Disadvantages of a latin square design are:

1. The number of classes of each blocking variable must equal the number of treatments. This restriction is often difficult to meet in practice.
2. The assumptions of the model are restrictive (e.g., that there are no interactions between either blocking variable and treatments, and also none between the two blocking variables).
3. The use of a latin square design will lead to a very small number of degrees of freedom for experimental error when only a few treatments are studied. On the other hand, when many treatments are studied, the degrees of freedom for experimental error may be larger than necessary.
4. The randomization required is somewhat more complex than that for earlier designs considered.

Because of the limitations on the degrees of freedom for experimental error just described, latin squares are rarely used when more than eight treatments are being investigated. For the same reason, when there are only a few treatments, say, four or less, additional replications are usually required when a latin square design is employed.

Randomization of Latin Square Design

There exist many latin squares for a given number of treatments. For example, for $r = 3$, there are altogether 12 different possible arrangements. Four of the 12 possible arrangements are (we omit the row and column blocking variable labels):

1			2			3			4		
A	*B*	*C*	*A*	*C*	*B*	*B*	*A*	*C*	*C*	*B*	*A*
B	*C*	*A*	*B*	*A*	*C*	*C*	*B*	*A*	*A*	*C*	*B*
C	*A*	*B*	*C*	*B*	*A*	*A*	*C*	*B*	*B*	*A*	*C*

The number of possible latin square designs increases rapidly as the number of treatments gets larger; for $r = 5$, there are 161,280 possible arrangements.

The objective of randomization is to select one of all possible latin squares for the given number of treatments r, such that each square has an equal probability of being selected. Clearly, it is not generally feasible to list all possible latin squares so that one can be selected at random. Instead, we utilize *standard latin squares*, which are latin squares in which the elements of the first row and the first column are arranged alphabetically. The earlier latin square 1 is a standard latin square. Table B.14 contains all the standard squares for $r = 3$ and 4, and a single selected standard square for $r = 5, 6, 7, 8$, and 9.

The randomization procedure usually employed with Table B.14 is as follows:

1. For $r = 3$, independently arrange the rows and columns of the standard square at random.
2. For $r = 4$, select one of the standard squares at random. Then, independently arrange its rows and columns at random.
3. For $r = 5$ and higher, independently arrange the rows, columns, and treatments of the given standard square at random.

It can be shown that this procedure selects one latin square at random from all possible squares for $r = 3$ and 4. For $r = 5$ or higher, the randomization procedure is not based on all possible latin squares, but rather on very large and suitable subsets thereof.

Example

An experiment was conducted to study the effects of different types of background music on the productivity of bank tellers. The treatments were defined as various combinations of tempo music (slow, medium, fast) and style of music (instrumental and vocal, instrumental only). The treatments and Latin letter designations were as follows:

Treatment	Latin Letter Designation	Tempo and Style of Music
1	*A*	Slow, instrumental and vocal
2	*B*	Medium, instrumental and vocal
3	*C*	Fast, instrumental and vocal
4	*D*	Medium, instrumental only
5	*E*	Fast, instrumental only

Table 28.4 contains the results of this experiment. The treatment in each cell is shown in parentheses. The experimental unit in this study is a working day for the crew of bank tellers: the productivity data pertain to the performance of the entire crew. Let Y_{ijk} denote the observation in the cell defined by the ith class of the row blocking variable and the jth class of the column blocking variable. The subscript k indicates the treatment assigned to this cell by the particular latin square design employed. For instance, $Y_{123} = 17$ is the productivity on Tuesday of the first week, and Table 28.4 indicates that the type of music on that day was C.

TABLE 28.4 Latin Square Design and Experimental Results—Background Music Example (productivity of crew—data coded).

Week	Day: M	T	W	Th	F	Mean
1	18 (*D*)	17 (*C*)	14 (*A*)	21 (*B*)	17 (*E*)	$\bar{Y}_{1..} = 17.4$
2	13 (*C*)	34 (*B*)	21 (*E*)	16 (*A*)	15 (*D*)	$\bar{Y}_{2..} = 19.8$
3	7 (*A*)	29 (*D*)	32 (*B*)	27 (*E*)	13 (*C*)	$\bar{Y}_{3..} = 21.6$
4	17 (*E*)	13 (*A*)	24 (*C*)	31 (*D*)	25 (*B*)	$\bar{Y}_{4..} = 22.0$
5	21 (*B*)	26 (*E*)	26 (*D*)	31 (*C*)	7 (*A*)	$\bar{Y}_{5..} = 22.2$
Mean	$\bar{Y}_{.1.} = 15.2$	$\bar{Y}_{.2.} = 23.8$	$\bar{Y}_{.3.} = 23.4$	$\bar{Y}_{.4.} = 25.2$	$\bar{Y}_{.5.} = 15.4$	$\bar{Y}_{...} = 20.6$

$\bar{Y}_{..1} = 11.4$ $\quad$ $\bar{Y}_{..4} = 23.8$
$\bar{Y}_{..2} = 26.6$ $\quad$ $\bar{Y}_{..5} = 21.6$
$\bar{Y}_{..3} = 19.6$

The subscript k in Y_{ijk} is actually redundant for a latin square design because the row and cell designation (i, j) determines the treatment for the particular latin square employed. However, we continue to use all three subscripts for ease of identification.

We shall analyze the results of this study in Section 28.5.

28.4 Latin Square Model

A latin square design model involves the main effect of the row blocking variable, denoted by ρ_i, the main effect of the column blocking variable, denoted by κ_j, and the treatment main effect, denoted by τ_k. It is assumed that no interactions exist between these three variables. Thus, the model employed is an additive one. For the case of fixed treatment and block effects, the model is:

$$Y_{ijk} = \mu_{...} + \rho_i + \kappa_j + \tau_k + \varepsilon_{ijk} \tag{28.12}$$

where:

$\mu_{...}$ is a constant
ρ_i, κ_j, τ_k are constants subject to the restrictions $\sum \rho_i = \sum \kappa_j = \sum \tau_k = 0$
ε_{ijk} are independent $N(0, \sigma^2)$
$i = 1, \ldots, r;\ j = 1, \ldots, r;\ k = 1, \ldots, r$

Note again that the number of classes for each of the two blocking variables is the same as the number of treatments, and that the total number of experimental trials is r^2.

Comment

If the treatment effects are random, the only change in model (28.12) is that the τ_k now are independent $N(0, \sigma_\tau^2)$ and are independent of the ε_{ijk}. ■

28.5 Analysis of Latin Square Experiments

Notation

We shall employ the usual notation for row, column, and treatment totals and means:

$$Y_{i\cdot\cdot} = \sum_j Y_{ijk} \qquad \bar{Y}_{i\cdot\cdot} = \frac{Y_{i\cdot\cdot}}{r} \tag{28.13a}$$

$$Y_{\cdot j\cdot} = \sum_i Y_{ijk} \qquad \bar{Y}_{\cdot j\cdot} = \frac{Y_{\cdot j\cdot}}{r} \tag{28.13b}$$

$$Y_{\cdot\cdot k} = \sum_{i,j} Y_{ijk} \qquad \bar{Y}_{\cdot\cdot k} = \frac{Y_{\cdot\cdot k}}{r} \tag{28.13c}$$

The overall total and mean are denoted as usual by:

$$Y_{\cdot\cdot\cdot} = \sum_i \sum_j Y_{ijk} \qquad \bar{Y}_{\cdot\cdot\cdot} = \frac{Y_{\cdot\cdot\cdot}}{r^2} \tag{28.13d}$$

Note the redundancy of any one of the three subscripts, arising from the fact that the treatment is uniquely determined by the row and column specifications for the latin square utilized. The various means for the background music example are shown in Table 28.4. The estimated treatment means are calculated by first collecting the data for each treatment and then averaging these values. For instance, we have:

$$\bar{Y}_{\cdot\cdot 1} = \frac{7 + 13 + 14 + 16 + 7}{5} = 11.4$$

Fitting of Model

The least squares and maximum likelihood estimators of the parameters in latin square model (28.12) are:

Parameter	Estimator	
$\mu_{\cdot\cdot\cdot}$	$\hat{\mu}_{\cdot\cdot\cdot} = \bar{Y}_{\cdot\cdot\cdot}$	(28.14a)
ρ_i	$\hat{\rho}_i = \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}$	(28.14b)
κ_j	$\hat{\kappa}_j = \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot}$	(28.14c)
τ_k	$\hat{\tau}_k = \bar{Y}_{\cdot\cdot k} - \bar{Y}_{\cdot\cdot\cdot}$	(28.14d)

The fitted values therefore are:

$$\hat{Y}_{ijk} = \bar{Y}_{i\cdot\cdot} + \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot k} - 2\bar{Y}_{\cdot\cdot\cdot} \tag{28.15}$$

and the residuals are:

$$e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot k} + 2\bar{Y}_{\cdot\cdot\cdot} \tag{28.16}$$

Analysis of Variance

Table 28.5 presents the ANOVA table for latin square model (28.12). The sums of squares can be obtained by the rules in Appendix D, remembering that one subscript is redundant.

TABLE 28.5 ANOVA Table for Latin Square Design Model (28.12) with Fixed Effects.

Source of Variation	*SS*	*df*	*MS*	$E\{MS\}$
Row blocking variable	*SSROW*	$r-1$	$MSROW = \frac{SSROW}{r-1}$	$\sigma^2 + r\frac{\sum \rho_i^2}{r-1}$
Column blocking variable	*SSCOL*	$r-1$	$MSCOL = \frac{SSCOL}{r-1}$	$\sigma^2 + r\frac{\sum \kappa_j^2}{r-1}$
Treatments	*SSTR*	$r-1$	$MSTR = \frac{SSTR}{r-1}$	$\sigma^2 + r\frac{\sum \tau_k^2}{r-1}$
Error	*SSRem*	$(r-1)(r-2)$	$MSRem = \frac{SSRem}{(r-1)(r-2)}$	σ^2
Total	*SSTO*	r^2-1		

The definitional forms of the sums of squares are as follows:

$$SSTO = \sum_i \sum_j (Y_{ijk} - \bar{Y}_{...})^2 \tag{28.17a}$$

$$SSROW = r \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2 \tag{28.17b}$$

$$SSCOL = r \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2 \tag{28.17c}$$

$$SSTR = r \sum_k (\bar{Y}_{..k} - \bar{Y}_{...})^2 \tag{28.17d}$$

$$SSRem = \sum_i \sum_j (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{..k} + 2\bar{Y}_{...})^2 \tag{28.17e}$$

SSROW is the *row sum of squares*. The more the row means $\bar{Y}_{i..}$ differ, the larger is *SSROW*. Similarly, *SSCOL* is the *column sum of squares* and measures the variability of the column means $\bar{Y}_{.j.}$. *SSTR* denotes, as usual, the treatment sum of squares. Finally, *SSRem* stands for the remainder sum of squares reflecting the error variability. We use this notation here since this sum of squares is made up of several different interaction components.

The degrees of freedom in Table 28.5 can be understood as follows. There are r^2 observations, and hence *SSTO* has $r^2 - 1$ degrees of freedom associated with it. Since there are r classes for the row and column blocking variables each, and also r treatments, each of the corresponding sums of squares has $r - 1$ degrees of freedom associated with it. The number of degrees of freedom associated with *SSRem* is the remainder, namely, $(r^2 - 1) - 3(r - 1) = (r - 1)(r - 2)$. Note that the addition of a second blocking variable has reduced the number of degrees of freedom for the error sum of squares from $(r - 1)^2$ for a randomized complete block design based on r blocks and r treatments to $(r - 1)(r - 2)$, a reduction of $r - 1$ degrees of freedom.

The $E\{MS\}$ column in Table 28.5 for latin square model (28.12) can be obtained by using the rules in Appendix D, remembering that one subscript is redundant, or by a computer package that provides expected mean squares.

Test for Treatment Effects

To test for treatment effects in latin square model (28.12) with fixed effects:

$$H_0\text{: all } \tau_k = 0$$
$$H_a\text{: not all } \tau_k \text{ equal zero} \tag{28.18a}$$

we see from the $E\{MS\}$ column in Table 28.5 that the appropriate test statistic is:

$$F^* = \frac{MSTR}{MSRem} \tag{28.18b}$$

The appropriate decision rule to control the risk of a Type I error at α is:

$$\text{If } F^* \leq F[1-\alpha; r-1, (r-1)(r-2)], \text{ conclude } H_0$$
$$\text{If } F^* > F[1-\alpha; r-1, (r-1)(r-2)], \text{ conclude } H_a \tag{28.18c}$$

Comments

1. If the presence of blocking variable effects is to be tested, we see from the $E\{MS\}$ column in Table 28.5 that the appropriate test statistics are:

$$F^* = \frac{MSROW}{MSRem} \tag{28.19a}$$

$$F^* = \frac{MSCOL}{MSRem} \tag{28.19b}$$

2. If the treatment effects are random, the alternatives to be considered are:

$$H_0\text{: } \sigma_\tau^2 = 0$$
$$H_a\text{: } \sigma_\tau^2 > 0 \tag{28.20}$$

but the test statistic and decision rule are the same as in (28.18) for the fixed treatment effects case. ■

Analysis of Treatment Effects

When differential treatment effects are found by the analysis of variance and the treatments have fixed effects, estimates of contrasts involving the treatment effects are usually desired, often utilizing multiple comparison procedures. The appropriate mean square to be used in the estimated variance of the contrast is *MSRem* obtained from (28.17e), and the multiples for the estimated standard deviation of the contrast are as follows:

Single comparison $\quad t[1-\alpha/2; (r-1)(r-2)]$ **(28.21a)**

Tukey procedure (for pairwise comparisons) $\quad T = \frac{1}{\sqrt{2}} q[1-\alpha; r, (r-1)(r-2)]$ **(28.21b)**

Scheffé procedure $\quad S^2 = (r-1)F[1-\alpha; r-1, (r-1)(r-2)]$ **(28.21c)**

Bonferroni procedure (g comparisons) $\quad B = t[1-\alpha/2g; (r-1)(r-2)]$ **(28.21d)**

Residual Analysis

The use of the residuals in (28.16) for examining the aptness of a latin square model presents no new issues; the basic points made earlier for other designs apply also to latin square designs. The Tukey test for additivity in a randomized complete block design, discussed in Section 21.4, can be extended to latin square designs. Reference 28.3 describes the extension.

Example

The analysis of variance calculations for the background music data in Table 28.4 were made by using a computer package and the results are shown in Table 28.6. The residuals were also obtained and analyzed. Figure 28.5a contains a plot of the residuals against the fitted values, and Figure 28.5b contains a normal probability plot of the residuals. These plots do not reveal any serious departures from the model assumptions, though they show one case that appears to be outlying. The Bonferroni outlier test, explained on page 396, was employed to test whether this case is an outlier but did not identify it as such. Based on these and other diagnostics, including the Tukey test for additivity, it was concluded that model (28.12) is appropriate for the data.

To test for treatment effects:

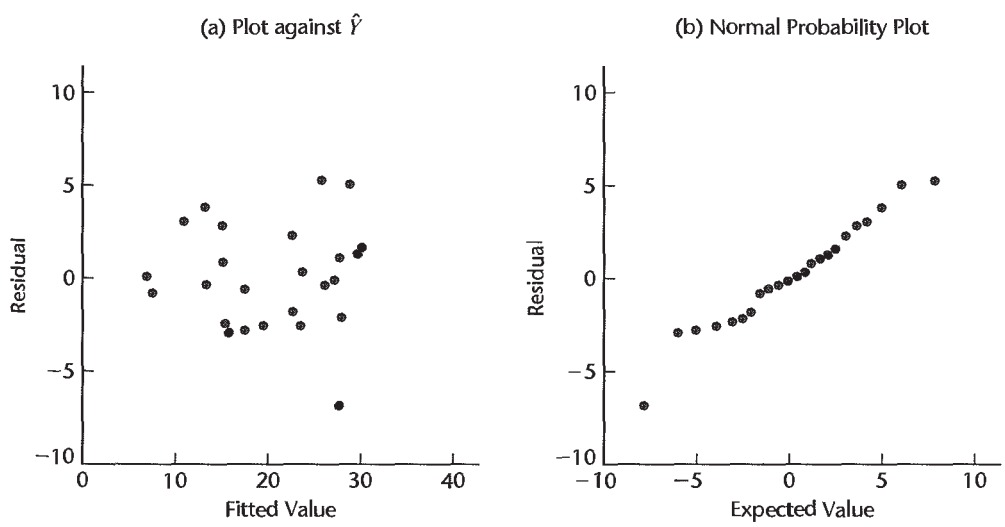
$$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$$

$$H_a: \text{not all } \tau_k \text{ equal zero}$$

TABLE 28.6 ANOVA Table—Background Mnsic Example.

Source of Variation	*SS*	*df*	*MS*
Weeks	82.0	4	20.5
Days within week	477.2	4	119.3
Type of music	664.4	4	166.1
Error	188.4	12	15.7
Total	1,412.0	24	

FIGURE 28.5 Diagnostic Residual Plots—Background Music Example.



we find from Table 28.6:

$$F^* = \frac{MSTR}{MSRem} = \frac{166.1}{15.7} = 10.6$$

To control the risk of making a Type I error at $\alpha = .01$, we require $F(.99; 4, 12) = 5.41$. Since $F^* = 10.6 > 5.41$, we conclude H_a, that the various types of background music have differential effects on the productivity of the bank tellers. The P-value of this test is .0007.

Pairwise comparisons between the different kinds of music were desired with a family confidence coefficient of .90, using the Tukey procedure. Substituting into (17.14) with $n_i = n_{i'} = r$ and using *MSRem* from Table 28.6 as the mean square, we obtain:

$$s^2\{\hat{L}\} = \frac{2MSRem}{r} = \frac{2(15.7)}{5} = 6.28 \qquad s\{\hat{L}\} = 2.51$$

Remember that each estimated treatment mean $\bar{Y}_{\cdot k}$ is based on five observations here. Next, we require the T multiple in (28.21b):

$$T = \frac{1}{\sqrt{2}}q(.90; 5, 12) = \frac{1}{\sqrt{2}}(3.92) = 2.77$$

so that:

$$Ts\{\hat{L}\} = 2.77(2.51) = 6.95$$

Conducting pairwise tests based on the confidence intervals, the treatments can be placed into three groups:

Group 1		Group 2		Group 3	
Music 2	$\bar{Y}_{\cdot 2} = 26.6$	Music 4	$\bar{Y}_{\cdot 4} = 23.8$	Music 1	$\bar{Y}_{\cdot 1} = 11.4$
Music 4	$\bar{Y}_{\cdot 4} = 23.8$	Music 5	$\bar{Y}_{\cdot 5} = 21.6$		
Music 5	$\bar{Y}_{\cdot 5} = 21.6$	Music 3	$\bar{Y}_{\cdot 3} = 19.6$		

The most promising treatment appears to be mixed instrumental-vocal music in medium tempo ($k = 2$). There is clear evidence that it is better than instrumental-vocal music in slow tempo ($k = 1$) or instrumental-vocal music in fast tempo ($k = 3$). The point estimates suggest it also is better than solely instrumental music in medium ($k = 4$) or fast ($k = 5$) tempo, but the experimental evidence on these latter two comparisons is inconclusive.

Factorial Treatments

If the treatments in a latin square design are factorial in nature, the treatment sum of squares *SSTR* is decomposed in the usual manner. For a two-factor experiment involving factors A and B, we have:

$$SSTR = SSA + SSB + SSAB \tag{28.22}$$

Estimates of fixed factor effects can be made readily since they are simply contrasts of the treatment means.

Random Blocking Variable Effects

If the row and/or column blocking variable(s) in a latin square design have classes that should be viewed as random selections from a population, the fixed effects latin square model (28.12) needs to be modified in the usual fashion. The analysis of variance is the same as for the fixed blocking variable effects model and all tests and estimates of treatment effects are conducted as for fixed blocking variable effects.

Missing Observations

While missing observations destroy the symmetry (orthogonality) of the latin square design and make the usual ANOVA calculations inappropriate, the regression approach ordinarily remains appropriate when observations in a latin square design are missing. We just set up the regression model for the available observations and then fit the model to the data. The procedure is analogous to that discussed in Section 23.4 for complete block designs. Tests are conducted by fitting the full and appropriate reduced regression models. Estimation of fixed treatment effects is done in terms of the regression coefficients for the full model in the usual manner.

28.6 Planning Latin Square Experiments

Power of *F* Test

The power of the F test for treatment effects in latin square model (28.12) involves the noncentrality parameter:

$$\phi = \frac{1}{\sigma}\sqrt{\sum \tau_k^2} \tag{28.23}$$

with degrees of freedom $r - 1$ for the numerator and $(r - 1)(r - 2)$ for the denominator. Other than these modifications, no new issues are encountered in obtaining the power of the test for treatment effects in a latin square design.

Necessary Number of Replications

A latin square design provides r replications for each treatment. Power and/or estimation considerations similar to those for randomized complete block designs may indicate that r replications are too few, particularly when r is small, say, 3, 4, or 5. Two methods of increasing the number of replications with a latin square design are discussed in Section 28.7. With either method, it is necessary to assess in advance the magnitude of the experimental error variance σ^2 in order to plan the necessary number of replications.

Efficiency of Blocking Variables

The efficiency of a latin square design can be assessed relative to a completely randomized design or relative to a randomized complete block design. The efficiency relative to a completely randomized design is defined by:

$$E_1 = \frac{\sigma_r^2}{\sigma_L^2} \tag{28.24a}$$

where σ_r^2 and σ_L^2 are the experimental error variances with a completely randomized design and a latin square design, respectively. The efficiency relative to a randomized complete

block design can be measured in two ways, depending on whether the row or the column blocking variable is used in the randomized block design:

$$E_2 = \frac{\sigma_{br}^2}{\sigma_L^2} \tag{28.24b}$$

$$E_3 = \frac{\sigma_{bc}^2}{\sigma_L^2} \tag{28.24c}$$

where σ_{br}^2 and σ_{bc}^2 are the experimental error variances with a randomized block design if the row blocking variable or the column blocking variable is utilized, respectively.

We can estimate σ_r^2, σ_{br}^2, and σ_{bc}^2 from the results for a latin square design as follows:

$$s_r^2 = \frac{MSROW + MSCOL + (r-1)MSRem}{r+1} \tag{28.25a}$$

$$s_{br}^2 = \frac{MSCOL + (r-1)MSRem}{r} \tag{28.25b}$$

$$s_{bc}^2 = \frac{MSROW + (r-1)MSRem}{r} \tag{28.25c}$$

Thus, the estimated measures of efficiency are:

$$\hat{E}_1 = \frac{MSROW + MSCOL + (r-1)MSRem}{(r+1)MSRem} \tag{28.26a}$$

$$\hat{E}_2 = \frac{MSCOL + (r-1)MSRem}{rMSRem} \tag{28.26b}$$

$$\hat{E}_3 = \frac{MSROW + (r-1)MSRem}{rMSRem} \tag{28.26c}$$

When r is small, the efficiency measures may be modified by means of (21.15) to account for differences in the number of degrees of freedom associated with the mean squares used for estimating the experimental error variances for the two designs being compared.

Example

For the background music example, we obtain the following efficiency measures from the results in Table 28.6:

$$\hat{E}_1 = \frac{20.5 + 119.3 + 4(15.7)}{6(15.7)} = 2.2$$

$$\hat{E}_2 = \frac{119.3 + 4(15.7)}{5(15.7)} = 2.3$$

$$\hat{E}_3 = \frac{20.5 + 4(15.7)}{5(15.7)} = 1.1$$

We see that the latin square design was efficient relative to a completely randomized design. The latter would have required over twice as many replications for each treatment as the latin square design so that the variance for any specified estimated treatment contrast would be the same with both designs. Most of this efficiency was gained by the column blocking variable (days within week), because the efficiency of the latin square design relative to a complete block design with the column blocking variable is poor, being close to 1. Hence, little was achieved by also blocking by the row blocking variable (week).

28.7 Additional Replications with Latin Square Designs

A latin square design, as noted earlier, provides r replications for each treatment. If power and/or estimation considerations indicate that these are too few replications, two basic methods are available for increasing the number of replications—replications within cells and additional latin squares. We consider each in turn.

Replications within Cells

This method of increasing the replications per treatment is feasible when two or more experimental units can be obtained for each cell defined by the row and column blocking variables. Consider, for instance, an experiment in which IQ (low, normal, high) and age (young, middle, old) are the blocking variables. In this type of situation, it is possible to obtain two or more experimental subjects for each cell, and each of the subjects in a cell will then receive the treatment assigned to that cell by the latin square employed.

Let n denote the number of experimental units available for each cell, and let Y_{ijkm} denote the observation for the mth unit ($m = 1, \ldots, n$) in the (i, j) cell for which the assigned treatment is k. The additive fixed effects model (28.12) is modified for the n replications in each cell as follows:

$$Y_{ijkm} = \mu_{\cdots} + \rho_i + \kappa_j + \tau_k + \varepsilon_{ijkm} \tag{28.27}$$

where:

$\mu_{\cdots}$ is a constant

ρ_i, κ_j, τ_k are constants subject to the restrictions $\sum \rho_i = \sum \kappa_j = \sum \tau_k = 0$

ε_{ijkm} are independent $N(0, \sigma^2)$

$i = 1, \ldots, r; j = 1, \ldots, r; k = 1, \ldots, r; m = 1, \ldots, n$

The ANOVA sums of squares and degrees of freedom for model (28.27) can be obtained by the rules in Appendix D, remembering that one subscript is redundant. The treatment, row, and column sums of squares are, respectively:

$$SSTR = rn \sum_k (\bar{Y}_{\cdot\cdot k\cdot} - \bar{Y}_{\cdots})^2 \tag{28.28a}$$

$$SSROW = rn \sum_i (\bar{Y}_{i\cdots} - \bar{Y}_{\cdots})^2 \tag{28.28b}$$

$$SSCOL = rn \sum_j (\bar{Y}_{\cdot j\cdot\cdot} - \bar{Y}_{\cdots})^2 \tag{28.28c}$$

The total sum of squares as usual is:

$$SSTO = \sum_i \sum_j \sum_m (Y_{ijkm} - \bar{Y}_{\cdots})^2 \tag{28.28d}$$

while $SSRem$ is obtained as a remainder:

$$SSRem = SSTO - SSROW - SSCOL - SSTR \tag{28.28e}$$

The degrees of freedom for row, column, and treatment sums of squares are unchanged, while those associated with $SSRem$ are increased from $(r-1)(r-2)$ to $nr^2 - 3r + 2$, an increase of $(n-1)r^2$ degrees of freedom.

TABLE 28.7 ANOVA Table for Latin Square Design Model (28.27) with n Replications per Cell.

Source of Variation	*SS*	*df*	*MS*
Row blocking variable	*SSROW*	$r-1$	*MSROW*
Column blocking variable	*SSCOL*	$r-1$	*MSCOL*
Treatments	*SSTR*	$r-1$	*MSTR*
Error	*SSRem*	nr^2-3r+2	*MSRem*
Total	*SSTO*	nr^2-1	

The analysis of variance is shown in Table 28.7. The expected mean squares can be obtained by the rules in Appendix D, remembering that one subscript is redundant, or from a suitable computer package. The test statistic for testing treatment effects is again $F^* = MSTR/MSRem$.

When n replications are present within a cell for a latin square, it is possible to obtain a pure error measure and conduct a test for lack of fit of model (28.27) in the usual manner.

Example

A state university, while developing a retraining program to teach general computer repair skills to persons displaced from their previous occupations, conducted an experiment to evaluate the effects of three different incentive methods on achievement during the program. The blocking variables were IQ and age of subject. Two replications per cell were utilized. Table 28.8a contains the achievement scores for the participants in the experiment, while Table 28.8b contains the analysis of variable table obtained from a computer package.

To test the appropriateness of additive model (28.27), we use the usual test statistic for lack of fit:

$$F^* = \frac{MSLF}{MSPE} = \frac{8.2}{4.0} = 2.05$$

For level of significance $\alpha = .05$, we need $F(.95; 2, 9) = 4.26$. Since $F^* = 2.05 \leq 4.26$, we conclude that additive model (28.27) is appropriate here. The P-value of the test is .18. The comparison of the three incentive methods was then carried out in the usual fashion.

Additional Latin Squares

At times, it is not possible to obtain additional experimental units within a cell. This is the case, for instance, in the background music example of Table 28.4, where only one type of music can be played in one day in a bank. When it is not possible to replicate within cells, additional replications for each treatment frequently can be obtained by adding one or more latin squares to one of the blocking variables. In the background music example of Table 28.4, for instance, the experiment could be run for another five weeks. In an experiment using plant crews as experimental units and employing as blocking variables plant shift (morning, afternoon, evening) and production department (1, 2, 3), additional replications can be obtained by running the experiment in other production departments.

The layout for the background music example of Table 28.4, when run over another five weeks, is shown in Table 28.9. The second latin square, and additional ones when required, is selected independently of the first.

TABLE 28.8 Example of Latin Square Design with Two Replications per Cell—Retraining Program Experiment.

(a) Data

IQ i	Age (j) Young	Middle	Old
	(*B*)	(*A*)	(C)
High	19	20	25
	16	24	21
	(C)	(*B*)	(*A*)
Normal	24	14	14
	22	15	14
	(*A*)	(C)	(*B*)
Low	10	12	7
	14	13	4

(b) Analysis of Variance

Source of Variation	*SS*	*df*	*MS*
IQ	364.3	2	182.2
Age	34.3	2	17.2
Treatments	147.0	2	73.5
Error	52.4	11	4.76
Lack of fit	16.4	2	8.2
Pure error	36.0	9	4.0
Total	598.0	17	

TABLE 28.9 Two-Latin-Squares Design—Background Music Example of Table 28.4.

Square	Week	Day M	T	W	Th	F
	1	*D*	C	*A*	*B*	*E*
	2	C	*B*	*E*	*A*	*D*
1	3	*A*	*D*	*B*	*E*	C
	4	*E*	*A*	C	*D*	*B*
	5	*B*	*E*	*D*	C	*A*
	6	*E*	*D*	C	*A*	*B*
	7	*B*	*A*	*E*	*D*	C
2	8	*D*	C	*A*	*B*	*E*
	9	*A*	*E*	*B*	C	*D*
	10	C	*B*	*D*	*E*	*A*

Frequently, the additional squares may be viewed as classes of a third blocking variable. For instance, in the background music example of Table 28.9, the two latin squares may be considered to be two levels of the blocking variable "time period." The first five weeks may be viewed as time period 1, and the second five weeks as time period 2. As another example, in the experiment with plant crews mentioned previously, the production departments for the first latin square may be on an hourly rate, while the departments for the second latin square may be on incentive pay. Thus, with additional latin squares, one can, in effect, introduce a third blocking variable. As a consequence, the variation associated with the third blocking variable can be removed from the experimental error variability. In addition, the interactions between the third blocking variable and the other variables can be studied.

28.8 Replications in Repeated Measures Studies

We noted earlier that a latin square design is highly suitable for repeated measures studies when there are r treatments and r subjects. If additional replications are needed, however, replications within cells cannot be used since a cell pertains to an individual subject. Instead, latin square crossover designs or independent latin squares may be used.

Latin Square Crossover Designs

These designs, also called *latin square changeover designs*, are often useful when a latin square is to be used in a repeated measures study to balance the order positions of treatments, yet more subjects are required than called for by a single latin square. With this type of design, the subjects are randomly assigned to the different treatment order patterns given by a latin square (several latin squares may be used at times). Consider an experiment in which treatments A, B, and C are to be administered to each subject, and the three treatment order patterns are given by the latin square:

	Order Position		
Pattern	**1**	**2**	**3**
1	A	B	C
2	B	C	A
3	C	A	B

Suppose that $3n$ subjects are available for the study. Then n subjects will be assigned at random to each of the three order patterns in a latin square crossover design. Note that this design is a mixture of repeated measures (within subjects) and latin square (order patterns form a latin square).

Assuming that all effects are additive and fixed except that the effects for subjects are random, a relatively simple model for latin square crossover designs can be developed for r treatments and n subjects per order pattern. In the following model, ρ_i denotes the effect of the ith treatment order pattern, κ_j denotes the effect of the jth order position, τ_k denotes the effect of the kth treatment, and $\eta_{m(i)}$ denotes the effect of subject m which is nested

within the ith treatment order pattern:

$$Y_{ijkm} = \mu_{...} + \rho_i + \kappa_j + \tau_k + \eta_{m(i)} + \varepsilon_{ijkm} \tag{28.29}$$

where:

$\mu_{...}$ is a constant
ρ_i, κ_j, τ_k are constants subject to the restrictions $\sum \rho_i = \sum \kappa_j = \sum \tau_k = 0$
$\eta_{m(i)}$ are independent $N(0, \sigma_\eta^2)$
ε_{ijkm} are independent $N(0, \sigma^2)$ and independent of the $\eta_{m(i)}$
$i = 1, \ldots, r; j = 1, \ldots, r; k = 1, \ldots, r; m = 1, \ldots, n$

The analysis of variance sums of squares, degrees of freedom, and expected mean squares for this model can be obtained by the rules in Appendix D, remembering that one subscript is redundant. The formulas for the sums of squares follow the usual pattern:

$$SSTO = \sum_i \sum_j \sum_m (Y_{ijkm} - \bar{Y}_{...})^2 \tag{28.30a}$$

$$SSP = nr \sum_i (\bar{Y}_{i...} - \bar{Y}_{...})^2 \tag{28.30b}$$

$$SSO = nr \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...})^2 \tag{28.30c}$$

$$SSTR = nr \sum_k (\bar{Y}_{..k} - \bar{Y}_{...})^2 \tag{28.30d}$$

$$SSS = r \sum_i \sum_m (\bar{Y}_{i..m} - \bar{Y}_{i...})^2 \tag{28.30e}$$

$$SSRem = SSTO - SSP - SSO - SSTR - SSS \tag{28.30f}$$

Here, SSP is the (treatment) *pattern sum of squares*, SSO is the *order position sum of squares*, SSS is the *subject sum of squares*, and the other sums of squares have their usual meanings. Table 28.10 contains the ANOVA table.

TABLE 28.10 **ANOVA Table for Latin Square Crossover Design Model (28.29).**

Source of Variation	SS	df	MS	$E\{MS\}$
Patterns (P)	SSP	$r-1$	MSP	$\sigma^2 + r\sigma_\eta^2 + nr\frac{\sum \rho_i^2}{r-1}$
Order positions (O)	SSO	$r-1$	MSO	$\sigma^2 + nr\frac{\sum \kappa_j^2}{r-1}$
Treatments (TR)	$SSTR$	$r-1$	$MSTR$	$\sigma^2 + nr\frac{\sum \tau_k^2}{r-1}$
Subjects (S) (within patterns)	SSS	$r(n-1)$	MSS	$\sigma^2 + r\sigma_\eta^2$
Error	$SSRem$	$(r-1)(nr-2)$	$MSRem$	σ^2
Total	$SSTO$	nr^2-1		

TABLE 28.11 Latin Square Crossover Design—Apple Sales Example.

(a) Data (coded)

Pattern i	Store	Two-Week Period (j) 1	2	3
1	$m=1$	9 (*B*)	12 (*C*)	15 (*A*)
	$m=2$	4 (*B*)	12 (*C*)	9 (*A*)
2	$m=1$	12 (*A*)	14 (*B*)	3 (*C*)
	$m=2$	13 (*A*)	14 (*B*)	3 (*C*)
3	$m=1$	7 (*C*)	18 (*A*)	6 (*B*)
	$m=2$	5 (*C*)	20 (*A*)	4 (*B*)

(b) Analysis of Variance

Source of Variation	*SS*	*df*	*MS*
Patterns	.33	2	.17
Order positions	233.33	2	116.67
Displays	189.00	2	94.50
Stores (within patterns)	21.00	3	7.00
Error	20.33	8	2.54
Total	464.0	17	

Example

Table 28.11a contains data for a study of the effects of three different displays on the sale of apples, using a latin square crossover design. Six stores were used, with two assigned at random to each of the three treatment order patterns shown. Each display was kept for two weeks, and the observed variable was sales per 100 customers. Table 28.11b contains the analysis of variance. The sums of squares were obtained from a computer run.

To test for treatment (display) effects, we use:

$$F^* = \frac{MSTR}{MSRem} = \frac{94.5}{2.54} = 37.2$$

For $\alpha = .05$, we require $F(.95; 2, 8) = 4.46$. Since $F^* = 37.2 > 4.46$, we conclude that there are differential sales effects for the three displays. The P-value of the test is 0+. Tests for pattern effects, order position effects, and store effects were also carried out. They indicated that order position effects were present, but no pattern or store effects. Order position effects here are associated with the three time periods in which the displays were studied, and may reflect seasonal effects as well as the results of special events, such as unusually hot weather in one period. The comparison of the three treatment effects was then carried out in the usual fashion.

Use of Independent Latin Squares

If the order position effects are not approximately constant for all subjects (stores, etc.), a crossover design is not fully effective. It may then be preferable to place the subjects into homogeneous groups with respect to the order position effects and use independent latin

squares for each group. Suppose that four treatments are to be administered to eight subjects each, four males and four females, and that the experimenter expects the fatigue effect to be stronger for females than for males. The use of two independent latin squares, one for male subjects and the other for female subjects, may then be advisable.

Carryover Effects

If carryover effects from one treatment to another are anticipated, that is, if not only the order position but also the preceding treatment has an effect, these carryover effects may be balanced out by choosing a latin square in which every treatment follows every other treatment an equal number of times. For $r = 4$, an example of such a latin square is:

	Period			
Subject	**1**	**2**	**3**	**4**
1	*A*	*B*	*D*	*C*
2	*B*	*C*	*A*	*D*
3	*C*	*D*	*B*	*A*
4	*D*	*A*	*C*	*B*

Note that treatment *A* follows each of the other treatments once, and similarly for the other treatments. This design is appropriate when the carryover effects do not persist for more than one period.

When r is odd, the sequence balance can be obtained by using a pair of latin squares with the property that the treatment sequences in one square are reversed in the other square. Indeed, even when r is even, it is usually desirable to use a pair of such squares so that the degrees of freedom associated with *MSRem* are reasonably large. Such a design is sometimes called a *latin square double crossover design.* This type of design retains the advantages of employing two blocking variables in a latin square, while enabling the experimenter also to balance and measure the carryover effects.

For the earlier apple display illustration in which three displays were studied in six stores, the two latin squares might be as shown in Table 28.12. The stores should first be placed into two homogeneous groups and these should then be assigned to the two latin squares.

TABLE 28.12 Illustration of a Latin Square Double Crossover Design.

		Two-Week Period		
Square	**Store**	**1**	**2**	**3**
1	1	*A*	*B*	*C*
	2	*B*	*C*	*A*
	3	*C*	*A*	*B*
2	4	*A*	*C*	*B*
	5	*B*	*A*	*C*
	6	*C*	*B*	*A*

Cited References

28.1. Cochran, W. G., and G. M. Cox. *Experimental Designs*. 2nd ed. New York: John Wiley & Sons, 1957.

28.2. Cook, R. D., and C. J. Nachtsheim. "Computer-Aided Blocking of Factorial and Response Surface Designs," *Technometrics* 31 (1989), pp. 339–346.

28.3. Dean, A., and D. Voss. *Design and Analysis of Experiments*. New York: Springer-Verlag, 1999.

28.4. Snedecor, G. W., and W. G. Cochran. *Statistical Methods*. 8th ed. Ames, Iowa: The Iowa State University Press, 1989.

Problems

28.1. Discuss the advantages and disadvantages of balanced incomplete block designs in comparison to randomized complete block designs.

28.2. What is meant by *balance* in a balanced incomplete block design? What are the advantages of balance? Under what circumstances might the use of an unbalanced incomplete block design be justified?

28.3. Construct a balanced incomplete block design for three treatments in blocks of size two. How many blocks n_b are required? What are n and n_p for your design?

28.4. Construct a balanced incomplete block design for seven treatments in blocks of size five. How many blocks n_b are required? What are n and n_p for your design?

28.5. Construct a balanced incomplete block design for eight treatments in blocks of size three. How many blocks n_b are required? What are n and n_p for your design?

28.6. **Detergent effectiveness.** A chemical engineer wished to evaluate the effectiveness of nine alternative formulations of a dishwashing detergent in terms of the extent to which each would maintain foam or suds while in use. Three sinks were available, and three people were instructed to use the sinks to wash plates at a constant rate. Each block consisted of three experimental units, where the experimental unit was a sink with a fixed amount of clean water and a fixed amount of soil added. Three detergent formulations were randomly assigned to the three sinks in each block. The response Y was foam duration, which was measured by the number of plates washed before the suds disappeared. BIBD number 18 from Table 28.1 was utilized for this experiment. Data for the randomized BIBD follow:

	Treatments			Responses		
Block	Sink 1	Sink 2	Sink 3	Sink 1	Sink 2	Sink 3
1	3	8	4	13	20	7
2	4	9	2	6	29	17
3	3	6	9	15	23	31
4	9	5	1	31	26	20
5	2	7	6	16	21	23
6	6	5	4	23	26	6
7	9	8	7	28	19	21
8	7	1	4	20	20	7
9	6	8	1	24	19	20
10	5	8	2	26	19	17
11	5	3	7	24	14	19
12	3	2	1	11	17	19

John, P. W. M. "An Application of a Balanced Incomplete Block Design," *Technometrics* 3 (1961), pp. 51–54.

Obtain the residuals for balanced incomplete block design model (28.2) and plot them against the fitted values. Also prepare a normal probability plot of the residuals and calculate the

coefficient of correlation between the ordered residuals and their expected values under normality. Summarize your findings about the appropriateness of model (28.2) here.

28.7. Refer to **Detergent Effectiveness** Problem 28.6. Assume that balanced incomplete block design model (28.2) is appropriate.

a. Obtain the least squares estimates of the treatment means and plot them against treatment number in the form of Figure 28.4. Does your plot suggest the presence of treatment effects?

b. Test whether or not treatment affects foam duration; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Test whether or not block effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Give a 95 percent confidence interval for the fifth treatment mean.

e. Analyze the nature of the treatment effects by making all pairwise comparisons among the treatment means. Use the Tukey procedure and a 90 percent family confidence coefficient. Summarize your findings using a line plot of the least squares treatment means.

*28.8. **Automobile tire wear.** An automotive engineer wished to evaluate the effects of four rubber compounds on the life of automobile tires. The manufacturing process permitted the use of up to three different compounds in a given tire. To do this, the tire is divided into three sections, and a different compound is used in each section. Because each segment of a tire would be subject to nearly identical road conditions, the investigator decided to use tires as blocks, with three of the four treatments (compounds) being applied to the three experimental units (tire segments) in each block. Four tires were tested. The response Y is a coded measure of wear. Design 2 from Table 28.1 was utilized; the experimental layout and response data follow:

	Compound			
Tire	A	B	C	D
1	238	238	279	
2	196	213		308
3	254		334	367
4		312	421	412

Davies, O. L., ed. *The Design and Analysis of Industrial Experiments*, London: Oliver and Boyd (1961)

Obtain the residuals for balanced incomplete block design model (28.2) and plot them against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Summarize your findings about the appropriateness of model (28.2) here.

*28.9. Refer to **Automobile tire wear** Problem 28.8. Assume that balanced incomplete block design model (28.2) is appropriate.

a. Obtain the least squares estimates of the treatment means and plot them against treatment number in the form of Figure 28.4. Does your plot suggest the presence of treatment effects?

b. Test whether or not the type of compound affects tire wear; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Test whether or not block effects are present; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

d. Give a 95 percent confidence interval for the mean wear for compound A.

e. Analyze the nature of the treatment effects by making all pairwise comparisons among the treatment means. Use the Tukey procedure and a 95 percent family confidence coefficient. Summarize your findings using a line plot of the least squares treatment means.

*28.10. Suppose that Tukey's method for all pairwise comparisons will be made using balanced incomplete block design number 2 in Table 28.1. Assume that σ^2 will be no larger than 2.0 and the widths of the simultaneous 95 percent confidence intervals are not to exceed 3.0. Determine n, the number of replicates, and n_b, the number of blocks, necessary to satisfy these requirements. How many repeats of design number 2 are required?

28.11. Suppose that Tukey's method for all pairwise comparisons will be made using balanced incomplete block design number 5 in Table 28.1. Assume that σ^2 will be no larger than 1.5 and the widths of the simultaneous 90 percent confidence intervals are not to exceed 2.5. Determine n, the number of replicates, and n_b, the number of blocks, necessary to satisfy these requirements. How many repeats of design number 5 are required?

28.12. A behavioral scientist explained why latin square designs are used so frequently: "Many times in behavioral science, we require the use of repeated measures designs because variability between human subjects is so great. Since an order effect may be present in this situation, we employ latin square designs to eliminate any bias due to order effects." Comment.

28.13. a. Using random permutations, select randomly a 3 by 3 latin square. Show all steps.

b. Using random permutations, select randomly a 6 by 6 latin square. Show all steps.

*28.14. **Hardware sales.** A manufacturer conducted a small pilot study of the effect of the price of one of its products on sales of this product in hardware stores. Since it might be confusing to customers if prices were switched repeatedly within a store, only one price was used for any one store during the six-month study period. Sixteen stores were employed in the study. To reduce experimental error variability, stores were chosen so that there would be one store for each sales volume-geographic location class. The four price levels (A: \$1.79; B: \$1.69; C: \$1.59; D: \$1.49) were assigned to the stores according to the latin square design shown below. Data on sales during the six-month period (in thousand dollars) follow.

Sales Volume Class	Geographic Location Class (j)			
i	Northeast	Northwest	Southeast	Southwest
1 (smallest)	1.2 (B)	1.5 (C)	1.0 (A)	1.7 (D)
2	1.4 (A)	1.9 (D)	1.6 (B)	1.5 (C)
3	2.8 (C)	2.1 (B)	2.7 (D)	2.0 (A)
4 (largest)	3.4 (D)	2.5 (A)	2.9 (C)	2.7 (B)

Obtain the residuals for latin square model (28.12) and plot them against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Summarize your findings about the appropriateness of model (28.12) here.

*28.15. Refer to **Hardware sales** Problem 28.14. Assume that latin square model (28.12) is appropriate.

a. Prepare a main effects plot of the estimated treatment means. What does the plot suggest about the effects of the four price levels on sales?

b. Test whether or not price level affects mean sales; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Analyze the nature of the price effect on sales by making all pairwise comparisons among the treatment means. Use the Tukey procedure and a 90 percent family confidence coefficient. Summarize your findings.

d. Does there appear to be a linear relationship between price level and mean sales? Could you formally test for linearity? Explain.

*28.16. Refer to **Hardware sales** Problems 28.14 and 28.15.

a. Calculate the three estimated efficiency measures in (28.26).

b. Would a randomized complete block design have been adequate here? If so, which blocking variable would have been best?

28.17. **Summary reports.** A management information systems consultant conducted a small-scale study of five different daily summary reports (A: greatest amount of detail; B; C; D; E: least amount of detail). Five sales executives were used in the study. Each was given one type of daily report for a month and then was asked to rate its helpfulness on a 25-point scale (0: no help; 25: extremely helpful). Over a five-month period, each executive received each type of report for one month according to the latin square design shown below. The helpfulness ratings follow.

Executive i	Month (j) March	April	May	June	July
Harrison	21 (D)	8 (A)	17 (C)	9 (B)	16 (E)
Smith	5 (A)	10 (E)	3 (B)	12 (C)	15 (D)
Carmichael	20 (C)	10 (B)	15 (E)	22 (D)	12 (A)
Loeb	4 (B)	17 (D)	3 (A)	9 (E)	10 (C)
Munch	17 (E)	16 (C)	20 (D)	7 (A)	11 (B)

Obtain the residuals for latin square model (28.12) and plot them against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Summarize your findings about the appropriateness of model (28.12) here.

28.18. Refer to **Summary reports** Problem 28.17. Assume that latin square model (28.12) is appropriate.

a. Prepare a main effects plot of the estimated treatment means. What does the plot suggest about the effects of the five types of reports?

b. Test whether or not the five types of reports differ in mean helpfulness; use significance level $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Analyze the effectiveness of the five types of reports by making all pairwise comparisons among the treatment means. Use the Tukey procedure and a 95 percent family confidence coefficient. Summarize your findings.

28.19. Refer to **Summary reports** Problems 28.17 and 28.18.

a. Calculate the three estimated efficiency measures in (28.26).

b. How effective was the use of the latin square design here?

*28.20. Refer to **Hardware sales** Problems 28.14 and 28.15. Assume that $\sigma = .15$. What is the power of the test for treatment effects in Problem 28.15b if $\tau_1 = -.4$, $\tau_2 = 0$, $\tau_3 = .1$, and $\tau_4 = .3$?

28.21. Refer to **Summary reports** Problems 28.17 and 28.18. Assume that $\sigma = 1.4$. What is the power of the test for treatment effects in Problem 28.18b if $\tau_1 = -2$, $\tau_2 = -1$, $\tau_3 = 0$, $\tau_4 = 1.5$, $\tau_5 = 1.5$?

28.22. **Drugs interaction.** A pilot study was undertaken on the interaction effects of two drugs to stimulate growth in girls who are of short stature because of a particular syndrome. Each drug was known to be modestly effective singly, but the combination of the two drugs had never been investigated. Blocking by both subject and time period was desired whereby repeated measures for different treatments applied to the same subject are obtained. A 4 by 4 latin square design, shown below, was utilized for four subjects, four time periods, and four treatments. The four time periods consisted of one month each, separated by an intervening month during which no treatment was given. The four treatments were A: no treatment (placebo); B: drug X alone; C: drug Y alone; D: both drugs X and Y. The response variable was the difference in the growth rates (in centimeters per month) during the treatment period and the base period before the experiment began. The results of the study follow.

Subject	Period (j)			
i	1	2	3	4
1	.02 (A)	.15 (B)	.45 (D)	.18 (C)
2	.27 (B)	.24 (C)	−.01 (A)	.58 (D)
3	.11 (C)	.35 (D)	.14 (B)	−.03 (A)
4	.48 (D)	.04 (A)	.18 (C)	.22 (B)

Obtain the residuals in (28.16) for latin square model (28.12) and plot them against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Summarize your findings.

28.23. Refer to **Drugs interaction** Problem 28.22. Assume that an appropriate model is latin square model (28.12), modified so that subjects have random effects and a factorial structure for the treatments is incorporated (factor A: drug X; factor B: drug Y).

a. State the model to be employed.

b. Test for interaction effects between the two drugs; use $\alpha = .10$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Estimate the interaction contrast:

$$L = \left(\frac{\mu_{\cdot\cdot 2} + \mu_{\cdot\cdot 3}}{2} - \mu_{\cdot\cdot 1}\right) - \left(\mu_{\cdot\cdot 4} - \frac{\mu_{\cdot\cdot 2} + \mu_{\cdot\cdot 3}}{2}\right) = \mu_{\cdot\cdot 2} - \mu_{\cdot\cdot 1} - \mu_{\cdot\cdot 4} + \mu_{\cdot\cdot 3}$$

using a 90 percent confidence interval. Interpret your result.

*28.24. Refer to **Hardware sales** Problem 28.14.

a. Set up the regression model equivalent to latin square model (28.12) using 1, −1, 0 indicator variables.

b. Test by means of the regression approach whether or not price level affects mean sales; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

c. Obtain a 95 percent confidence interval by the regression approach for $L = \tau_3 - \tau_4$. Interpret your interval estimate.

d. Suppose that observation $Y_{232} = 1.6$ were missing.

i. Use the regression approach to test whether price level affects mean sales; control the α risk at .05. State the alternatives, decision rule, and conclusion.

ii. Use the regression approach to estimate $L = \tau_1 - \tau_2$ by means of a 95 percent confidence interval.

28.25. Refer to **Summary reports** Problem 28.17. Suppose that observations $Y_{114} = 21$ and $Y_{453} = 10$ were missing.

a. Use the regression approach to test whether the five types of reports differ in mean effectiveness; employ significance level $\alpha = .01$. State the alternatives, decision rule, and conclusion.

b. Use the regression approach to estimate $L = \tau_4 - \tau_1$ by means of a 99 percent confidence interval.

28.26. **TV commercials.** A study was undertaken to determine whether the volume of sound of a television commercial affects recall and whether this effect varies by product. Thirty-two subjects were chosen, two each for 16 groups defined according to age (class 1: youngest; 2; 3; 4: oldest) and amount of education (class 1: lowest education level; 2; 3; 4: highest education level). Each subject was exposed to one of four television commercial showings (A: high volume, product X; B: low volume, product X; C: high volume, product Y; D: low volume, product Y) according to the latin square design shown below. Two different commercials were involved, one for each product. During the following week, the subjects were asked to mention everything they could remember about the advertisement. Scores were based on the number of learning points mentioned, suitably standardized. The results follow.

Age Class i:	1	2	3	4
Education Level				
$j = 1$:	83 86 (D)	70 76 (B)	67 74 (C)	56 60 (A)
$j = 2$:	64 69 (A)	81 75 (C)	67 61 (B)	72 67 (D)
$j = 3$:	78 75 (C)	64 60 (A)	76 81 (D)	63 67 (B)
$j = 4$:	76 74 (B)	87 81 (D)	64 57 (A)	64 66 (C)

Obtain the residuals for latin square model (28.27) and plot them against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Summarize your findings about the appropriateness of the model utilized here.

28.27. Refer to **TV commercials** Problem 28.26. Assume that latin square model (28.27), modified to allow for factorial treatments (factor A: volume; factor B: product), is appropriate.

a. State the model to be employed.

b. Test for volume-product interaction effects; use $\alpha = .01$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Test for volume main effects and product main effects. For each test, use $\alpha = .01$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

d. To study the nature of the volume and product main effects, estimate the difference between the two factor level means for each factor. Use the Bonferroni procedure and a 95 percent family confidence coefficient. State your findings.

28.28. **Recall decay.** In an experiment to study recall decay with three different questionnaires (A, B, C), nine subjects were questioned at three different times three months apart about the number of trips to a shopping center during the preceding three months. Each time a different questionnaire was used. The latin square design shown on the following page used to determine the questionnaire order for each subject, with three subjects assigned randomly to each of the three treatment order patterns. The data on number of shopping trips reported follow.

Pattern i	Subject	Time Period (j) 1	2	3
	$m=1$	40 (C)	18 (A)	30 (B)
1	$m=2$	35 (C)	25 (A)	37 (B)
	$m=3$	31 (C)	22 (A)	28 (B)
	$m=1$	10 (B)	43 (C)	33 (A)
2	$m=2$	18 (B)	49 (C)	37 (A)
	$m=3$	15 (B)	48 (C)	29 (A)
	$m=1$	7 (A)	19 (B)	59 (C)
3	$m=2$	11 (A)	24 (B)	51 (C)
	$m=3$	19 (A)	21 (B)	62 (C)

Obtain the residuals for latin square crossover model (28.29) and plot them against the fitted values. Also prepare a normal probability plot of the residuals and calculate the coefficient of correlation between the ordered residuals and their expected values under normality. Summarize your findings about the appropriateness of model (28.29) here.

28.29. Refer to **Recall decay** Problem 28.28. Assume that latin square crossover model (28.29) is appropriate.

a. Test for the presence of treatment order pattern, time period, and questionnaire effects. For each test, use level of significance $\alpha = .05$ and state the alternatives, decision rule, and conclusion. What is the P-value of each test?

b. Analyze the questionnaire main effects by estimating all pairwise comparisons of treatment means. Use the Tukey procedure and a 90 percent family confidence coefficient. Summarize your findings.

Chapter 9

Exploratory Experiments: Two-Level Factorial and Fractional Factorial Designs

Up to this point, much of our discussion of the design of experiments has focused on the planning of *confirmatory* experiments. Generally, confirmatory experiments employ a relatively small number of explanatory factors. The factors under investigation usually are suggested by existing theory or by previous experimental findings. *Exploratory* experimental studies are typically encountered during the early stages of a new research study, when little is known about the set of important or *active* explanatory factors. At this stage of the investigation, the experimenter often needs to consider a large number of factors in order to identify the factors that are the most important. One means of including a large number of factors in an experiment while keeping the total number of treatment combinations at a manageable level is to study each factor at only two levels. For example, in a four-factor experiment, one replication of a two-level factorial experiment consists of just $2^4 = 16$ treatment trials. In contrast, if each factor were studied at three levels, a single replication would require $3^4 = 81$ treatment trials—over five times that required by the two-level experiment.

Even when only two levels are employed for each factor, the size of the experiment can still become prohibitively large when a large number of factors are to be studied. In such cases, a carefully selected subset, or *fraction,* of the treatments can be used with little or no loss of information about the main effects and key low-order interactions. *Fractional factorial designs* permit the study of a large number of factors with relatively few experimental trials.

Another means of keeping the number of trials small in exploratory experiments is to use a single replication or to employ replications for only one or a few of the treatments.

In this chapter, we first discuss the use of two-level factorial experiments and then consider two-level experiments with only one replication. We then take up fractional factorial designs and their analysis, including designs for screening a large number of factors. In Section 29.5 we discuss briefly the use of blocking in two-level experiments. We conclude the chapter by introducing robust product and process design experiments and illustrate their use with a case study from the automotive industry. Unless explicitly stated otherwise,

we assume throughout the chapter that *all treatment sample sizes are equal and all factor effects are fixed.*

29.1 Two-Level Full Factorial Experiments

Design of Two-Level Studies

Experimental studies involving k factors, each at two levels, are often referred to as 2^k factorial studies. The choice of the two levels for each factor in a two-level factorial experiment at times is automatic. Some factors exist naturally at two levels. For instance, in a marketing research study of the effects of including or excluding special features, such as antilock brakes and automatic headlight dimmers in an automobile, the factors automatically have two levels. At other times, a deliberate choice of the two levels must be made. For instance, in a study of a rubber extrusion process, curing time was one of the factors of interest. Economic and engineering considerations dictated that curing time be at least 30 minutes and not longer than 45 minutes. The two levels selected here were 30 and 45 minutes to provide information at the limits of the range of the factor.

An example of a two-level factorial study involving three factors with three replications is the stress test study in Table 24.4. There, the gender levels were male and female, and subjects were classified as having low or high body fat and being light or heavy smokers.

Since two-level factorial studies are a special case of the factorial studies discussed in earlier chapters, we already know how to analyze such studies. For our purposes here, however, we need to modify our earlier notation because it becomes awkward when there are many factors. Also, we shall see that some simplifications arise in the calculational formulas when all factors have two levels.

Notation

Consider our usual formulation of the regression version of a three-factor ANOVA model for a balanced study where each factor has two levels:

$$\begin{aligned} Y_{ijkm} = \mu_{\cdots} &+ \alpha_1 X_{ijkm1} + \beta_1 X_{ijkm2} + \gamma_1 X_{ijkm3} \\ &+ (\alpha\beta)_{11} X_{ijkm1} X_{ijkm2} + (\alpha\gamma)_{11} X_{ijkm1} X_{ijkm3} \\ &+ (\beta\gamma)_{11} X_{ijkm2} X_{ijkm3} + (\alpha\beta\gamma)_{111} X_{ijkm1} X_{ijkm2} X_{ijkm3} + \varepsilon_{ijkm} \end{aligned} \tag{29.1}$$

where X_1, X_2, X_3 take on the values 1 and -1 for the two factor levels. Even though in a two-level factorial study there is only one main effect term for each factor, one two-factor interaction for each pair of factors, and so on, it is evident that with more factors the notation used in model (29.1) will become very cumbersome.

We therefore will change the notation as follows, using the conventions for polynomial regression in Section 8.1:

1. The main effects will be represented by β_1, β_2, etc. The overall constant will be represented by β_0.
2. The two-factor interaction effects will be represented by β_{12}, β_{13}, etc.
3. Three-factor and higher-order interaction effects will be represented correspondingly; for instance, by β_{123} and β_{1234}.

4. The index i will be used to denote the observation number, running from 1 to n_T.
5. Cross-product terms will be represented by a single X. For instance, X_1X_2 will be represented by X_{12}; $X_1X_2X_3$ will be represented by X_{123}; and so on. The value of X_1X_2 for the ith observation will be represented by X_{i12}.
6. When the factor is quantitative, the low level will be the first level and will be coded -1, and the high level will be the second level and will be coded 1. This coding for quantitative factors is equivalent to standardizing the levels by subtracting the mean and dividing by half of the range. For a qualitative factor, the first level correspondingly will be coded -1 and the second level coded 1. Note that the $-1, 1$ coding here is the opposite of the convention followed earlier.

With these conventions, model (29.1) is now stated as follows, using $X_0 \equiv 1$ as the dummy variable associated with β_0:

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \beta_{12} X_{i12} + \beta_{13} X_{i13} + \beta_{23} X_{i23} + \beta_{123} X_{i123} + \varepsilon_i \tag{29.2}$$

where:

ε_i are independent $N(0, \sigma^2)$

$X_0 \equiv 1$

$$X_1 = \begin{cases} -1 & \text{if case from first level of factor 1} \\ 1 & \text{if case from second level of factor 1} \end{cases}$$

$$X_2 = \begin{cases} -1 & \text{if case from first level of factor 2} \\ 1 & \text{if case from second level of factor 2} \end{cases}$$

$$X_3 = \begin{cases} -1 & \text{if case from first level of factor 3} \\ 1 & \text{if case from second level of factor 3} \end{cases}$$

β_0 in model (29.2) corresponds to $\mu_{\cdot\cdot\cdot}$ in model (29.1). Because the codes -1, 1 are now reversed from our earlier convention, β_1 corresponds to $-\alpha_1 = \alpha_2$. Similarly, β_2 corresponds to $-\beta_1 = \beta_2$, and β_3 corresponds to $-\gamma_1 = \gamma_2$. The parameter β_{12} corresponds to $(\alpha\beta)_{11} = (\alpha\beta)_{22}$ because of two reversals in the signs of the indicator variables.

For k factors, model (29.2) is extended as follows:

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \beta_{12} X_{i12} + \cdots + \beta_{12\cdots k} X_{i12\cdots k} + \varepsilon_i \tag{29.2a}$$

where:

$$X_{ij} = \begin{cases} -1 & \text{if case } i \text{ from first level of factor } j \\ 1 & \text{if case } i \text{ from second level of factor } j \end{cases}$$

and X_0 and ε_i are defined as in (29.2).

It is often helpful to list the treatments in a two-level factorial experiment in a standard order. We shall use as the *standard order* a listing of the treatments such that the level of factor 1, X_1, changes most frequently, the level of factor 2, X_2, changes with second greatest frequency, and so on. In a three-factor study, for instance, the standard order of the

treatments is obtained by listing factor levels in the following sequence:

Treatment	X_1	X_2	X_3
1	−1	−1	−1
2	1	−1	−1
3	−1	1	−1
4	1	1	−1
5	−1	−1	1
6	1	−1	1
7	−1	1	1
8	1	1	1

Note that treatment 1 consists of all three factors at their first levels, treatment 2 consists of factor A at its second level and factors B and C at their first levels, and so on. The matrix consisting of the X_1, X_2, and X_3 columns is called the *design matrix* because it identifies the treatments in the experimental study.

A standard order for treatments is simply a convention for listing treatments in two-level factorial experiments; the actual ordering of the treatment trials in the experiment and the assignment of the treatments to experimental units are determined by randomization.

Estimation of Factor Effects

When a balanced factorial experiment is carried out at two levels for each factor and a −1, 1 coding is employed, the $\mathbf{X}'\mathbf{X}$ matrix is greatly simplified. Consider a two-factor study with $n = 1$ replication. The $\mathbf{X}$ matrix, using the coding in (29.2), is as follows (treatments are in standard order):

$$\mathbf{X} = \begin{array}{c} \begin{array}{cccc} X_0 & X_1 & X_2 & X_{12} \end{array} \\ \begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

The simplifications in the $\mathbf{X}'\mathbf{X}$ matrix arise because:

1. Any two columns of the $\mathbf{X}$ matrix are orthogonal; that is, $\mathbf{X}_q'\mathbf{X}_{q'} = 0$. In our simple example, for instance:

$$\mathbf{X}_1'\mathbf{X}_2 = [-1 \quad 1 \quad -1 \quad 1] \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} = 0$$

2. The sum of squares of the elements in each column, $\mathbf{X}_q'\mathbf{X}_q$, is always n_T. In our simple example, for instance:

$$\mathbf{X}_1'\mathbf{X}_1 = [-1 \quad 1 \quad -1 \quad 1] \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} = 4$$

Consequently, the elements on the main diagonal of the $\mathbf{X}'\mathbf{X}$ matrix are all n_T and the elements off the main diagonal are all zero so that $\mathbf{X}'\mathbf{X}$ is a diagonal matrix:

$$\mathbf{X}'\mathbf{X} = n_T\mathbf{I} \tag{29.3}$$

The inverse of $\mathbf{X}'\mathbf{X}$ therefore is a diagonal matrix with the diagonal elements being the reciprocals of the elements in (29.3):

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{n_T}\mathbf{I} \tag{29.4}$$

The least squares and maximum likelihood estimators in (6.25) therefore become simple in form:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \frac{1}{n_T}\mathbf{X}'\mathbf{Y} \tag{29.5}$$

Letting $\mathbf{X}_q$ denote the column vector containing the qth column of the $\mathbf{X}$ matrix, the estimated regression coefficient b_q therefore is:

$$b_q = \frac{1}{n_T}\mathbf{X}_q'\mathbf{Y} \tag{29.6}$$

Since each column vector $\mathbf{X}_q$ contains only 1s and -1s, the estimated coefficients b_q are very simple linear combinations of the observations.

We illustrate this in Table 29.1, which contains the $\mathbf{Y}$ vector and the $\mathbf{X}$ matrix for the stress test example of Table 24.4, with the observations listed in standard order. The Y observations are shown both in the earlier notation and the current notation to facilitate recognition of the treatments involved. (Note that the coding of the factor levels in the $\mathbf{X}$ matrix is the opposite of that in Table 24.7 and that the ordering of the observations also

TABLE 29.1 **Y and X Data Matrices in Standard Order—Stress Test Example of Table 24.4.**

$$\mathbf{Y} = \begin{bmatrix} Y_{1111} \\ Y_{1112} \\ Y_{1113} \\ Y_{2111} \\ Y_{2112} \\ Y_{2113} \\ Y_{1211} \\ \vdots \\ Y_{2221} \\ Y_{2222} \\ Y_{2223} \end{bmatrix} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ Y_5 \\ Y_6 \\ Y_7 \\ \vdots \\ Y_{22} \\ Y_{23} \\ Y_{24} \end{bmatrix} = \begin{bmatrix} 24.1 \\ 29.2 \\ 24.6 \\ 20.0 \\ 21.9 \\ 17.6 \\ 14.6 \\ \vdots \\ 10.1 \\ 14.4 \\ 6.1 \end{bmatrix} \qquad \mathbf{X} = \begin{array}{c} \begin{array}{cccccccc} X_0 & X_1 & X_2 & X_3 & X_{12} & X_{13} & X_{23} & X_{123} \end{array} \\ \begin{bmatrix} 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ & & & & & & & \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \end{array}$$

differs.) We see that the estimated coefficient b_{12}, for instance, is simply:

$$b_{12} = \frac{1}{n_T}\mathbf{X}'_{12}\mathbf{Y} = \frac{1}{24}[1 \quad 1 \quad 1 \quad -1 \quad \cdots \quad 1]\begin{bmatrix} 24.1 \\ 29.2 \\ 24.6 \\ 20.0 \\ \vdots \\ 6.1 \end{bmatrix} = .754 \tag{29.7}$$

The variance-covariance matrix of **b** in (6.46) is also greatly simplified:

$$\boldsymbol{\sigma}^2\{\mathbf{b}\} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} = \frac{\sigma^2}{n_T}\mathbf{I} \tag{29.8}$$

Note from this matrix that the estimated regression coefficients here are uncorrelated and have constant variance:

$$\sigma^2\{b_q\} = \frac{\sigma^2}{n_T} \tag{29.9}$$

The estimated variance-covariance matrix in (6.48) becomes:

$$\mathbf{s}^2\{\mathbf{b}\} = \frac{MSE}{n_T}\mathbf{I} \tag{29.10}$$

so that the estimated variance of b_q is simply:

$$s^2\{b_q\} = \frac{MSE}{n_T} \tag{29.11}$$

Comments

1. Some texts and software packages define the effect of a factor as an observed difference between responses when that factor changes from its first level to its second level. For example, the estimated main effect of factor 1 (factor A) is defined as:

$$A = \begin{pmatrix}\text{Average response for all} \\ \text{trials in which } X_1 = 1\end{pmatrix} - \begin{pmatrix}\text{Average response for all} \\ \text{trials in which } X_1 = -1\end{pmatrix} \tag{29.12}$$

A is an estimate of $\alpha_2 - \alpha_1 = 2\alpha_2$; recall that $\alpha_1 = -\alpha_2$ when the factors are at two levels. Consequently, the relation between A and our estimate b_1 (which now estimates $\alpha_2 = -\alpha_1$) is:

$$A = 2b_1 \tag{29.13}$$

The relations for the other main effects and interaction effects are similar.

2. The -1, 1 coding used for the predictor variables in (29.2) is sometimes referred to as an *orthogonal coding* because it leads for balanced two-level factorial designs to a diagonal $\mathbf{X}'\mathbf{X}$ matrix. ■

Inferences about Factor Effects

As noted earlier, a main objective in two-level exploratory studies is usually the identification of active effects. An effect is considered active if the corresponding factor effect coefficient is nonzero. Since all estimated factor effects have the same variance for balanced studies, as

noted in (29.9), a normal probability plot can be made of all estimated main and interaction effects to identify those that appear to be active. We shall illustrate this plot shortly.

Formal tests for a regression coefficient, with the alternatives H_0: $\beta_q = 0$, H_a: $\beta_q \neq 0$, are carried out in the usual manner, based on either the t^* statistic in (7.25) or the F^* statistic in (7.24). In many instances, the testing procedure will be used for each of the factor effects. The family level of significance then can be controlled at α by either the Bonferroni inequality (4.4) or the Kimball inequality (19.53).

Example

Figure 29.1 contains the MINITAB FFactorial output for the stress test example of Table 24.4. In this study, the effects of gender of subject (factor A), body fat of subject (factor B), and smoking history of subject (factor C) on exercise tolerance were studied. The MINITAB ANOVA output is based on the coding of the factor levels in Table 29.1. The estimated factor effect coefficients b_q are shown in the column marked "Coef." The column marked "Effect" contains the alternative definition of effects in (29.12). Notice that when each entry in this column is divided by 2, as shown in (29.13), the estimated coefficients b_q are obtained. Also note that the estimated standard deviations in the column labeled "Std Coef" are all the same, as required by (29.11):

$$s\{b_q\} = \left(\frac{MSE}{n_T}\right)^{1/2} = \left(\frac{9.335}{24}\right)^{1/2} = .6237$$

Using a significance level of .015 for each of the seven tests on the estimated factor effect coefficients so as to assure a family level of significance of .10 by the Kimball inequality, we see from the P-values in Figure 29.1 that the set of active factor effects consists of the gender, body fat, and smoking main effects, and the body fat–smoking interaction.

FIGURE 29.1 MINITAB FFactorial Output—Stress Test Example of Table 24.4.

Estimated Effects and Coefficients for TOLERANCE

Term	Effect	Coef	Std Coef	t-value	P
Constant		16.271	0.6237	26.09	0.000
GENDER	−5.425	−2.713	0.6237	−4.35	0.000
BODYFAT	−6.358	−3.179	0.6237	−5.10	0.000
SMOKING	−3.425	−1.713	0.6237	−2.75	0.014
GENDER*BODYFAT	1.508	0.754	0.6237	1.21	0.244
GENDER*SMOKING	−1.358	−0.679	0.6237	−1.09	0.292
BODYFAT*SMOKING	3.475	1.737	0.6237	2.79	0.013
GENDER*BODYFAT*SMOKING	−0.558	−0.279	0.6237	−0.45	0.660

Analysis of Variance for TOLERANCE

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	3	489.538	489.538	163.179	17.48	0.000
2-Way Interactions	3	97.175	97.175	32.392	3.47	0.041
3-Way Interactions	1	1.870	1.870	1.870	0.20	0.660
Residual Error	16	149.367	149.367	9.335		
Pure Error	16	149.367	149.367	9.335		
Total	23	737.950				

29.2 Analysis of Unreplicated Two-Level Studies

In many applications of two-level factorial experiments, particularly when many factors are included, only a single replication can be run because of time, budgetary, or other resource limitations. As discussed in Chapter 20, no degrees of freedom are available for obtaining an estimate of the error variance σ^2 when only one replication is employed. Special procedures instead must be used for statistical analysis.

We shall now describe three approaches for analyzing unreplicated experiments:

1. The pooling of higher-order interactions to obtain an estimate of the variance.
2. The use of graphical procedures for identifying active effects.
3. The use of replications at the center point to obtain a pure error estimate of the error variance σ^2.

First, however, we shall describe an unreplicated 2^4 factorial experiment that will be used as an illustration.

Example

The Pecos Foods Corporation initiated an experimental study to characterize the effects of process temperature (factor 1 or A), an antimicrobial agent or preservative (factor 2 or B), moisture level (factor 3 or C), and acidity (factor 4 or D) on the microbial growth in a fruit bar. Microbial growth is measured by counting microbes in a sample of the product following three months in storage. The four factors were studied at the following low and high levels:

Factor	Low Level	High Level
Process temperature	152	178
Preservative	0.0	.1
Moisture	.65	.85
Acidity	4.8	6.8

One replication of a 2^4 factorial experiment was run. The $\mathbf{X}$ matrix for the standard 2^4 factorial ANOVA model and the response vector are shown in Table 29.2, in standard order. Note that the columns X_1, X_2, X_3, and X_4 constitute the design matrix, identifying each of the treatments. The response, denoted for simplicity by Y, is the natural logarithm of the microbial count. This transformation was chosen partly because the actual counts ranged from 87 to 104,410—i.e., over several orders of magnitude. In addition, the Box-Cox procedure (3.36) supported the use of the logarithmic transformation.

The regression model version of the four-factor ANOVA model was fitted, using the X variables in Table 29.2. The MINITAB regression results for the full ANOVA model ($p = n_T = 16$) are presented in Figure 29.2. Because there are no degrees of freedom available for error, no estimate of the error variance and no t statistics and P-values for the estimated regression coefficients are shown. Note that three estimated factor effect coefficients, $b_2 = -1.25$, $b_3 = 1.40$, and $b_{23} = -1.40$ are substantially larger in absolute value than the next largest coefficient, $b_{134} = -.24$. Consequently, the preservative and moisture factors (2 and 3) may be active. We shall now consider the use of pooling, Pareto

TABLE 29.2 Y Vector and X Matrix—Pecos Foods Corporation Example.

Treatment	Y	X_0	X_1	X_2	X_3	X_4	X_{12}	X_{13}	X_{14}	X_{23}	X_{24}	X_{34}	X_{123}	X_{124}	X_{134}	X_{234}	X_{1234}
1	5.55	1	−1	−1	−1	−1	1	1	1	1	1	1	−1	−1	−1	−1	1
2	4.47	1	1	−1	−1	−1	−1	−1	−1	1	1	1	1	1	1	−1	−1
3	5.19	1	−1	1	−1	−1	−1	1	1	−1	−1	1	1	1	−1	1	−1
4	5.32	1	1	1	−1	−1	1	−1	−1	−1	−1	1	−1	−1	1	1	1
5	10.54	1	−1	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1	1	−1
6	11.56	1	1	−1	1	−1	−1	1	−1	−1	1	−1	−1	1	−1	1	1
7	5.08	1	−1	1	1	−1	−1	−1	1	1	−1	−1	−1	1	1	−1	1
8	5.45	1	1	1	1	−1	1	1	−1	1	−1	−1	1	−1	−1	−1	−1
9	5.12	1	−1	−1	−1	1	1	1	−1	1	−1	−1	−1	1	1	1	−1
10	5.63	1	1	−1	−1	1	−1	−1	1	1	−1	−1	1	−1	−1	1	1
11	6.18	1	−1	1	−1	1	−1	1	−1	−1	1	−1	1	−1	1	−1	1
12	5.24	1	1	1	−1	1	1	−1	1	−1	1	−1	−1	1	−1	−1	−1
13	10.73	1	−1	−1	1	1	1	−1	−1	−1	−1	1	1	1	−1	−1	1
14	10.33	1	1	−1	1	1	−1	1	1	−1	−1	1	−1	−1	1	−1	−1
15	6.53	1	−1	1	1	1	−1	−1	−1	1	1	1	−1	−1	−1	1	−1
16	4.93	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

FIGURE 29.2 MINITAB Regression Results for Full ANOVA Model—Pecos Foods Corporation Example.

The regression equation is

lnmicrob = 6.74 − 0.124 x1 − 1.25 x2 + 1.40 x3 + 0.0956 x4 − 0.131 x12
+ 0.0481 x13 − 0.179 x14 − 1.40 x23 + 0.134 x24 − 0.109 x34
− 0.101 x123 − 0.201 x124 − 0.244 x134 + 0.112 x234 + 0.132 x1234

Predictor	Coef	Stdev	t-ratio	p
Constant	6.74062	0.00000	*	*
x1	−0.124375	0.000000	*	*
x2	−1.25063	0.00000	*	*
x3	1.40313	0.00000	*	*
x4	0.0956249	0.0000000	*	*
x12	−0.130625	0.000000	*	*
x13	0.0481250	0.0000000	*	*
x14	−0.179375	0.000000	*	*
x23	−1.39562	0.00000	*	*
x24	0.134375	0.000000	*	*
x34	−0.109375	0.000000	*	*
x123	−0.100625	0.000000	*	*
x124	−0.200625	0.000000	*	*
x134	−0.244375	0.000000	*	*
x234	0.111875	0.000000	*	*
x1234	0.131875	0.000000	*	*

s = *

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	15	91.62849	6.10857	*	*
Error	0	*	*		
Total	15	91.62849			

plots, dot plots, and normal probability plots in an effort to identify more definitively the set of active effects.

Pooling of Interactions

A common approach to analyzing unreplicated experiments is to assume that some higher-order interactions are inactive. The extra sums of squares corresponding to these interaction terms are then used to form an estimate of the error variance σ^2. For example, in a 2^4 factorial experiment, it may be reasonable to assume that all three-factor and four-factor interactions are small or negligible in relation to main effects and two-factor interactions. This implies that $\beta_{123} = \beta_{124} = \beta_{134} = \beta_{234} = \beta_{1234} = 0$. By dropping the corresponding terms from the model, five degrees of freedom will be available for an estimate of σ^2. For balanced two-level experiments, it can be shown that the extra sum of squares for X_q is:

$$SSR(X_q) = n_T b_q^2 \tag{29.14}$$

Since for balanced two-level factorial studies, the columns of the **X** matrix are orthogonal, any extra sum of squares does not depend on the order of the variables and the extra sums of squares are additive. Hence, the pooled estimator of σ^2 is as follows:

$$MSE = n_T \left(\frac{\sum b_q^2 \text{ for pooled estimated coefficients}}{\text{Number of pooled coefficients}} \right) \tag{29.15}$$

Inferences can then be made in customary fashion.

Example

In the Pecos Foods Corporation example, it was decided that all three-factor and four-factor interactions are unimportant. Using (29.15) and the results in Figure 29.2, an estimate of the error variance based on five degrees of freedom is:

$$MSE = n_T \left(\frac{b_{123}^2 + b_{124}^2 + b_{134}^2 + b_{234}^2 + b_{1234}^2}{5} \right)$$

$$= 16 \left[\frac{(-.101)^2 + (-.201)^2 + (-.244)^2 + (.112)^2 + (.132)^2}{5} \right] = .448$$

MINITAB regression results for the model based on main effects and two-factor interactions are presented in Figure 29.3. Notice that $MSE = .448$, as just calculated. Residual analysis (not shown) did not reveal any violations in assumptions.

The P-values in Figure 29.3 indicate that the main effects for preservative and moisture (factors 2 and 3) and the preservative-moisture interaction effect are statistically significant; each of the associated P-values is .001 or less. The active factors in the Pecos Foods Corporation example are therefore preservative and moisture. Figure 29.4 presents a MINITAB interaction plot of the estimated means $\bar{Y}_{jk\cdot}$ for the two active factors. We see that increasing preservative at high levels of moisture decreases microbial growth. At low moisture levels, however, preservative has little effect. Correspondingly, at low preservative levels, decreasing moisture decreases microbial growth while at high preservative levels, changing the moisture level has little effect on microbial growth.

FIGURE 29.3 MINITAB Regression Results for ANOVA Model without Higher-Order Interactions—Pecos Foods Corporation Example.

The regression equation is
lnmicrob = 6.74 − 0.124 x1 − 1.25 x2 + 1.40 x3 + 0.096 x4 − 0.131 x12
+ 0.048 x13 − 0.179 x14 − 1.40 x23 + 0.134 x24 − 0.109 x34

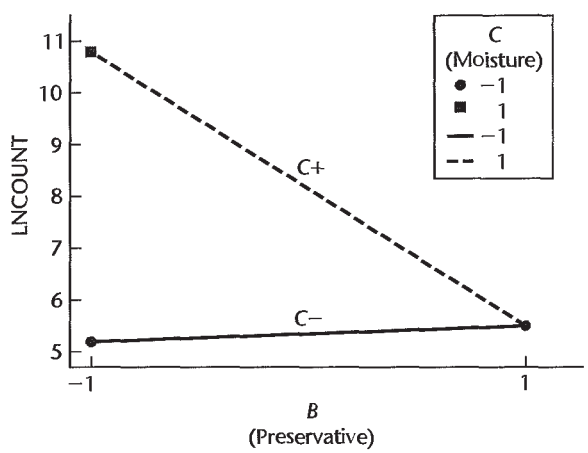
Predictor	Coef	Stdev	t-ratio	p
Constant	6.7406	0.1673	40.28	0.000
x1	−0.1244	0.1673	−0.74	0.491
x2	−1.2506	0.1673	−7.47	0.001
x3	1.4031	0.1673	8.39	0.000
x4	0.0956	0.1673	0.57	0.592
x12	−0.1306	0.1673	−0.78	0.470
x13	0.0481	0.1673	0.29	0.785
x14	−0.1794	0.1673	−1.07	0.333
x23	−1.3956	0.1673	−8.34	0.000
x24	0.1344	0.1673	0.80	0.458
x34	−0.1094	0.1673	−0.65	0.542

s = 0.6693 R-sq = 97.6% R-sq(adj) = 92.7%

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	10	89.3885	8.9388	19.95	0.002
Error	5	2.2400	0.4480		
Total	15	91.6285			

FIGURE 29.4 MINITAB Interaction Plot—Pecos Foods Corporation Example.

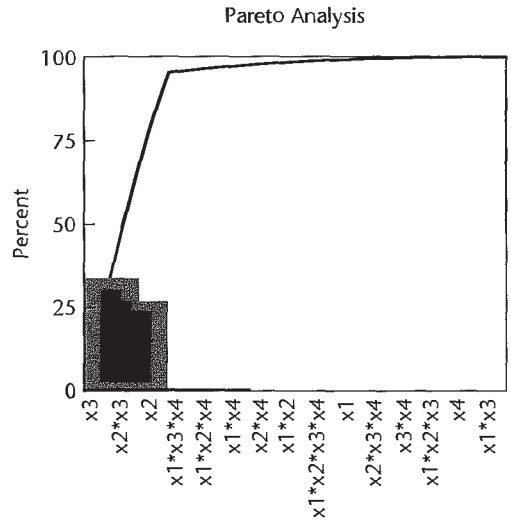


Pareto Plot

The Pareto plot is a qualitative tool for visually identifying important effects in unreplicated two-level studies. It shows the percentage of the total sum of squares *SSTO* that is associated with each estimated effect in the full factorial model. (Remember that for an unreplicated full factorial model, *SSTO* = *SSR*.) From (29.14), this percentage is:

$$\frac{n_T b_q^2}{SSTO}(100) \tag{29.16}$$

FIGURE 29.5 JMP Pareto Plot—Pecos Foods Corporation Example.



Large percentage contributions correspond to large (absolute) estimated coefficients, and therefore to active factor effects. Pareto plots present the percent contributions to *SSTO* in decreasing order, either as a bar plot, a cumulative line plot, or both.

Example

To calculate the percent contribution to the total sum of squares for each factor effect in the Pecos Foods Corporation example, we use (29.16) and the regression results in Figure 29.2 for the full factorial model. For example, the percent contribution associated with X_3 is:

$$\frac{n_T b_3^2}{SSTO}(100) = \frac{16(1.40)^2}{91.63}(100) = 34.2\%$$

A JMP Pareto plot shown in Figure 29.5 contains both a bar plot and a cumulative line plot. Notice that the effects X_2, X_3, and X_2X_3 account for nearly all of the total variation in the data. Thus, the Pareto plot identifies the same factor effects as active as does pooling of higher-order interactions.

Comments

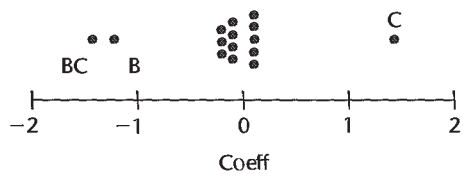
1. Other forms of Pareto plots are also used. For example, some statistics packages provide a Pareto plot of estimated effects. In these plots, the bars correspond to the absolute magnitudes of the estimated effect coefficients. Such plots are sometimes referred to as *scree* plots.

2. While Pareto plots are useful for identifying active effects, they can be misused. For example, a Pareto plot is sometimes used to identify the smallest effects for pooling to estimate σ^2. This approach often will lead to an estimate of the error variance that is too small, making the Type I error rates for tests for active effects larger than desired. ■

Dot Plot

Another graphic plot often used in the analysis of unreplicated factorial studies to help identify active effects is a simple dot plot of estimated factor effect coefficients. This plot will show whether any estimated coefficients are far outlying. We know from (29.9) that the variances for all estimated effect coefficients are the same for unreplicated 2^k factorial studies so that the estimated effect coefficients will follow the same normal distribution if no

FIGURE 29.6 Dot Plot of Estimated Factor Effect Coefficients—Pecos Foods Corporation Example.



effects are present. Inactive factors will tend to be clustered in the middle of the distribution. A large departure from the middle of the distribution suggests that the factor may be active.

Example

A dot plot of the estimated factor effect coefficients for the Pecos Foods Corporation example is presented in Figure 29.6. Note that most factor effects fall near zero; these presumably are the inactive factor effects. The three outlying coefficients, for factors B and C, correspond to the three factor effects identified already by the other techniques as the active effects.

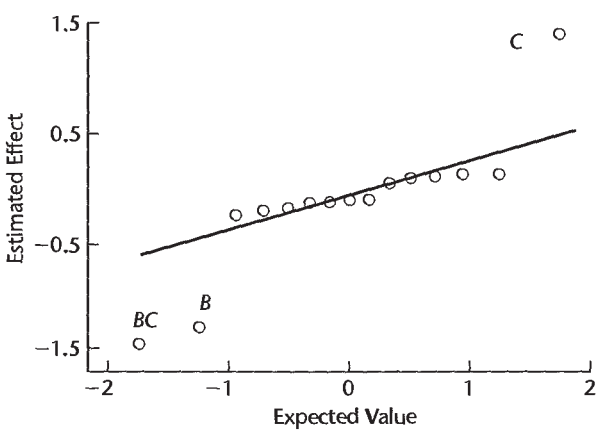
Normal Probability Plot

A normal probability plot of the estimated factor effect coefficients in an unreplicated 2^k factorial study can be constructed in the same fashion as a normal probability plot of residuals, as described on page 110. This is possible because the estimated factor effect coefficients are independent with constant variance σ^2/n_T. Since no estimate of σ^2 is available, we set $MSE = 1$ in (3.6). If no effects are present, all estimated coefficients follow the same normal distribution $N(0, \sigma^2/n_T)$ and should fall along a straight line in the plot. Strong deviations from a straight line are indicative of active effects, in which case all estimated coefficients do not come from the same normal distribution. Typically, the middle points represent inactive effects and fall along a straight line. If they do not, it may be an indication that the error terms are not normally distributed.

Example

Figure 29.7 shows a normal probability plot of the estimated effect coefficients for the Pecos Foods Corporation example. A line has been fitted judgmentally to the center points that appear to represent inactive effects. Notice that the estimated effect coefficients for the

FIGURE 29.7 Normal Probability Plot of Estimated Effect Coefficients—Pecos Foods Corporation Example.



factor B and factor C main effects and for the BC interaction effect fall away from the line fitted to the inactive effects.

Comments

1. When many factor effects are active and only a few are inactive, it may be difficult to fit a line to the few inactive effects at the center. Consequently, a normal probability plot with many active factor effects is often difficult to interpret.

2. Half-normal probability plots, as described in Section 14.8, are often used in place of (full) normal probability plots discussed here. One advantage of half-normal probability plots is that identification of active effects is sometimes facilitated. This is because the active effects in a half-normal plot all fall at the right upper end of the plot whereas in a (full) normal plot active effects may be at both ends of the plot.

3. A normal probability plot containing all factor effects is also appropriate for 2^k factorial experiments with replications provided that there are equal numbers of replications for each treatment. ■

Center Point Replications

When all factors are quantitative, two-level experiments can be augmented by replications at the *center point*. A center point is a new treatment in which each of the factors is set at the midpoint of its range. For example, in the Pecos Foods Corporation example, the center point treatment levels are:

$$\text{Temperature} = \frac{152 + 178}{2} = 165$$

$$\text{Preservative} = \frac{0 + .1}{2} = .05$$

$$\text{Moisture} = \frac{.65 + .85}{2} = .75$$

$$\text{Acidity} = \frac{4.8 + 6.8}{2} = 5.8$$

We shall use n_0 to denote the number of center point replicates. Two important advantages stem from the inclusion of two or more such replicates:

1. A pure error estimate of σ^2 based on $n_0 - 1$ degrees of freedom can be obtained, avoiding any bias that otherwise might be associated with inferential procedures based on the pooling of what appear to be small higher-order effects.
2. With replications at the center point, it is possible to test whether or not the model is a good fit.

Pure Error Estimate of σ^2. Let Y_{0i} denote the response associated with the ith replicate at the center point, and let $\bar{Y}_{0\cdot}$ denote the mean of the n_0 responses at the center point. A pure error estimate of σ^2 is given by the sample variance of the center point replicates:

$$MSPE = \frac{\sum(Y_{0i} - \bar{Y}_{0\cdot})^2}{n_0 - 1} \tag{29.17}$$

Test for Lack of Fit. Once a pure error mean square has been obtained, the test for lack of fit in (6.68) proceeds as usual. For a two-level factorial study with no replications that is

augmented by n_0 observations at the center point, one degree of freedom is associated with *SSLF* and $n_0 - 1$ degrees of freedom with *SSPE*.

A conclusion of lack of fit indicates that curvature is present in one or more of the factor effects, but it is not possible to attribute the curvature effect to a specific factor without further experimentation. Methods for augmenting two-level factorial experiments for assessment of curvature effects are discussed in Chapter 30.

Example

Suppose that four center point replicates had been included in the Pecos Foods Corporation study and that these responses are:

$$Y_{01} = 7.23 \qquad Y_{02} = 7.89 \qquad Y_{03} = 7.80 \qquad Y_{04} = 7.39$$

We then find $\bar{Y}_{0\cdot} = 7.578$, $SSPE = .303$, and $MSPE = .101$. From the regression analysis of the augmented data set (output not shown), we find that $SSE = 2.544$. Hence, using (3.24), we obtain:

$$SSLF = SSE - SSPE = 2.544 - .303 = 2.241$$

Hence, test statistic (6.68b) here is:

$$F^* = \frac{2.241}{1} \div \frac{.303}{3} = 22.2$$

For $\alpha = .05$, we require $F(.95; 1, 3) = 10.1$. Since $F^* = 22.2 > 10.1$ we conclude H_a, that curvature is present. The P-value of the test is .018.

We can obtain some information about the nature of the curvature by comparing the average of the responses at the center point, $\bar{Y}_{0\cdot} = 7.578$, with the average of the responses at the corner points, which is 6.74. Since the mean response is higher at the center point than would be expected from a linear interpolation of the corner points, a mound-shaped surface may be required to model the response adequately in the interior of the experimental region.

Comment

When a lack of fit test is conducted after the ANOVA model has been revised by dropping effects that appear to be unimportant, a conclusion of lack of fit does not necessarily imply the presence of curvature effects. Lack of fit could then also be due, for instance, to the absence of important interaction effects. ■

29.3 Two-Level Fractional Factorial Designs

Even when each factor is studied at only two levels, the number of treatments grows rapidly with the number of factors, as the following table demonstrates:

Number of Factors	Number of Treatments
2	4
4	16
6	64
8	256
10	1,024

The use of 1,024 experimental trials for just one replication to study 10 factors will be prohibitive in most instances. In this situation, a subset of all factorial treatments can often be used with little loss of information. The use of fractional factorial designs is the subject of this section.

A basic notion that underlies the use of fractional factorial designs is the *sparsity of effects* principle. This principle states that in most systems, responses are driven largely by a limited number of main effects and lower-order interactions, and that higher-order interactions usually are relatively unimportant. For example, information concerning three-factor and higher-order interactions is often not important compared to main effects and two-factor interactions. Under these conditions, a full factorial design can be very wasteful when many factors are of interest. For instance, in the analysis of a full six-factor, two-level factorial experiment, the degrees of freedom associated with the various factor effects are as follows:

Model Terms	Degrees of Freedom
Intercept term	1
Main effect coefficients	6
Two-factor interaction coefficients	15
Three-factor interaction coefficients	20
Four-factor interaction coefficients	15
Five-factor interaction coefficients	6
Six-factor interaction coefficients	1

Note that 42 degrees of freedom will be devoted to the study of three-factor and higher-order interactions. Thus, about 2/3 (42/64) of the degrees of freedom for the study of factor effects in this experiment will be used to estimate factor effects that are ordinarily of little interest. In contrast, in a fractional factorial design, a subset of the treatments is selected in such a way that most of the degrees of freedom for the study of factor effects are devoted to main effects and low-order interactions, with only some loss of information about higher-order interactions.

Confounding

A fractional factorial design achieves the efficiency of providing full information about main effects and low-order interactions with fewer experimental trials by confounding these effects with unimportant higher-order interactions. To understand the concept of confounding, consider again the **X** matrix of the Pecos Foods Corporation example in Table 29.2. A single replication of a 2^4 full factorial design was employed here, requiring 16 experimental trials. Suppose that in advance of the experiment, it had been determined that only half of the 16 treatments could be used due to budgetary constraints. Which eight of the 16 treatments should be eliminated? Suppose that the experimenter considered dropping treatments 2, 3, 6, 7, 10, 11, 14, 15. The **X** matrix for the remaining eight treatments is given in Table 29.3a.

This choice of treatments to be dropped involves a number of potentially serious problems. Notice first that column vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ are identical in the eight-run design of

TABLE 29.3 **X Matrices for Two Half-Fraction Designs of the 2^4 Full Factorial Design in Table 29.2—Pecos Foods Corporation Example.**

(a) Treatments 2, 3, 6, 7, 10, 11, 14, 15 deleted

Treatment	X_0	X_1	X_2	X_3	X_4	X_{12}	X_{13}	X_{14}	X_{23}	X_{24}	X_{34}	X_{123}	X_{124}	X_{134}	X_{234}	X_{1234}
1	1	−1	−1	−1	−1	1	1	1	1	1	1	−1	−1	−1	−1	1
4	1	1	1	−1	−1	1	−1	−1	−1	−1	1	−1	−1	1	1	1
5	1	−1	−1	1	−1	1	−1	1	−1	1	−1	1	−1	1	1	−1
8	1	1	1	1	−1	1	1	−1	1	−1	−1	1	−1	−1	−1	−1
9	1	−1	−1	−1	1	1	1	−1	1	−1	−1	−1	1	1	1	−1
12	1	1	1	−1	1	1	−1	1	−1	1	−1	−1	1	−1	−1	−1
13	1	−1	−1	1	1	1	−1	−1	−1	−1	1	1	1	−1	−1	1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

(b) Treatments 2, 3, 5, 8, 9, 12, 14, 15 deleted

Treatment	X_0	X_1	X_2	X_3	X_4	X_{12}	X_{13}	X_{14}	X_{23}	X_{24}	X_{34}	X_{123}	X_{124}	X_{134}	X_{234}	X_{1234}
1	1	−1	−1	−1	−1	1	1	1	1	1	1	−1	−1	−1	−1	1
4	1	1	1	−1	−1	1	−1	−1	−1	−1	1	−1	−1	1	1	1
6	1	1	−1	1	−1	−1	1	−1	−1	1	−1	−1	1	−1	1	1
7	1	−1	1	1	−1	−1	−1	1	1	−1	−1	−1	1	1	−1	1
10	1	1	−1	−1	1	−1	−1	1	1	−1	−1	1	−1	−1	1	1
11	1	−1	1	−1	1	−1	1	−1	−1	1	−1	1	−1	1	−1	1
13	1	−1	−1	1	1	1	−1	−1	−1	−1	1	1	1	−1	−1	1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 29.3a; i.e., $\mathbf{X}_1 = \mathbf{X}_2$. Because the columns of this $\mathbf{X}$ matrix are linearly dependent, the matrix $\mathbf{X}'\mathbf{X}$ is singular and does not have an inverse. To be able to obtain least squares and maximum likelihood estimates, we must remove the redundancy resulting from the equality of the $\mathbf{X}_1$ and $\mathbf{X}_2$ column vectors. We do this by retaining only one of the two column vectors. Suppose that we drop the $\mathbf{X}_2$ column vector. In our original model, the main effects for factors 1 and 2 were represented by:

$$\beta_1 X_1 + \beta_2 X_2 \tag{29.18}$$

When $X_1 = X_2$, the model terms become:

$$\beta_1 X_1 + \beta_2 X_1 = (\beta_1 + \beta_2) X_1 \qquad \text{when} \qquad X_1 = X_2 \tag{29.18a}$$

Thus, with the experimental design in Table 29.3a, we will not be able to estimate the factor 1 and factor 2 main effects separately but only their combined main effects. If the experimental results indicate that the effect associated with X_1 is active, we will not know whether the result is due to the effect of factor 1, the effect of factor 2, or to a combination of the effects of these two factors. Factors 1 and 2 are said to be *confounded* or *aliased* in this experiment.

Upon further inspection of Table 29.3a, we find seven more pairs of identical columns, resulting in the following correspondences among the columns of **X**:

$$\begin{array}{llll} \mathbf{X}_1 = \mathbf{X}_2 & \mathbf{X}_3 = \mathbf{X}_{123} & \mathbf{X}_4 = \mathbf{X}_{124} & \mathbf{X}_{13} = \mathbf{X}_{23} \\ \mathbf{X}_{14} = \mathbf{X}_{24} & \mathbf{X}_{234} = \mathbf{X}_{134} & \mathbf{X}_{1234} = \mathbf{X}_{34} & \mathbf{X}_{12} = \mathbf{X}_0 \end{array} \tag{29.19}$$

Consequently, the two effects in each of the following pairs will be confounded with each other:

$$\begin{array}{llll} \beta_1 + \beta_2 & \beta_3 + \beta_{123} & \beta_4 + \beta_{124} & \beta_{13} + \beta_{23} \\ \beta_{14} + \beta_{24} & \beta_{234} + \beta_{134} & \beta_{1234} + \beta_{34} & \beta_{12} + \beta_0 \end{array} \tag{29.20}$$

Since β_{12} is confounded with β_0, the overall mean, β_{12} is sometimes said to be *unmeasurable*.

The relations in either (29.19) or (29.20) define the complete *confounding scheme* for this fractional factorial design. We shall generally describe a confounding scheme in the form of (29.19) and, for simplicity, shall show the column correspondences by means of the subscripts of the column vectors. For our example in Table 29.3a, the confounding scheme is represented in this fashion as follows:

$$\begin{array}{llll} 1 = 2 & 3 = 123 & 4 = 124 & 13 = 23 \\ 14 = 24 & 234 = 134 & 1234 = 34 & 12 = 0 \end{array}$$

The subscript numbers are now shown in italics as a reminder that the equality sign applies not to the numbers shown but to the column vectors for which the numbers are the subscripts.

The proposed eight-treatment design in Table 29.3a is clearly undesirable since main effects are confounded with each other. Suppose instead that the investigator chose to eliminate treatments 2, 3, 5, 8, 9, 12, 14, 15. The resulting **X** matrix is given in Table 29.3b. Notice that the correspondences among the columns of the **X** matrix now are:

$$\begin{array}{llll} 1 = 234 & 2 = 134 & 3 = 124 & 4 = 123 \\ 12 = 34 & 13 = 24 & 14 = 23 & 0 = 1234 \end{array} \tag{29.21}$$

We see that main effects are now confounded only with three-factor interactions and that two-factor interactions are confounded with other two-factor interactions, while the four-factor interaction is confounded with the overall mean. If three-factor and four-factor interactions are negligible, this design could be quite useful. In that case, if $\beta_1 + \beta_{234}$ were found to be statistically significant, we could safely conclude that the observed effect is due to factor 1 and not to the three-factor interaction among factors 2, 3, and 4.

A potential drawback of the design in Table 29.3b is that the two-factor interactions are confounded with other two-factor interactions. If any effects associated with two-factor interactions turn out to be active, additional experimental trials will be required to separate these effects.

An abbreviated ANOVA table for the fractional factorial design in Table 29.3b showing only source of variation and degrees of freedom is given in Table 29.4. Notice that only eight factor effect coefficients can be estimated, corresponding to the eight confounded pairs of effects. Since no degrees of freedom are available for estimation of σ^2, the tools described in Section 29.2 for the analysis of unreplicated factorial studies need to be employed for the analysis of factor effects.

TABLE 29.4 Abbreviated ANOVA Table for Fractional Factorial Design in Table 29.3b.

Source of Variation	*df*
$X_0 = X_{1234}$	1
$X_1 = X_{234}$	1
$X_2 = X_{134}$	1
$X_3 = X_{124}$	1
$X_4 = X_{123}$	1
$X_{12} = X_{34}$	1
$X_{13} = X_{24}$	1
$X_{14} = X_{23}$	1
Error	0
Total	8

Defining Relation

In our explanation of confounding, we began with a full factorial design, arbitrarily dropped some treatments from the experiment, and then examined whether the choice of the dropped treatments was a good one by considering the confounding scheme of the resulting fractional factorial design. Finding an appropriate fractional factorial design is actually done in reverse order by first specifying an acceptable confounding scheme. In order to proceed from this specification to find the corresponding fractional factorial design, we need to utilize the defining relation of the confounding scheme.

Consider again the fractional factorial design in Table 29.3b. The defining relation for this design is the correspondence in (29.21) involving the X_0 column:

$$0 = 1234 \tag{29.22}$$

Recall that (29.22) is a shorthand stating that the X_0 column equals the X_{1234} column. Hence, $X_{i0} = X_{i1234}$ for all column entries. The confounding scheme for the design can be determined from this defining relation by multiplying the column on each side of the defining relation by successive columns of the **X** matrix, the multiplication being carried out term by term.

Since all column entries for a two-level factorial design are either 1 or -1, some general column multiplication results are useful.

1. When multiplying the X_0 column by the X_0 column (the resulting column entries being $X_{i0}X_{i0}$), all entries remain 1 since $X_{i0} \equiv 1$ and $(1)^2 = 1$. We state this in the following fashion:

$$0 \times 0 = 0^2 = 0 \tag{29.23}$$

2. Multiplying any column X_q by X_0 (the resulting column entries being $X_{i0}X_{iq}$) leaves the column entries unchanged because $X_{i0} \equiv 1$:

$$0 \times q = q \tag{29.24}$$

3. Multiplying any column by itself (the resulting column entries being $X_{iq}X_{iq}$) yields the X_0 column since $(1)^2 = (-1)^2 = 1$:

$$q \times q = q^2 = 0 \tag{29.25}$$

Returning now to the defining relation in (29.22), let us multiply the columns on both sides of the defining relation by the X_1 column. On the left side we obtain by (29.24):

$$1 \times 0 = 1 \tag{29.26}$$

and on the right side we find:

$$1 \times 1234 = 1^2 234 = 0234 = 234 \tag{29.27}$$

The result in (29.27) follows because we obtain for each column entry:

$$X_{i1} X_{i1234} = X_{i1}(X_{i1} X_{i2} X_{i3} X_{i4}) = X_{i1}^2 X_{i2} X_{i3} X_{i4} = X_{i2} X_{i3} X_{i4}$$

Combining the results in (29.26) and (29.27), we have found:

$$1 \times 0 = 1 = 1 \times 1234 = 234 \qquad (1 = 234) \tag{29.28a}$$

Continuing the process of multiplying both sides of (29.22) by successive columns of the **X** matrix we find:

$$\begin{aligned}
2 \times 0 &= 2 \times 1234 = 12^2 34 = 134 && (2 = 134)\\
3 \times 0 &= 3 \times 1234 = 123^2 4 = 124 && (3 = 124)\\
4 \times 0 &= 4 \times 1234 = 1234^2 = 123 && (4 = 123)\\
12 \times 0 &= 12 \times 1234 = 1^2 2^2 34 = 34 && (12 = 34)\\
13 \times 0 &= 13 \times 1234 = 1^2 23^2 4 = 24 && (13 = 24)\\
14 \times 0 &= 14 \times 1234 = 1^2 234^2 = 23 && (14 = 23)
\end{aligned} \tag{29.28b}$$

We stop at this point because multiplication by succeeding columns will yield no new confounding relations.

Notice that the operations in (29.28a) and (29.28b) have reproduced the complete confounding scheme in (29.21). The relation on which these operations were based, $0 = 1234$, is called the *defining relation*. The defining relation is always the one that shows the equality with the X_0 column.

Half-Fraction Designs

Once the desired defining relation (and hence, the confounding scheme) is specified, the fractional factorial design corresponding to the desired confounding scheme can be constructed in the following manner:

Step 1. Construct the **X** matrix for the full factorial design.

Step 2. Choose those rows (treatments) for which the defining relation holds.

To illustrate the use of this procedure, consider again the Pecos Foods Corporation example in Table 29.2. The desired defining relation is that in (29.22), namely, $0 = 1234$. Hence, we need to select those treatments for which $X_{i1234} = X_{i0}$. We see from Table 29.2 that $X_{i1234} = 1$ for treatments 1, 4, 6, 7, 10, 11, 13, 16. This is the design in Table 29.3b. It is called a 2^{4-1} fractional factorial design. As noted before, the 4 in the exponent refers to the number of factors. The 1 indicates the level of fractionation; here the full factorial design was fractionated one time. In general, we shall refer to a 2^{k-f} fractional factorial design, where k denotes the number of factors and f the fraction.

An equally useful half-fraction design can be constructed from the eight treatments that were omitted (2, 3, 5, 8, 9, 12, 14, 15). Notice from Table 29.2 that $X_{i0} = -X_{i1234}$ for these treatments. The defining relation for this alternate half-fraction design is therefore:

$$0 = -1234 \tag{29.29}$$

It is easy to verify that the complete confounding scheme for this design is:

$$\begin{array}{llll} 1 = -234 & 2 = -134 & 3 = -124 & 4 = -123 \\ 12 = -34 & 13 = -24 & 14 = -23 & 0 = -1234 \end{array} \tag{29.30}$$

We see that confounding scheme (29.30) for the omitted treatments is the same as that of the retained treatments in (29.28) except that the sign of the second term has changed. Statistically, both of these half-fractions provide similar information, and either one can be used. The choice is sometimes based on the investigator's desire to include one or more specific treatments in the experiment. For example, when the treatment consisting of all runs at the first level (−1) is the control treatment, the investigator who wishes to include this treatment would select the half-fraction corresponding to the defining relation $0 = 1234$.

Comment

The identification of the treatments to be included in a 2^{k-f} fractional factorial design can be carried out without first constructing the **X** matrix for the full 2^k factorial study. The use of *design generators* permits the construction of a 2^{k-f} fractional factorial design by constructing the **X** matrix for a full factorial study in only $k - f$ factors and then augmenting this matrix. Details are provided in Reference 29.1. ■

Quarter-Fraction and Smaller-Fraction Designs

When the number of factors is large, the number of treatments in even a one-half fraction design may still be prohibitively large. In such cases, smaller fractions may be obtained by continuing the process of fractionation. For example, in the Pecos Foods Corporation example, a single replication of a full factorial study involves $2^4 = 16$ experimental trials and a half-fraction design involves $2^{4-1} = 8$ trials. A single replication of a quarter-fraction design will involve only $2^{4-2} = 4$ trials and an eighth-fraction design will consist of $2^{4-3} = 2$ trials. We shall now describe the construction and analysis of 2^{k-f} fractional factorial designs. The number of treatments in such a design is 2^{k-f}.

We shall illustrate how to obtain the confounding scheme for a quarter-fraction design by returning to the Pecos Foods Corporation example in Table 29.3b, where the half-fraction design is based on the defining relation $0 = 1234$. Let us fractionate this design in half by using the defining relation $0 = 12$. From Table 29.3b, we see that $X_{i12} = X_{i0} = 1$ for treatments 1, 4, 13, 16. The **X** matrix for this new quarter-fraction design is given in Table 29.5. Notice that the confounding of effects has become quite severe. From an inspection of the columns of the **X** matrix, we find that the complete confounding scheme is:

$$\begin{aligned} 0 &= 1234 = 12 = 34 \qquad \text{(defining relation)} \\ 1 &= 234 = 2 = 134 \\ 3 &= 124 = 123 = 4 \\ 13 &= 24 = 23 = 14 \end{aligned}$$

TABLE 29.5 Quarter-Fraction Design of 2^4 Full Factorial Design in Table 29.2, Based on Defining Relation: 0 = 1234 = 12 = 34—Pecos Foods Corporation Example.

Treatment	X_0	X_1	X_2	X_3	X_4	X_{12}	X_{13}	X_{14}	X_{23}	X_{24}	X_{34}	X_{123}	X_{124}	X_{134}	X_{234}	X_{1234}
1	1	−1	−1	−1	−1	1	1	1	1	1	1	−1	−1	−1	−1	1
4	1	1	1	−1	−1	1	−1	−1	−1	−1	1	−1	−1	1	1	1
13	1	−1	−1	1	1	1	−1	−1	−1	−1	1	1	1	−1	−1	1
16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Since main effects are confounded with each other (1 with 2, and 3 with 4) this design is clearly undesirable.

As in the case of a half-fraction design, the confounding scheme for a quarter-fraction design can be determined directly without constructing the **X** matrix. We begin with the half-fraction defining relation:

$$0 = 1234 \tag{29.31a}$$

We then augment this with the defining relation for the second fractionation:

$$0 = 1234 = 12 \tag{29.31b}$$

Finally, we need to add a term to recognize that the X_{34} column is also equal to the X_{1234} and X_{12} columns:

$$0 = 1234 = 12 = 34 \tag{29.31c}$$

34 is called the *generalized interaction*. It can be automatically identified by multiplying the two interaction columns X_{1234} and X_{12} in the augmented defining relation in (29.31b):

$$1234 \times 12 = 1^2 2^2 34 = 34$$

In general, for a 2^{k-f} fractional factorial design, there are 2^f terms in the defining relation. These consist of:

1. The constant term, *0*.
2. The f interaction terms used to define the f successive fractionations.
3. The $2^f - f - 1$ generalized interactions, constructed from the cross products involving pairs, triples, and so on, of the f interaction terms used to define the f successive fractionations. Since there are 2^f terms in the defining relation for a 2^{k-f} fractional factorial design, we see that each factor effect is confounded with $2^f - 1$ other factor effects.

Once the defining relation has been obtained for a 2^{k-f} design, the complete confounding scheme can be found by multiplying all terms in the defining relation successively by the main effect and interaction columns in the **X** matrix.

Example

A two-level, five-factor experiment is to be fractionated, first on the basis of the relation:

$$0 = 124 \tag{29.32a}$$

and a second time using:

$$0 = -135 \tag{29.32b}$$

We shall now determine the complete confounding scheme for the experiment. Combining (29.32a) and (29.32b), we obtain:

$$0 = 124 = -135$$

The generalized interaction is therefore:

$$124 \times -135 = -1^2 2345 = -2345$$

The defining relation consequently is:

$$0 = 124 = -135 = -2345 \tag{29.33}$$

The complete confounding scheme is determined by multiplying the terms in (29.33) successively by each of the $2^5 - 1 = 31$ main effect and interaction columns. For example:

$$1 \times 0 = 1 \times 124 = 1 \times -135 = 1 \times -2345 \quad \text{or} \quad 1 = 24 = -35 = -12345$$

In summary we find (omitting any redundant entries):

$$\begin{aligned}
0 &= 124 = -135 = -2345 \\
1 &= 24 = -35 = -12345 \\
2 &= 14 = -1235 = -345 \\
3 &= 1234 = -15 = -245 \\
4 &= 12 = -1345 = -235 \\
5 &= 1245 = -13 = -234 \\
23 &= 134 = -125 = -45 \\
34 &= 123 = -145 = -25
\end{aligned}$$

The eight treatments to be included in this fractional factorial design are those for which $X_{i124} = 1$, $X_{i135} = -1$, and $X_{i2345} = -1$ simultaneously.

Resolution

The resolution of a two-level fractional factorial design, denoted by R, is the number of factors involved in the lowest-order effect in the defining relation, excluding the constant term (*0*). This is a critical characteristic of a design because it indicates the severity of the confounding scheme. For instance, recall that the defining relation of the 2^{4-1} fractional factorial design of Table 29.3b is:

$$0 = 1234$$

The resolution of this half-fraction design is $R = 4$ because there are four factors involved in the term *1234*. The resolution $R = 4$ tells us that the most severe cases of confounding will involve:

1. A main effect and a three-factor interaction (e.g., $1 = 234$)
2. A two-factor interaction and another two-factor interaction (e.g., $12 = 34$)

Roman numerals are commonly used to denote the resolution to avoid confusion with the number of factors. We characterize the design in Table 29.3b as a 2_{IV}^{4-1} fractional factorial design to indicate that it has resolution $R = IV$.

In general, the higher is the resolution of a design, the less severe is the degree of confounding. The resolution should never be less than III. In a resolution II design, at least one pair of main effects will be confounded together. For example, consider the 2_{II}^{5-2} quarter-fraction design with defining relation:

$$0 = 123 = 45 = 12345 \tag{29.34}$$

Since the lowest-order effect in this defining relation is *45*, the design has resolution II. Here the factor 4 main effect is confounded with the factor 5 main effect ($4 = 5$), which is clearly most undesirable. Fractional factorial designs of resolution III, IV, and V are most commonly used. The relationship between resolution and degree of confounding for these three classes of designs can be summarized as follows:

Design Resolution	Worst-Case Degree of Confounding
III	Some main effects are confounded with two-factor interactions.
IV	Some main effects are confounded with three-factor interactions. Some two-factor interactions are confounded with other two-factor interactions.
V	Some main effects are confounded with four-factor interactions. Some two-factor interactions are confounded with three-factor interactions.

Projection Property. A useful property of fractional factorial designs is that any design of resolution R contains complete factorial designs in any subset of $R - 1$ factors. For example, consider the resolution IV half-fraction design in Table 29.3b. Note that if we were to drop the fourth factor, for instance, a full factorial eight-run design would result for the first three factors. This has important design implications. Suppose that an experimenter expects that at most three of the five factors in a study will turn out to be active. By choosing a fractional design with resolution IV, the experimenter will be assured that once the inactive factors are identified and dropped from the analysis, the experimental design for the remaining active factors will be a full factorial design with no confounding.

Selecting a Fraction of Highest Resolution

Clearly, it is desirable that a defining relation be chosen so that the resolution of the design is as large as possible. For half-fraction designs, this is easy: equate the highest-order interaction column with the X_0 column. For example, to provide the maximum resolution (V) in a five-factor study, set the defining relation as follows:

$$0 = 12345$$

In general, the resulting resolution is equal to the number of factors in the study.

For quarter replicates, eighth replicates, and so on, identifying the defining relation that yields the maximum resolution is not so simple. For example, consider the choice of a defining relation for a 2^{6-2} design. If we fractionate first on the basis of:

$$0 = 123456$$

the highest resolution possible will be III:

$$0 = 123456 = 123 = 456$$

However, an alternative defining relation leads to a resolution IV design:

$$0 = 1235 = 2346 = 1456$$

This is, in fact, the highest possible resolution in a 2^{6-2} fractional factorial design.

The 2^{k-f} fractional factorial designs that have highest resolution have been identified and catalogued for choices of k and f that are of general interest (Ref. 29.1). Table 29.6 lists the defining relations for these designs for $3 \leq k \leq 9$; the generalized interactions have been omitted in this listing for the sake of brevity. A number of software packages also will construct fractional factorial designs with highest possible resolution for specified numbers of factors and experimental trials. Most of these packages construct fractional factorial designs employing the defining relations in Table 29.6.

Example

The Iowa Aluminum Corporation manufactures sheet aluminum from recycled aluminum beverage containers. The manufacturing process first casts molten aluminum onto a conveyor belt in a continuous strip. The strip is then sprayed with a coolant comprised of a mixture of water and oil as it enters each of three mills. After the processing in the third mill, the strip is automatically coiled and packaged for shipping. The surface of the aluminum sheets must be sufficiently clean and free of defects or the product will not be shipped. Historically, the rejection rate was about 25 percent.

In an effort to reduce the percentage of rejected coils, an experimental study was undertaken. Management committed two days of production to the experiment, which permitted about 20 experimental trials. Six factors that might affect the quality of the aluminum were identified: (1) temperature of the coolant; (2) percentage of oil in coolant; (3, 4, 5) volume of coolant applied to the strip at each of the three mills (as a percentage of full volume); and (6) strip speed. Low and high limits for each of the factors were identified for the two-level six-factor experiment. Since a 2^6 experiment involves 64 factor level combinations or treatments and since only about 20 experimental trials were feasible, a one-quarter fractional factorial design was needed.

Figure 29.8 contains a summary of the quarter-fraction design for the two-level six-factor experiment provided by the MINITAB Fractional Factorial procedure. We see that a resolution IV design is the highest-resolution design that can be attained in a 16-run, six-factor fractional factorial study. For this resolution, we know that all main effects are clear of other main effects and two-factor interactions and that some main effects will be confounded with three-factor interactions. Also, some two-factor interactions will be confounded with other two-factor interactions. The complete confounding scheme is shown in Figure 29.8, where the factors are denoted A through F (instead of *1* through *6*) and the symbol I is used (instead of *0*) to denote the constant term. Also, MINITAB uses the format in

TABLE 29.6 Two-Level Fractional Factorial Designs with Maximum Resolution for Three to Nine Factors.

Number of Factors	Fraction	Number of Runs	Defining Relation (omitting generalized interactions)
3	2_{III}^{3-1}	4	$0 = 123$
4	2_{IV}^{4-1}	8	$0 = 1234$
5	2_{V}^{5-1}	16	$0 = 12345$
	2_{III}^{5-2}	8	$0 = 124 = 135$
6	2_{VI}^{6-1}	32	$0 = 123456$
	2_{IV}^{6-2}	16	$0 = 1235 = 2346$
	2_{III}^{6-3}	8	$0 = 124 = 135 = 236$
7	2_{VII}^{7-1}	64	$0 = 1234567$
	2_{IV}^{7-2}	32	$0 = 12346 = 12457$
	2_{IV}^{7-3}	16	$0 = 1235 = 2346 = 1347$
	2_{III}^{7-4}	8	$0 = 124 = 135 = 236 = 1237$
8	2_{V}^{8-2}	64	$0 = 12347 = 12568$
	2_{IV}^{8-3}	32	$0 = 1236 = 1247 = 23458$
	2_{IV}^{8-4}	16	$0 = 2345 = 1346 = 1237 = 1248$
9	2_{VI}^{9-2}	128	$0 = 134678 = 235679$
	2_{IV}^{9-3}	64	$0 = 12347 = 13568 = 34569$
	2_{IV}^{9-4}	32	$0 = 23456 = 13457 = 12458 = 12359$
	2_{III}^{9-5}	16	$0 = 1235 = 2346 = 1347 = 1248 = 12349$

(29.20) to represent the confounding scheme. For example, the defining relation is listed by MINITAB as:

$$I + ABCE + ADEF + BCDF$$

In our representation, the defining relation is expressed as follows:

$$0 = 1235 = 1456 = 2346$$

Management was willing to assume that all three-factor interactions would be quite small in relation to main effects and two-factor interactions. It also recognized that if important two-factor interactions are found to be present, it may be necessary to conduct additional experimental trials to separate confounded interaction effects. Management therefore decided to use the fractional factorial design in Figure 29.8, with four replications added at the center point to provide a rough estimate of the error variance and a test of the fit of the model.

FIGURE 29.8 MINITAB Fractional Factorial Design Summary—Iowa Aluminum Corporation Example.

Fractional Factorial Design

Factors:	6	Design:	6, 16	Resolution:	IV
Runs:	16	Replicates:	1	Fraction:	1/4
Blocks:	none	Center points:	0		

Design Generators: E = ABC F = BCD

Alias Structure

I + ABCE + ADEF + BCDF

A + BCE + DEF + ABCDF
B + ACE + CDF + ABDEF
C + ABE + BDF + ACDEF
D + AEF + BCF + ABCDE
E + ABC + ADF + BCDEF
F + ADE + BCD + ABCEF
AB + CE + ACDF + BDEF
AC + BE + ABDF + CDEF
AD + EF + ABCF + BCDE
AE + BC + DF + ABCDEF
AF + DE + ABCD + BCEF
BD + CF + ABEF + ACDE
BF + CD + ABDE + ACEF
ABD + ACF + BEF + CDE
ABF + ACD + BDE + CEF

Table 29.7 contains the design matrix listed in standard order for the MINITAB fractional factorial design augmented by four replications at the center point. In the right column are shown the results of the experiment. The response of interest is the surface impurity score, where surface impurities are rated on a 0–10 scale (0 = no impurity, 10 = high impurity). The MINITAB output for an initial factorial ANOVA fit is shown in Figure 29.9. Because four replications at the center point were made, an estimate of σ^2 is available. From an initial inspection of the absolute size of the factor effect coefficients and their associated P-values, it appears that the active effects are main effects for oil percentage, coolant volume 3, and strip speed, and the two-factor interaction between coolant temperature and coolant volume 1 (which is confounded with the two-factor interaction between oil percentage and coolant volume 3).

Since this study was exploratory in nature, a new model was developed in which only the factor effects identified as active (X_2, X_5, X_6, $X_{13} = X_{25}$) are retained. An ANOVA model containing the three main effects and one interaction effect was fitted. Residual analysis (not shown) did not reveal any serious departures from the model assumptions. Figure 29.10 contains the MINITAB output for a regression fit of the revised ANOVA model. Note that the lack of fit statistic is shown, $F^* = MSLF/MSPE = .04$, for which the P-value is .9958. Hence, the fit of the revised model appears to be good. We see from the ANOVA output that the statistical significance of the estimated factor effect coefficients b_2, b_5, b_6, and $b_{13} + b_{25}$ is confirmed.

We turn now to the interpretation of the experimental results. Because the β_{13} and β_{25} interaction terms are confounded, the source of this effect cannot be determined on the basis of the experimental results. Notice, however, that both the factor 2 and factor 5 main effects

TABLE 29.7 Experimental Design Matrix and Y Observations—Iowa Aluminum Corporation Example.

	Design Matrix						
Treatment	Coolant Temperature X_1	Oil Percentage X_2	Coolant Volume 1 X_3	Coolant Volume 2 X_4	Coolant Volume 3 X_5	Strip Speed X_6	Impurity Score Y
1	−1	−1	−1	−1	−1	−1	4
2	1	−1	−1	−1	1	−1	6
3	−1	1	−1	−1	1	1	7
4	1	1	−1	−1	−1	1	2
5	−1	−1	1	−1	1	1	3
6	1	−1	1	−1	−1	1	1
7	−1	1	1	−1	−1	−1	5
8	1	1	1	−1	1	−1	9
9	−1	−1	−1	1	−1	1	3
10	1	−1	−1	1	1	1	2
11	−1	1	−1	1	1	−1	8
12	1	1	−1	1	−1	−1	5
13	−1	−1	1	1	1	−1	4
14	1	−1	1	1	−1	−1	4
15	−1	1	1	1	−1	1	4
16	1	1	1	1	1	1	6
17	0	0	0	0	0	0	3
18	0	0	0	0	0	0	5
19	0	0	0	0	0	0	4
20	0	0	0	0	0	0	6

were identified as active and that neither the factor 1 nor the factor 3 main effects were statistically significant. These results suggest (but do not prove) that the observed effect is likely due to the β_{25} interaction. To investigate this further, a small follow-up 2×2 factorial experiment was run involving only factors 1 and 3. No β_{13} interaction effect was found, and it was therefore concluded that the $\beta_{13} + \beta_{25}$ confounded interaction effect in the original experiment is due to the β_{25} interaction effect.

The results of the experiment are summarized in Figure 29.11 by a main effects plot for factor 6 (strip speed) and an interactions plot for factors 2 (oil percentage) and 5 (coolant volume 3). The results can be qualitatively summarized as follows:

1. Figure 29.11a shows that increasing strip speed decreases observed surface impurities. Strip speed should therefore be set at its high level ($X_6 = 1$).

2. Figure 29.11b shows that when oil percentage is at its high level, increasing coolant volume 3 increases surface impurities. When oil percentage is at its low level, increasing coolant volume 3 has relatively little effect on surface impurities. We also see that increasing the oil percentage increases surface impurities; the effect is particularly strong when coolant volume 3 is at its high level. Thus, both oil percentage and coolant volume 3 should be set at their low levels ($X_2 = -1$ and $X_5 = -1$).

FIGURE 29.9 MINITAB Fractional Factorial Output for Initial Model—Iowa Aluminum Corporation Example.

Estimated Effects and Coefficients for Defects

Term	Effect	Coef	Std Coef	t-value	P
Constant		4.550	0.2503	18.18	0.000
Cooltemp	−0.375	−0.187	0.2799	−0.67	0.540
Oilpct	2.375	1.187	0.2799	4.24	0.013
Coolvol1	−0.125	−0.062	0.2799	−0.22	0.834
Coolvol2	−0.125	−0.062	0.2799	−0.22	0.834
Coolvol3	2.125	1.062	0.2799	3.80	0.019
Stripspd	−2.125	−1.062	0.2799	−3.80	0.019
Cooltemp*Oilpct	−0.125	−0.062	0.2799	−0.22	0.834
Cooltemp*Coolvol1	1.375	0.687	0.2799	2.46	0.070
Cooltemp*Coolvol2	−0.125	−0.062	0.2799	−0.22	0.834
Cooltemp*Coolvol3	0.625	0.312	0.2799	1.12	0.327
Cooltemp*Stripspd	−1.125	−0.563	0.2799	−2.01	0.115
Oilpct*Coolvol2	0.125	0.062	0.2799	0.22	0.834
Oilpct*Stripspd	0.125	0.062	0.2799	0.22	0.834
Cooltemp*Oilpct*Coolvol2	0.125	0.062	0.2799	0.22	0.834
Cooltemp*Oilpct*Stripspd	0.125	0.062	0.2799	0.22	0.834

Analysis of Variance for Defects

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	6	59.3750	59.3750	9.89583	7.90	0.033
2-Way Interactions	7	14.4375	14.4375	2.06250	1.65	0.330
3-Way Interactions	2	0.1250	0.1250	0.06250	0.05	0.952
Residual Error	4	5.0125	5.0125	1.25312		
Curvature	1	0.0125	0.0125	0.01250	0.01	0.936
Pure Error	3	5.0000	5.0000	1.66667		
Total	19	78.9500				

We can predict the mean impurity level produced by the process at the optimum (coded) settings of the control variables:

$$\begin{aligned} X_2 &= \text{Oil percentage } = -1 \\ X_5 &= \text{Coolant volume 3} = -1 \\ X_6 &= \text{Strip speed } = 1 \end{aligned} \tag{29.35}$$

by using the fitted regression model equivalent to the final ANOVA model in Figure 29.10:

$$\hat{Y} = 4.5500 + 1.1875X_2 + 1.0625X_5 - 1.0625X_6 + .6875X_{25} \tag{29.36}$$

The estimated impurity response for process setting (29.35) is:

$$\hat{Y}_h = 4.5500 + 1.1875(-1) + 1.0625(-1) - 1.0625(1) + .6875(-1)(-1) = 1.925$$

A confirmation run at the optimum setting can be carried out to assess the validity of the estimated regression function. The validity is supported if the new response falls inside the $1 - \alpha$ prediction limits (6.63). The 95 percent limits turn out to be (see Figure 29.10):

$$-.312 \leq Y_{h(\text{new})} \leq 4.162$$

FIGURE 29.10 MINITAB Fractional Factorial Regression Output for Revised Model—Iowa Aluminum Corporation Example.

The regression equation is
Defects = 4.55 + 1.19 Oilpct + 1.06 Coolvol3 − 1.06 Stripspd + 0.687 Tmp*vol 1

Predictor	Coef	Stdev	t-ratio	p
constant	4.5500	0.2058	22.11	0.000
Oilpct	1.1875	0.2300	5.16	0.000
Coolvol 3	1.0625	0.2300	4.62	0.000
Stripspd	−1.0625	0.2300	−4.62	0.000
Tmp*vol 1	0.6875	0.2300	2.99	0.009

s = 0.9201 R-sq = 83.9% R-sq(adj) = 79.6%

Analysis of Variance

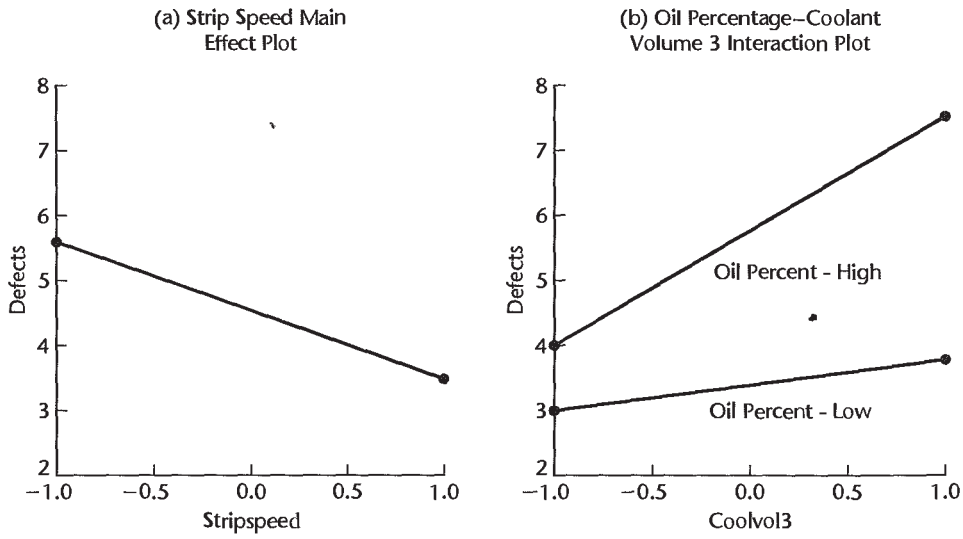
SOURCE	DF	SS	MS	F	p
Regression	4	66.250	16.562	19.56	0.000
Error	15	12.700	0.847		
Total	19	78.950			

SOURCE	DF	SEQ SS
Oilpct	1	22.562
Coolvol 3	1	18.062
Stripspd	1	18.062
Tmp*vol 1	1	7.563

Fit	Stdev. Fit	95.0% C. I.	95.0% P. I.
1.925	0.504	(0.851, 2.999)	(−0.312, 4.162)

Pure error test − F = 0.04 P = 0.9958 DF(pure error) = 11

FIGURE 29.11 Main Effect and Interaction Plots—Iowa Aluminum Corporation Example.



Since the impurity response cannot be negative, the prediction limits should be modified as follows:

$$0 \leq Y_{h(\text{new})} \leq 4.162$$

A new response at the optimum levels less than 4.162 will be consistent with the model's prediction.

29.4 Screening Experiments

In the early stages of an investigation, it is not uncommon for investigators to identify a large number of potential explanatory variables. Unfortunately, the number of model terms required for a large number of factors becomes enormous. For example, in a manufacturing process optimization study, a brainstorming session involving manufacturing engineers, product development scientists, and line operators resulted in the identification of 28 potentially important factors. In addition to 28 parameters for main effects, there would be $28(27)/2 = 378$ parameters for two-factor interactions, $[28(27)(26)]/[2(3)] = 3{,}276$ parameters for three-factor interactions, and there would be many additional parameters for higher-order interactions. Even an investigation of just the main effects and two-factor interactions for 28 factors by use of a resolution IV or a resolution V fractional factorial design would be impossible here.

For these circumstances, screening designs are useful. With these designs, the objective is simply to identify the set of active factors. No information about interactions or curvature is typically obtained. In this section, we shall discuss the use of resolution III fractional factorial designs and Plackett-Burman designs for the purpose of screening large numbers of factors.

2_{III}^{k-f} Fractional Factorial Designs

Recall that in a resolution III fractional factorial design, main effects are confounded with two-factor interactions. If it can be assumed that first-order interactions are small relative to the main effects, then a resolution III design can be used to identify the active factors.

As a simple example, consider a study of three factors, each at two levels, to be conducted with four experimental trials. A half-fraction of highest resolution is obtained by fractionating the 2^3 factorial on the basis of the defining relation:

$$0 = 123$$

The confounding scheme is therefore:

$$1 = 23$$
$$2 = 13$$
$$3 = 12$$

If it can be safely assumed that the two-factor interactions β_{12}, β_{13}, and β_{23} are small in relation to the main effects β_1, β_2, and β_3, then this half-fraction design can be used for identifying the set of active factors.

The use of resolution III designs for initial screening is typically followed by one or more experiments involving those factors that are identified as important. For example, a

10-factor, resolution III experiment (2_{III}^{10-6}) involving 16 experimental trials was used to study the effects of six process variables and four ingredient variables on the extent of crystallization in ice cream. Three factors were identified as important. The interactions among these three factors were then studied in a follow-up 2^3 factorial experiment.

Comment

Any resolution III fractional factorial design can be augmented by a second fraction of the same size to yield a new design of resolution IV or higher. The design matrix for the second fraction is obtained from that for the first fraction by simply reversing all signs. This process is sometimes called *folding over* the first fraction, and the resulting, combined design is sometimes referred to as a *foldover* design. ■

Plackett-Burman Designs

One limitation of resolution III fractional factorial designs is the requirement that the number of treatment combinations be a power of 2. The total experimental trials must therefore be 4, 8, 16, 32, 64, and so on. Plackett-Burman designs are two-level, resolution III designs that can be used for studying up to $n_T - 1$ factors in n_T experimental trials, where n_T is a multiple of 4. Valid run sizes for Plackett-Burman designs are therefore 4, 8, 12, 16, 20, and so on. Plackett-Burman designs for $n_T \leq 100$ are given (with the exception $n_T = 92$) in Reference 29.2. When n_T is a power of 2, the Plackett-Burman designs correspond to the resolution III fractional factorial designs already discussed. When n_T is not a power of 2, the confounding structure of the Plackett-Burman designs is very complicated. Plackett-Burman designs are available in many statistical software packages that provide capabilities for the design of experiments.

The analysis of Plackett-Burman designs is carried out in the same manner as for fractional factorial designs. Since these designs are usually run in a single replication, the various graphical procedures discussed in Section 29.2 can be used to identify active effects. Center point replications can also be added to provide an estimate of the error variance σ^2 and a test for lack of fit.

29.5 Incomplete Block Designs for Two-Level Factorial Experiments

When we considered randomized complete block designs in Chapter 21 and incomplete block designs in Chapter 28, we noted that blocks are chosen so that the experimental units within a block are homogeneous while they differ from block to block. When the number of treatments is large, it may be difficult to find blocks of sufficient size to permit the use of a complete block design. For example, if a block is a mold of four plastic parts, an experiment with eight treatments cannot be run using a mold as a complete block. However, an incomplete block design can be used here, with one-half of the treatments placed in one mold and the other four treatments in a second mold. Incomplete block designs are frequently required in factorial studies with a large number of factors. In this section, we discuss the use of incomplete block designs in two-level factorial experiments. The only

restriction is that the incomplete block size must be a power of 2. We shall start with an example for purposes of illustration.

Example

Steichen Bakeries was developing a partially baked French bread for national distribution. A study was undertaken to investigate the effects of proofing time, proofing temperature, baking time, and baking temperature on the volume and texture of the final product. A two-level, four-factor experiment was under consideration, involving 16 treatments. The production facility could produce from 8 to 10 batches of bread in a given day. Since ambient temperature and humidity in the plant can change significantly from day to day, blocking by day was considered to be important. Hence, an incomplete block design was required such that the 16 treatments are placed into two blocks of size eight. We will now consider how to place the 16 treatments into two blocks.

Assignment of Treatments to Blocks

The design matrix for the 2^4 full factorial study in the Steichen Bakeries example is shown in Table 29.8a. Suppose that the treatments are allocated to blocks in accordance with the level of the 1234 interaction column (X_{1234}). That is, all treatments for which $X_{1234} = -1$ are allocated to block 1 (day 1), and all treatments for which $X_{1234} = 1$ are assigned to block 2 (day 2). With this arrangement, it can be seen that the block effect (i.e., the day effect) will be completely confounded with the four-factor interaction effect. We thus forfeit the ability to obtain an estimate of the four-factor interaction effect β_{1234} that is free of block (day) effects. However, estimates of all main effects, two-factor interactions, and three-factor interactions will be independent of the block effect.

The blocking arrangement chosen by confounding the block effect with the 1234 interaction effect is displayed in Table 29.8a. Notice that each of the four factors appears four times at its low level and four times at its high level within each block. Thus, if ambient temperature is exceptionally high on day 1, causing the loaves of bread baked on that day to have volumes that are larger than usual, this effect will not bias the estimates of any of the main effects. It can be verified that the same balance of high and low levels (1s and -1s) within each block is also present for all interaction columns except for the X_{1234} column.

The analysis of the experiment is identical to that of a full 2^4 factorial study. The only difference concerns the interpretation of results, where it must be remembered that the four-factor interaction effect is confounded with the block (day) effect.

In general, blocking of factorial and fractional factorial designs is accomplished by confounding block effects with carefully chosen, high-order interaction effects. The division of treatments into blocks is performed in three steps:

1. Identify the high-order interaction effects to be confounded with the block effects. If the number of desired blocks is $b = 2^v$, v interaction effects need to be identified.
2. Construct the v columns of the X matrix that correspond to the interaction effects chosen. The patterns of 1s and -1s in these columns are used to identify the blocks.
3. The v interaction effects chosen, along with their generalized interactions, are confounded with the block effects. In all, $b - 1$ effects are so confounded.

TABLE 29.8 Blocking Arrangements—Steichen Bakeries Example.

(a) 2^4 Experiment in Two Blocks

Block (Day)	Treatment	Proofing Time X_1	Proofing Temperature X_2	Baking Time X_3	Baking Temperature X_4
1	1	1	−1	−1	−1
1	2	−1	1	−1	−1
1	3	−1	−1	1	−1
1	4	1	1	1	−1
1	5	−1	−1	−1	1
1	6	1	1	−1	1
1	7	1	−1	1	1
1	8	−1	1	1	1
2	9	−1	−1	−1	−1
2	10	1	1	−1	−1
2	11	1	−1	1	−1
2	12	−1	1	1	−1
2	13	1	−1	−1	1
2	14	−1	1	−1	1
2	15	−1	−1	1	1
2	16	1	1	1	1

(b) 2^4 Experiment in Four Blocks

Block (Day)	Treatment	Proofing Time X_1	Proofing Temperature X_2	Baking Time X_3	Baking Temperature X_4
1	1	1	1	−1	−1
1	2	−1	−1	1	−1
1	3	−1	1	−1	1
1	4	1	−1	1	1
2	5	−1	−1	−1	−1
2	6	1	1	1	−1
2	7	1	−1	−1	1
2	8	−1	1	1	1
3	9	−1	1	−1	−1
3	10	1	−1	1	−1
3	11	1	1	−1	1
3	12	−1	−1	1	1
4	13	1	−1	−1	−1
4	14	−1	1	1	−1
4	15	−1	−1	−1	1
4	16	1	1	1	1

In effect, this procedure fractionates the chosen design v times, and the 2^v resulting fractions define the divisions of treatments into blocks. This will result in 2^v blocks of size 2^{k-v} in the case of a full factorial study, or 2^v blocks of size 2^{k-f-v} in the case of a 2^{k-f} fractional factorial study.

As when constructing fractional factorial designs, the v interactions selected to define the blocks must be carefully chosen so that, to the greatest extent possible, low-order effects remain clear of block effects. Useful blocking arrangements have been catalogued (Ref. 29.1). They are also usually provided by statistical software packages that have capabilities for the design of experiments.

Example

In the Steichen Bakeries example, the investigator wished to run the 2^4 factorial study in four blocks. Here, the number of blocks is $b = 4 = 2^v$, so that $v = 2$. Thus, two higher-order interaction effects that are to be confounded with block effects need to be chosen for identifying the treatments assigned to the blocks. The investigator chose interactions 23 and 124. The treatments were then assigned to blocks in the following fashion:

Value of X_{23}	Value of X_{124}	Treatment Assigned to
−1	−1	Block 1
1	−1	Block 2
−1	1	Block 3
1	1	Block 4

Since there are $b = 4$ blocks, $b - 1 = 3$ factor effects are confounded with block effects. These are the 23 interaction, the 124 interaction, and their generalized interaction:

$$23 \times 124 = 12^2 34 = 134$$

The resulting design is shown in Table 29.8b. Notice again the balance of levels within each block: each factor appears twice at its high level and twice at its low level. This will also be true for all interaction columns except X_{23}, X_{124}, and X_{134}; these columns will be constant within each block. An abbreviated ANOVA table is shown in Table 29.9. Note that this table shows the confounding of the three interaction effects, 23, 124, and 134, with blocks.

Use of Center Point Replications

We noted earlier that two or more replications are often added at the center point when the factors are quantitative to provide an estimate of the error variance σ^2 and a test for lack of fit. When blocking is used, center point replications must be placed within the same block to obtain a valid measure of pure error. Otherwise, differences in responses will be due to both experimental error and block-to-block differences. Use of an equal number of center point replications in each block leads to all estimated factor (and block) effect coefficients being uncorrelated.

TABLE 29.9 Abbreviated ANOVA Table—Steichen Bakeries Example.

Source of Variation	*df*
X_0	1
X_1	1
X_2	1
X_3	1
X_4	1
X_{12}	1
X_{13}	1
X_{14}	1
X_{24}	1
X_{34}	1
X_{123}	1
X_{234}	1
X_{1234}	1
Blocks (confounded with X_{23}, X_{124}, X_{134})	3
Error	0
Total	16

29.6 Robust Product and Process Design

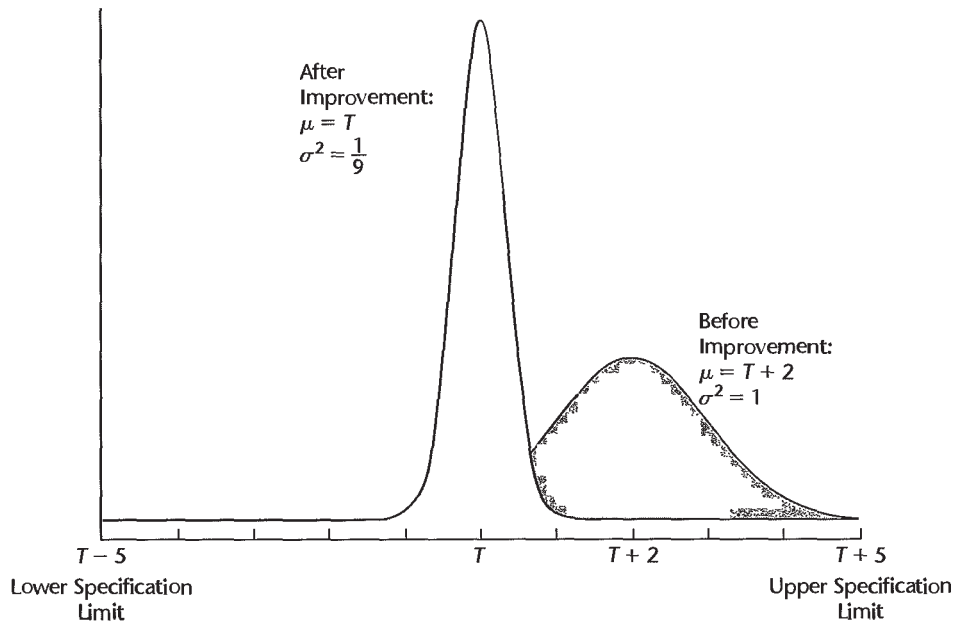
In recent years, the importance of reducing variation in products and processes has been widely recognized. Uncontrolled variation leads to waste, disruption, duplication of effort, decreased consumer satisfaction, and/or the need for inspection and rework. Thus, experimental studies are often designed to identify process or product designs that exhibit low levels of variation. Such product designs are called *robust,* because they produce a desired result in a consistent, repeatable fashion. The basic framework for using designed experimentation to develop robust products and processes was popularized by Dr. Genichi Taguchi, a Japanese quality consultant, in the 1980s. It is sometimes referred to generally as the "Taguchi Method" (Ref. 29.3).

For instance, in the manufacture of color television sets, an important performance characteristic or outcome measurement is the color density. We will assume that there is a best, or *target,* color density T. Ideally, all televisions would be produced with color density T. However, due to natural variations in materials, equipment, operators, or other aspects of the manufacturing process, the actual color densities Y will deviate from the target. While any television with a color density within ± 5 units of T was considered acceptable, the manufacturer found that *any* deviation from target decreased customer satisfaction. For this reason the manufacturer concluded that manufacturing televisions within specification was not sufficient. Customer satisfaction would be maximized if the absolute deviations from actual color density to target, $Dev = |Y - T|$, or the squared deviations $Dev^2 = |Y - T|^2$, were consistently small.

Taguchi observed that the average squared deviation from target is given by the mean squared error:

$$E\{Dev^2\} = E\{(Y - T)^2\} \tag{29.37}$$

FIGURE 29.12 Process Distributions Before and After Product Design Experiment—Color Television Example.



We encountered the mean squared error in Chapter 9 in connection with Mallow's C_p and again in Chapter 10 in connection with ridge regression. It can be shown [as we did earlier in (9.6)] that the mean squared error can be written as a sum of the variance of Y and the square of the *off-target distance* or bias, $(\mu - T)^2$:

$$\begin{aligned} E\{Y - T\}^2 &= \sigma^2\{Y\} + (E\{Y\} - T)^2 = \sigma^2 + (\mu - T)^2 \\ &= \text{Variance} + (\text{Off-Target Distance})^2 \end{aligned} \tag{29.38}$$

Figure 29.12 shows two process distributions for television color density. The distribution on the right is the process distribution of color density before a robust product design experiment was performed. In this case, color density, Y, follows a normal distribution with mean $\mu = T + 2$ and variance $\sigma^2 = 1$. The distribution on the left shows the process distribution following the experiment. Here, color density follows a normal distribution with mean $\mu = T$ and variance $\sigma^2 = 1/9$. Note that both distributions fall largely within the product specification limits $T \pm 5$; however, prior to experimentation, the mean squared error was:

$$E\{Dev^2\} = \sigma^2 + (\mu - T)^2 = 1 + 2^2 = 5$$

After the product design experiment, the mean squared error was reduced to:

$$E\{Dev^2\} = \sigma^2 + (\mu - T)^2 = \frac{1}{9} + 0^2 = \frac{1}{9}$$

Thus on average, the color densities of television sets for the robust design are much closer to target than those based on the previous design.

The implication of (29.38) for designed experimentation is as follows. In any test of alternative process or product designs, a "best" treatment combination will lead to a treatment mean that is close to target with minimal variance. Experiments are therefore conducted in such a way that two linear statistical models—one for the mean and one for the variance of the response—can be estimated. These estimated models are then used to identify robust factor-level settings—those that lead to a process mean μ that is close to target, with small process variance σ^2.

In this section, we first introduce a strategy for developing models for both the mean and the variance of the response. We then consider the use of special nuisance factors, called *noise factors,* in the construction of robust product design experiments. Noise factors are used to develop products and processes that are robust to specific, known sources of variation.

Location and Dispersion Modeling

As already noted, in robust product design experiment, a "best" factor-level combination leads to a response distribution with a small variance and a mean that is close to target. We shall assume that k-factor model (29.2a) is applicable, except that we will no longer assume that the error variance is constant. In addition, because one of our objectives is to model the variance response, we will assume that $n > 1$ complete replicates of the experiment have been conducted. Let Y_{ij} denote the response of the jth replicate for the ith treatment combination, for $i = 1, \ldots, r$ and $j = 1, \ldots, n$. Our model is now:

$$Y_{ij} = \beta_0 X_{i0} + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + \beta_{12} X_{i12} + \cdots + \beta_{12\cdots k} X_{i12\cdots k} + \varepsilon_{ij} \qquad \textbf{(29.39)}$$

where:

$$X_{il} = \begin{cases} -1 & \text{if case } i \text{ from first level of factor } l \\ 1 & \text{if case } i \text{ from second level of factor } l \end{cases}$$

$$X_{ik\cdots m} = X_{ik} X_{il} \cdots X_{im}$$

and ε_{ij} are independent $N(0, \sigma_i^2)$.

Denote the sample variance obtained for the ith treatment combination by s_i^2:

$$s_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{i\cdot})^2 \qquad \textbf{(29.40)}$$

The sample variance is the response to be modeled in the *dispersion model.* The raw responses, Y_{ij} are modeled directly using (29.39). We refer here to (29.39) as the *location model* because it provides for estimates of the mean response as a function of the control-factor-level settings. We now consider the development of these models, beginning with the dispersion model.

Dispersion Model. The dispersion model is based on (29.39), where the response Y_i is replaced by the logarithm of the ith sample variance. We also attach the superscript D to the regression parameters and to the error terms as a reminder that these quantities pertain only to the dispersion model:

$$\log_e s_i^2 = \beta_0^D X_{i0} + \beta_1^D X_{i1} + \cdots + \beta_k^D X_{ik} + \beta_{12}^D X_{i12} + \cdots + \beta_{12\cdots k}^D X_{i12\cdots k} + \varepsilon_i^D \qquad \textbf{(29.41)}$$

The regression parameters $\beta^D_{i\ldots k}$ are referred to as the *dispersion effects*. The reason that we use the $\log_e s_i^2$ as the response rather than s_i^2 is that the latter do not follow normal distribution with constant variance. Since the ε_{ij} are normally distributed with zero mean and variance σ_i^2 it follows from (A.70) that $(n-1)s_i^2/\sigma_i^2$ is distributed as χ^2 with $(n-1)$ degrees of freedom. It can be shown that $\log_e s_i^2$ is approximately normally distributed with mean $\log_e \sigma_i^2$ and constant variance $2/(n-1)$ (see, e.g., Reference 29.4). Thus, the ε_i^D are approximately independent and normally distributed with constant variance. Model (29.41) can then be estimated using ordinary least squares and the methods discussed in Section 29.2 for the analysis of unreplicated two-level studies.

Location Model. The location model is given by (29.39). However, because the variance is not constant, the parameters are most efficiently estimated using weighted least squares as described in Section 11.1. Specifically, we obtain an estimate of the variance for each factor-level combination using:

$$\hat{v}_i = \exp\left(\widehat{\log_e s_i^2}\right) \tag{29.42}$$

where $\widehat{\log_e s_i^2}$ is obtained from the estimated dispersion model (29.41). Then the weights are given by (11.16b) on page 425:

$$w_i = \frac{1}{\hat{v}_i} \tag{29.43}$$

Alternatively, an approximate analysis can be conducted based on ordinary least squares.

Strategy for Analysis. We suggest the following strategy for analyzing the location and dispersion models:

1. Fit dispersion model (29.41) and determine whether or not dispersion effects are present. This can be done using methods discussed in Section 29.2 for the analysis of unreplicated two-level factorials, or the Breusch-Pagan test (3.11) for constancy of error variance.
2. If the variance is constant, there is no need to fit the dispersion model (29.41). The location model (29.39) can then be analyzed using ordinary least squares and the methods described in previous sections.
3. If dispersion effects are present, fit location model (29.39) using weighted least squares, or conduct an approximate unweighted analysis.
4. Use the resulting models based on the active location and dispersion effects to identify factor-level combinations that move the predicted mean close to target while minimizing the predicted variance. If no dispersion effects are present, only the location model is employed. Similarly, if no location effects are present, only the dispersion model is employed.

In Step 4, if a factor is active—either through its main effect or through interactions involving the factor—in only one of the two models, the selection of optimal level setting can be conducted according to the model in which the factor is active. If a factor is active in both models, it might not be possible to find a factor-level combination that simultaneously produces an optimal mean and an optimal variance. In this case, a compromise setting is identified that leads to "good" (but not necessarily optimal) results for both the mean and the variance.

We illustrate the use of the modeling strategy with an adaptation of an example due to Taguchi.

Example

A food company investigated alternative recipes for a type of caramel. The performance characteristic of interest was the plasticity of the caramel. When subjected to sufficient shearing stress, any given caramel will be deformed. If, after the stress is removed, there is no recovery, the caramel is completely plastic. On the other hand, if recovery is complete and instantaneous, the caramel is completely elastic. A proper balance between these two factors is required. In the experiment, the plasticity was measured on a scale of 1 to 100, where 100 implies the complete plasticity. The target value of the caramel was 70.

Three ingredients were thought to be potentially important: brown sugar (X_1), sweetened condensed milk (X_2), and light corn syrup (X_3). The first three columns in Table 29.10 list the coded treatment combinations for the 2^3 full factorial design in standard order, and columns 4 through 7 provide the the levels of the interaction columns X_{12}, X_{13}, X_{23}, and X_{123}. Four replicates of the experiment were obtained, and the four Y_{ij} responses for each treatment combination are listed in columns 8–11. Also listed in Table 29.10 in columns 12 and 13 are the sample variances s_i^2 and their logarithms $\log_e s_i^2$.

The first step in the analysis was to fit dispersion model (29.41). Results obtained from a regression of column 13 in Table 29.10 on columns 1–7 are shown in Figure 29.13a. Since there are no replicates for dispersion model (29.41), t-values and P-values cannot be obtained. Figures 29.13b provides a normal probability plot of the estimated dispersion effect coefficients. The plot clearly suggests the presence of one nonzero dispersion effect, namely β_{13}^D. Ignoring inactive effects, the estimated dispersion model is:

$$\widehat{\log_e s_i^2} = 4.0098 + .5748X_{i13} \tag{29.44}$$

Since dispersion effects are present, we move to Step 3 of the strategy for analysis, which calls for the use of weighted least squares (or an approximate analysis using ordinary least squares) to estimate the parameters in the location effects model. We will illustrate the use of weighted least squares using an estimated variance function, as described in Section 11.1.

A model-based estimate of the variance for the ith treatment combination is, from (29.44):

$$\begin{aligned}\hat{v}_i &= \exp\left(\widehat{\log_e s_i^2}\right)\\ &= \exp(4.0098 + .5748X_{i13})\end{aligned}$$

TABLE 29.10 Experimental Design Matrix and Y Observations—Caramel Example.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Design Matrix							Replicates						
i	X_1	X_2	X_3	X_{12}	X_{13}	X_{23}	X_{123}	Y_{i1}	Y_{i2}	Y_{i3}	Y_{i4}	s_i^2	$\log_e s_i^2$	w_i
1	−1	−1	−1	1	1	1	−1	42	65	70	73	197.67	5.287	.0102
2	1	−1	−1	−1	−1	1	1	50	52	55	63	32.67	3.486	.0322
3	−1	1	−1	−1	1	−1	1	61	70	78	79	70.00	4.248	.0102
4	1	1	−1	1	−1	−1	−1	48	51	55	60	27.00	3.296	.0322
5	−1	−1	1	1	−1	−1	1	65	74	74	77	27.00	3.296	.0322
6	1	−1	1	−1	1	−1	−1	40	59	63	66	136.67	4.918	.0102
7	−1	1	1	−1	−1	1	−1	70	72	77	84	38.92	3.662	.0322
8	1	1	1	1	1	1	1	48	49	56	63	48.67	3.885	.0102

FIGURE 29.13 MINITAB Regression Output and Normal Probability Plot of Estimated Effect Coefficients for Dispersion Model—Caramel Example.

(a) Regression Coefficients

Term	Effect	Coef
Constant		4.009B
X1	−0.2270	−0.1135
X2	−0.4740	−0.2370
X3	−0.1390	−0.0695
X1*X2	−0.1375	−0.0687
X1*X3	1.1495	0.5748
X2*X3	0.1405	0.0702
X1*X2*X3	−0.5620	−0.2810

(b) Normal Probability Plot

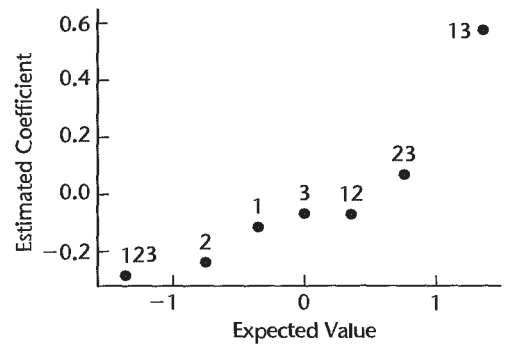


FIGURE 29.14 MINITAB Regression Output and Normal Probability Plot of Estimated Effect Coefficients for Location Model—Caramel Example.

(a) Regression Output

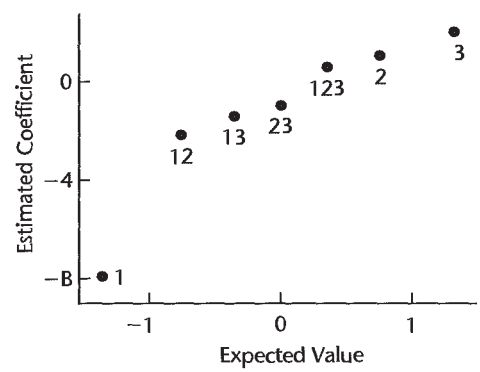
Predictor	Coef	SE Coef	T	P
Constant	62.781	1.478	42.48	0.000
x1	−7.906	1.478	−5.35	0.000
x2	1.031	1.478	0.70	0.492
x3	2.031	1.478	1.37	0.182
x12	−2.156	1.478	−1.46	0.158
x13	−1.406	1.478	−0.95	0.351
x23	−0.969	1.478	−0.66	0.518
x123	0.594	1.478	0.40	0.691

S = 1.041 R-Sq = 69.9% R-Sq(adj) = 61.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	7	60.341	8.620	7.96	0.000
Residual Error	24	25.993	1.083		
Total	31	86.334			

(b) Normal Probability Plot



From (11.16b), the ith estimated weight is $w_i = 1/\hat{v}_i$. For example, for the first treatment combination in Table 29.10, we obtain:

$$\hat{v}_1 = \exp(4.0098 + .5748X_{113}) = \exp(4.0098 + .5748(1)) = 97.96$$

from which we obtain the first weight: $w_1 = 1/97.96 = .0102$.

Use of the estimated weights listed in column 14 of Table 29.10 in a regression of the Y_{ij} responses in columns 8–10 on the predictors in columns 1–7 led to the weighted least squares location effects estimates summarized in the regression output in Figure 29.14a. Note that the P-value for b_1 is 0+, while the P-values for the remaining effects are all greater than 0.1. The normal probability plot of the estimated location effects in Figure 29.14b also suggests that β_1 is nonzero. Using weighted least squares to estimate the reduced location

model, we obtain (output not shown):

$$\hat{Y}_i = 63.511 - 8.960X_{i1} \tag{29.45}$$

With the estimated dispersion and location models in hand, we now turn to Step 4 in the strategy for analysis—the identification of robust factor-level combinations. From (29.44), two possible optimal settings that minimize the dispersion effect are $(X_1, X_3) = (+1, -1)$ and $(X_1, X_3) = (-1, +1)$. However, the result from the location model in (29.45) shows that, in order to move the estimated mean response to $T = 70$, the optimal setting for X_1 is -1. Thus, the optimal setting in the caramel example is: $(X_1, X_3) = (-1, +1)$. These settings lead to the following estimated mean and variance of caramel plasticity:

$$\hat{Y}_i = 63.511 - 8.960(-1) = 72.5$$

$$\widehat{\log_e s^2} = 4.0098 + .5748(-1)(+1) = 3.435$$

Thus the estimated mean has been moved to within 2.5 of the target $T = 70$. The estimated variance for this setting is exp(3.435)= 31.03.

Comments

1. In some cases, there are factors that are active only in the location model and not in the dispersion model. These factors are called *adjustment factors*. A common strategy is to select optimal settings according to the dispersion model, and then use the adjustment factors to bring the location to the target. Of course, there is no guarantee that adjustment factors exist.

2. The location model can be classified into three groups with respect to the target value: *the-smaller-the-better, the-larger-the-better,* and *the-nominal-the-better*. For instance, an automotive company conducted an experiment to study the effect of four factors on the braking distance in different driving conditions. Since the braking distance should be minimized, it is an example of the-smaller-the-better case. In another study, the response was the pull strength of truck seat belts following a crimping operation. The pull strength needs to be maximized to ensure that the seat belt does not break in an accident. Thus, it is an example of the-larger-the-better case. The procedures of the analysis in these two cases are the same as those shown in the caramel example, which is the-nominal-the-better.

3. The approach to weighted least squares described here for fitting the location model used a model-based estimate of the variance, $\hat{v}_i$, to obtain weights. A simple alternative is to use the sample variances s_i^2, in which case the weights are $w_i = 1/s_i^2$. This approach is discussed in Section 11.1.

■

Incorporating Noise Factors

As we have seen, dual-response modeling can be a powerful tool for identifying product or process designs that have low levels of variation. Recall from our discussion of blocking that variation is often caused by changes in background nuisance factors that cannot be controlled. In the caramel example, plasticity is affected by the ambient temperature. If the temperature changes during the course of the experiment, this would likely contribute to the variation in plasticity observed for each factor-level combination. In a manufacturing process control experiment, if different operators are responsible for different parts of the experiment, they may contribute to variation in the quality of the parts or products produced. In robust product design experiments, the investigator often is interested in reducing variation attributable to one or more specific nuisance factors. In simple terms, this is

accomplished by deliberately changing the levels of the nuisance factors during the course of the experiment, and then identifying settings of the experimental factors for which the response is relatively unaffected by changes to the nuisance factors.

In robust product design terminology, a nuisance factor that is deliberately varied during the experiment is called a *noise factor*. The standard (non-nuisance) experimental factors are termed *control factors*. Generally, control factors are variables that are easy or inexpensive to control in the design of the product or process. Noise factors are variables that are hard or expensive to control during manufacturing or during product use.

Consider again the caramel example. Suppose that the investigator was concerned specifically with the effect of temperature on the plasticity of the product when used by the consumer. Suppose also that the investigator was interested in four temperature levels, namely, 60°F, 70°F, 80°F, and 90°F. In this case, temperature would simply be added as a fourth (four-level) factor (X_4) in the experiment. The three control factors would be brown sugar (X_1), sweetened condensed milk (X_2), and light corn syrup (X_3). In the experiment, each factor-level combination of the control factors (X_1, X_2, X_3) is tested at the four levels of temperature. This is accomplished by crossing the levels of the control factors with the levels of the noise factors, leading here to the use of a $2^3 \times 4$ full factorial design. For purposes of analysis, the four responses obtained for each combination of the control factors are treated simply as replicates, and a dual-response analysis, as already described, is carried out. Control factor settings that lead to a small variance s_i^2 are unaffected by—and therefore robust to—the changes to levels of the noise factor.

Noise factors can arise during the manufacturing process or when the product is in use. *Internal noise* refers to variations that occur during the production process. Examples include raw material variation, manufacturing variation, unit-to-unit variation, and so on. Making product performance insensitive to these variations can improve the quality of the product while lowering the cost of production.

External noise refers to variations that occur when the product is used by the customer. Examples include the environment in which a product works, the load to which it is subjected, and natural deterioration. For instance, a reliable automobile should perform consistently whether it is used in Florida in the summer or Minnesota in the winter. A good washer should be robust to the laundry load. Making product performance insensitive to external variations will improve the reliability of the product and increase the customer satisfaction.

In summary, the basic procedure for incorporating noise factors into a robust product design experiment is as follows.

1. Identify the experimental layout for the control factors. This may be a full factorial or a fractional factorial, blocked or unblocked, depending on the experimenter's objectives, as discussed in Sections 29.1–29.6.
2. Identify the noise factors and associated noise-factor levels to be included in the experiment. If there is more than one noise factor, identify the factor-level combinations of the noise factors to be included. Generally these are obtained from a full factorial layout among the noise factors. However, fractional factorial arrangements of the noise factors are sometimes employed if many noise factors are present.
3. The full experimental design is obtained by crossing the control-factor-level combinations with the noise-factor-level combinations. As always, the resulting treatment

TABLE 29.11 Layout of the Experimental Design with Noise Factor—Caramel Example.

				Noise Factor					
Run	X_1	X_2	X_3	60°F	70°F	80°F	90°F	s_i^2	$\log_e s_i^2$
1	−1	−1	−1	42	65	70	73	197.67	5.287
2	1	−1	−1	50	52	55	63	32.67	3.486
3	−1	1	−1	61	70	78	79	70.00	4.248
4	1	1	−1	48	51	55	60	27.00	3.296
5	−1	−1	1	65	74	74	77	27.00	3.296
6	1	−1	1	40	59	63	66	136.67	4.918
7	−1	1	1	70	72	77	84	38.92	3.662
8	1	1	1	48	49	56	63	48.67	3.885

combinations are randomly assigned to the experimental units. Note that if there are n_c control-factor-level combinations and there are n_n noise-factor-level combinations, there will be $n_c n_n$ treatment combinations in all.

4. The analysis is conducted using the dual-response-optimization strategy outlined on page 1247. The n_n responses obtained for each control-factor-level combination are treated as replicates.

We illustrate the use of a single noise factor by continuing our discussion of the caramel example. We then move on to a more extensive case study from the automotive industry, which employed five control factors and two noise factors.

Caramel Example. In the caramel example, the four responses at a given control-factor-level combination were actually obtained at the four temperatures: 60°F, 70°F, 80°F, and 90°F. Note that we cannot control the temperature in the field, but by controlling it during the experiment, we can identify the settings of the control factors that lead to the desired plasticity across all levels of temperature—that is, with small variance.

The layout of the experimental design matrix is shown in Table 29.11. This is essentially the same design layout as the one shown in Table 29.10. The only difference is that the replications of each control-factor-level combination are conducted deliberately at different levels of the temperature. The steps in the analysis are identical to those shown previously, leading to (29.44) for the dispersion model, and (29.45) for the location model. Thus the setting with brown sugar at the low level ($X_1 = -1$) and light corn syrup at the high level ($X_3 = 1$) leads to a product that has the desired mean plasticity and is relatively unaffected by or robust to changes in temperature.

We now turn to a discussion of a robust product design experiment from the automotive industry.

Case Study—Clutch Slave Cylinder Experiment

A research project in a major automotive company was conducted to develop a design for a clutch slave cylinder that would minimize fluid leakage. Five two-level control factors and two two-level noise factors were identified. The five control factors are body inner diameter (X_1), body outer diameter (X_2), seal inner diameter (X_3), seal outer diameter (X_4), and seal design ($X_5 = -1$: lip seal; $X_5 = 1$: quads seal). Two noise factors are: temperature (X_6) and load ($X_7 = -1$: light; $X_7 = 1$: heavy). The response is leakage, which is to be minimized.

TABLE 29.12 Experimental Design and Responses—Clutch Slave Cylinder Example.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Control Factors					Noise Factors (X_6, X_7)					
i	X_1	X_2	X_3	X_4	X_5	(−1, −1)	(+1, −1)	(−1, +1)	(+1, +1)	$\log_e s_i^2$	w_i
1	−1	−1	−1	−1	−1	.8	.4	0	0	−1.920	1.245
2	1	−1	−1	1	−1	3.2	0	0	0	.940	.195
3	−1	1	−1	1	−1	0	0	0	2.4	.365	1.245
4	1	1	−1	−1	−1	5.8	0	0	2.8	2.036	.195
5	−1	−1	1	1	−1	0	3.0	0	2.4	.912	1.245
6	1	−1	1	−1	−1	0	1.2	0	4.0	1.270	.195
7	−1	1	1	−1	−1	0	2.6	0	1.2	.425	1.245
8	1	1	1	1	−1	1.0	2.3	5.2	0	1.627	.195
9	−1	−1	−1	−1	1	9.8	2.5	13.8	2.0	3.500	.009
10	1	−1	−1	1	1	6.4	3.0	13.0	0	3.440	.058
11	−1	1	−1	1	1	8.8	2.0	31.0	.4	5.294	.009
12	1	1	−1	−1	1	1.8	3.4	6.9	0	2.152	.058
13	−1	−1	1	1	1	6.8	2.4	26.4	0	4.970	.009
14	1	−1	1	−1	1	4.0	2.2	12.6	3.4	3.120	.058
15	−1	1	1	−1	1	10.2	1.8	38.8	3.2	5.697	.009
16	1	1	1	1	1	7.8	1.4	6.4	5.6	2.026	.058

The experimental plan is shown in Table 29.12. For the five control factors, a resolution IV design was used, in which the defining relation is $0 = 1234$. For each control factor setting, four responses were obtained, corresponding to the $2^2 = 4$ noise-factor-level settings. When the control-factor-level combinations are crossed with the noise-factor-level combinations we obtain a $2^{5-1} \times 2^2$ robust product design experiment.

Again following the dual-response modeling strategy on page 1247, we first estimate the dispersion model. Because the design in the control factors is a resolution IV fractional factorial design based on the defining relation $0 = 1234$, the following dispersion effects are confounded:

$$\begin{array}{llll} \beta_0^D + \beta_{1234}^D & \beta_1^D + \beta_{234}^D & \beta_2^D + \beta_{134}^D & \beta_3^D + \beta_{124}^D \\ \beta_4^D + \beta_{123}^D & \beta_5^D + \beta_{12345}^D & \beta_{12}^D + \beta_{34}^D & \beta_{13}^D + \beta_{24}^D \\ \beta_{14}^D + \beta_{23}^D & \beta_{15}^D + \beta_{2345}^D & \beta_{25}^D + \beta_{1345}^D & \beta_{35}^D + \beta_{1245}^D \\ \beta_{45}^D + \beta_{1235}^D & \beta_{125}^D + \beta_{345}^D & \beta_{135}^D + \beta_{245}^D & \beta_{145}^D + \beta_{235}^D \end{array} \tag{29.46}$$

We will form dispersion model (29.41) here by choosing the first dispersion effect from each of the 16 pairs in (29.46):

$$\log_e s_i^2 = \beta_0^D + \beta_1^D X_{i1} + \cdots + \beta_{145}^D X_{i45} + \varepsilon_i^D \tag{29.47}$$

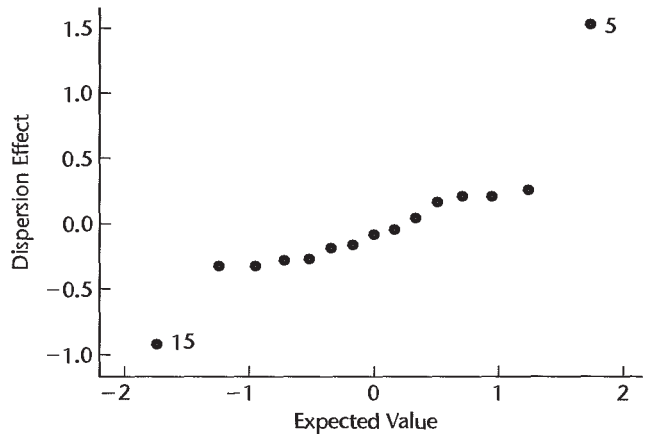
Regressing the $\log_e s_i^2$ values in column 10 of Table 29.12 on the predictors indicated by (29.47), we obtain the estimated dispersion effects shown in Figure 29.15a. A normal probability plot of the estimated dispersion effects is shown in Figure 29.15b. It can be seen that the main dispersion effect of factor X_5 and two-factor interaction X_{15} appear to

FIGURE 29.15 MINITAB Regression Output and Normal Probability Plot of Estimated Effect Coefficients for Dispersion Model—Clutch Slave Cylinder Example.

(a) Regression Output

Term	Effect	Coef
Constant		2.2409
X1	−0.3290	−0.1645
X2	0.4237	0.2119
X3	0.5300	0.2650
X4	0.4117	0.2059
X5	3.0680	1.5340
X1*X2	−0.6560	−0.3280
X1*X3	−0.6612	−0.3306
X1*X4	−0.5480	−0.2740
X1*X5	−1.8517	−0.9259
X2*X5	−0.3890	−0.1945
X3*X5	−0.1733	−0.0866
X4*X5	−0.0965	−0.0482
X1*X2*X5	−0.5698	−0.2849
X1*X3*X5	0.0815	0.0407
X1*X4*X5	0.3298	0.1649

(b) Normal Probability Plot of the Effects



be active. Eliminating the inactive effects leads to the estimated subset dispersion model:

$$\widehat{\log_e s_i^2} = 2.241 + 1.534X_{i5} - .926X_{i15} \qquad \textbf{(29.48)}$$

from which we obtain the model-based variance estimates: $\hat{v}_i = \exp(\widehat{\log_e s_i^2})$.

Since significant dispersion effects are present, we turn now to the estimation of the location model using weighted least squares. The estimated weights, which as before are the inverses of the estimated variances in (29.43), are shown in column 11 of Table 29.12. Use of these estimated weights in a regression of the Y_{ij} responses in columns 6–9 on the predictors indicated in (29.47) leads to the weighted least squares location effects estimates summarized in the regression output in Figure 29.16a. The output indicates that only one estimated location effect, b_5 is significant at the $\alpha = .05$ level of significance. The normal probability plot of the estimated location effects in Figure 29.16b also clearly suggests that β_5 is the only active location effect. Using weighted least squares to estimate the reduced location model, we obtain:

$$\hat{Y}_i = 3.232 + 2.235X_{i5} \qquad \textbf{(29.49)}$$

We now turn to Step 4 in the analysis strategy—the identification of robust control-factor-level combinations. Note that factor X_5 enters both the dispersion model (29.48) and location model (29.49) as a main effect with a positive coefficient. The predicted dispersion and location are both to be minimized in this example; we therefore set $X_5 = -1$. For dispersion model (29.48), X_5 also enters through the interaction term X_1X_5. Since the estimated dispersion interaction effect is $b_{15} = -.926$, minimization is accomplished by setting $X_1X_5 = 1$. With $X_5 = -1$ from the location model, we have $X_1(-1) = 1$, implying $X_1 = -1$. These settings lead to predicted mean fluid leakage:

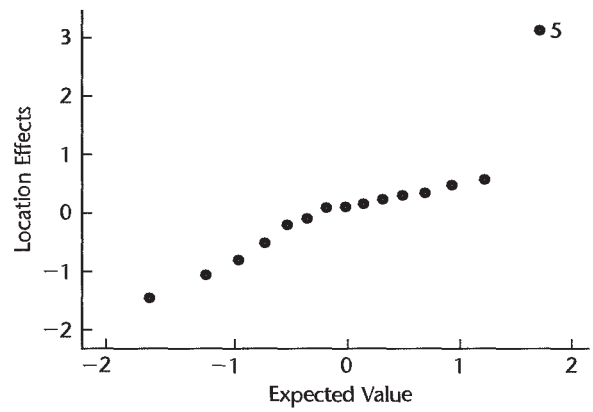
$$\hat{Y} = 3.232 + 2.235(-1) = .997$$

FIGURE 29.16 MINITAB Regression Output and Normal Probability Plot of Estimated Effect Coefficients for Location Model—Clutch Slave Cylinder Example.

(a) Regression Results

Predictor	Coef	SE Coef	T	P
Constant	4.3141	0.8256	5.23	0.000
X1	−1.0828	0.8256	−1.31	0.196
X2	0.4609	0.8256	0.56	0.579
X3	0.5578	0.8256	0.68	0.502
X4	0.0891	0.8256	0.11	0.915
X5	3.1172	0.8256	3.78	0.000
x12	−0.5422	0.8256	−0.66	0.514
x13	−0.2203	0.8256	−0.27	0.791
x14	0.1359	0.8256	0.16	0.870
x15	−1.4797	0.8256	−1.79	0.079
x25	0.2016	0.8256	0.24	0.808
x35	0.3234	0.8256	0.39	0.697
x45	0.0672	0.8256	0.08	0.935
x125	−0.8266	0.8256	−1.00	0.322
x135	−0.1047	0.8256	−0.13	0.900
x145	0.2891	0.8256	0.35	0.728

(b) Normal Probability Plot



with predicted variance:

$$\hat{v} = \exp[2.241 + 1.534(-1) - .926(-1)(-1)] = .803$$

Note that prediction intervals for these quantities can be obtained in the usual way. Often, a confirmation test is carried out at the suggested factor-level combination as a check on the validity of the model. The model is said to be confirmed if the results of the confirmation run fall within the calculated prediction limits.

Comments

1. An alternative approach to the dual-response optimization approach discussed here, called the *response modeling approach,* was proposed by Welch et al. (Ref. 29.5) and Shoemaker et al. (Ref. 29.6). This approach advocates, as a first step, the usual analysis of the experiment, making no distinction between noise and control factors. If significant interactions exist that involve both noise and control factors, these interactions are analyzed through graphical or other means to determine which control-factor-level combinations lead to the desired mean responses and are relatively unaffected by changes to the noise factors.

2. In the framework proposed by Taguchi, the analysis of a robust design model involves the *signal-to-noise ratio,* which is a transformation based on $\bar{Y}_{i.}$ and s_i^2 (Ref. 29.3). Since then, many other analysis methods have been proposed, but the location-dispersion modeling and the response-modeling approaches are often preferred by statisticians. For a more detailed discussion, see Reference 29.4.

3. The control factor layout chosen by the engineer in the clutch slave cylinder example was a resolution IV design. Table 29.6 indicates that a design with higher resolution was available, namely the 2_{V}^{5-1} design based on the defining relation $0 = 12345$. ■

Cited References

29.1. Box, G. E. P., W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*. New York: John Wiley & Sons, 1978.

29.2. Plackett, R. L., and J. P. Burman. "The Design of Optimum Multifactorial Experiments," *Biometrika* 33 (1946), pp. 305–25.

29.3. Taguchi, G. *Introduction to Quality Engineering*. Tokyo, Japan: Asian Productivity Organization, 1986.

29.4. Wu, C. F. J., and M. Hamada. *Experiments: Planning, Analysis, and Parameter Design Optimization*. New York: John Wiley & Sons, 2000.

29.5. Welch, W. J., T. K. Yu, S. M. Kang, and J. Sacks. "Computer Experiments for Quality Control by Parameter Designs," *Journal of Quality Technology* 40 (1990), pp. 62–71.

29.6. Shoemaker, A. C., K. L. Tsui, and C. F. J. Wu. "Economical Experimentation Methods for Robust Design," *Technometrics* 33 (1991), pp. 415–27.

Problems

29.1. A plant manager used a 2^4 factorial design with two replicates for each treatment to study the effects of four process variables $(X_1, \ldots, X_4)$ on product quality (Y). State the response model in the form of (29.2a). How many two-factor interaction terms are there? How many three-factor interaction terms? How many four-factor interaction terms?

29.2. A scientist observed: "Two-level factorial designs are useful if the number of factors is small. But I am concerned when there are 10 or more factors; the number of trials required for a 2^{10} experiment is simply too large." Discuss.

*29.3. **Reaction yield.** A chemical engineer decided to employ a single replicate of a 2^6 factorial design to study the effects of the process variables on the yield of a chemical reaction.

a. How may factors are involved? How many levels are there for each factor? How many experimental trials will be required for the single replicate of the experiment?

b. Can a test for lack of fit be obtained here?

29.4. A biologist considered studying the effects of various environmental pollutants on the health of mice by using a 2^{7-4} fractional factorial design.

a. How many factors are involved? How many levels are there for each factor? How many trials will be required for a single replicate of the experiment? Can a test for lack of fit be obtained?

b. The biologist decided to augment the design with six center-point replicates. Can a test for lack of fit now be obtained? If so, can the biologist determine which factors caused a curvature effect?

29.5. State the **X** matrix (including all main effects and interaction columns) for a single replicate of a 2^3 factorial design, with the rows listed in standard order. Show numerically that (29.3) holds for your **X** matrix.

*29.6. Refer to **Reaction yield** Problem 29.3. Past experience indicates that the standard deviation of reaction yield is $\sigma = 5$.

a. Find the variance of the estimated main effect coefficient b_1. Is the variance of the interaction effect coefficient b_{12} the same? Should it be?

b. How many replicates of the experiment are required in order to estimate factor effect coefficent b_1 within $\pm .5$ with 95 percent confidence?

*29.7. **Pilot training.** An unreplicated 2^5 full factorial design was used to investigate the effects of five factors on the learning rates of flight trainees when using flight simulators. The factors were display type ($X_1 = -1$: symbolic; $X_1 = 1$: pictorial), display orientation ($X_2 = -1$:

outside in; $X_2 = 1$: inside out), crosswind ($X_3 = -1$: no wind present; $X_3 = 1$: crosswind present), command guidance ($X_4 = -1$: constant guidance; $X_4 = 1$: guidance only when trainee strays far from best flight path), and flight path prediction ($X_5 = -1$: no prediction; $X_5 = 1$: constant prediction). The response Y is the average squared distance from the optimal flight path for 12 landing attempts by the trainee. The smaller is Y, the better is the trainee's performance. Thirty-two subjects (trainees) were selected at random from a large group of trainees with no prior flying experience. The design matrix for the experiment and the observed trainee flight scores (Y) follow.

Y	X_1	X_2	X_3	X_4	X_5
8.69	−1	−1	−1	−1	−1
7.71	1	−1	−1	−1	−1
9.03	−1	1	−1	−1	−1
...	...	...	...	...	...
6.67	1	−1	1	1	1
2.78	−1	1	1	1	1
7.45	1	1	1	1	1

Adapted in part from L. Lintern et al., "Display Principles, Control Dynamics, and Environmental Factors in Pilot Training and Transfer," *Human Factors* 32 (1990), pp. 64–69.

a. State the regression model in the form (29.2a). Fit this model and obtain the estimated factor effect coefficients. Does it appear from the magnitudes of the estimated coefficients that some factors may be active here?

b. Prepare a dot plot of the estimated factor effect coefficients. Which effects appear to be active?

c. Obtain a normal probability plot of the estimated factor effect coefficients. Which effects appear to be active? Do the estimated factor effects appear to be normally distributed? How do your results compare with those in parts (a) and (b)?

*29.8. Refer to **Pilot Training** Problem 29.7. The regression model was revised by dropping all three-factor and higher-order interactions.

a. State the revised regression model. Fit the revised regression model and prepare a plot of the residuals against the fitted values. Do the standard regression assumptions appear to be satisfied?

b. Obtain a normal probability plot of the residuals. Also conduct the correlation test for normality; use $\alpha = .05$. Does the assumption of normality appear to be reasonable here?

c. Using the P-values for the estimated factor effect coefficients, test for the significance of each factor effect. Control the family level of significance at $\alpha = .05$ using the Kimball inequality. Which effects appear to be active?

d. Summarize the results of the experiment with an appropriate set of plots of main effects and interactions. Interpret the results.

29.9. **Computer monitors.** A single replicate of a 2^4 full factorial design, augmented by three replicates at the center point, was used to determine the most reliable design of a computer monitor base. Factors of interest were clearance under the base (X_1), interface board height (X_2), side vent size (X_3), and interface board angle (X_4). All factors are quantitative and are coded with $X_i = -1$ for the low level of the factor and $X_i = 1$ for the high level. The response (Y) is the failure rate of the interface board, with lower failure rates representing higher product quality. The design matrix for the experiment and the observed design failure

rates (Y) follow.

Y	X_1	X_2	X_3	X_4
3.88	−1	−1	−1	−1
3.17	1	−1	−1	−1
4.07	−1	1	−1	−1
...	...	...	...	...
3.80	0	0	0	0
3.99	0	0	0	0
4.16	0	0	0	0

a. State the regression model in the form (29.2a). Fit this model and obtain the estimated factor effect coefficients. Does it appear from the magnitudes of the estimated coefficients that some factors may be active here?

b. Prepare a dot plot of the estimated factor effect coefficients. Which effects appear to be active?

c. Obtain a normal probability plot of the estimated factor effect coefficients. Which effects appear to be active? Do the estimated factor effects appear to be normally distributed? How do your results compare with those in parts (a) and (b)?

d. Obtain *MSPE* using the three center-point replicates and (29.17). Use this estimate to determine the P-value for each estimated factor effect coefficient. Determine which effects are active; use $\alpha = .05$ for each test.

29.10. Refer to **Computer monitors** Problem 29.9. The regression model was revised by including only the main effects of factors 1, 3, and 4 and the 34 interaction.

a. Fit the revised model and prepare a plot of the residuals against the fitted values. Do the standard regression assumptions appear to be satisfied?

b. Obtain a normal probability plot of the residuals. Also conduct the correlation test for normality; use $\alpha = .05$. Does the assumption of normality appear to be a reasonable one here?

c. Using the P-values for the estimated factor effect coefficients, test for the significance of each effect; use $\alpha = .01$ for each test. Which effects are active?

d. Conduct a test for lack of fit; use $\alpha = .05$. State the decision rule and conclusion.

e. Summarize the results of the experiment with an appropriate set of plots of main effects and interactions. Interpret the results. How should the monitor base be designed to achieve a minimum failure rate?

29.11. Refer to the **X** matrix for a 2^4 full factorial design in Table 29.2.

a. Identify the defining relation for the fractional design obtained by dropping treatments 3 to 6, 9, 10, 15, and 16. What is the resolution of the fractional design so obtained?

b. Give the complete confounding scheme for the fractional design obtained in part (a).

29.12. a. Construct a design for four two-level factors with eight experimental trials that has the highest possible resolution. What is the resolution of this design?

b. Verify the projection property for the design constructed in part (a) that any subset of three (or fewer) factors yields a full factorial design in those factors.

29.13. Is it possible to construct a resolution III design for four two-level factors with four experimental trials? If so, construct such a design. If not, indicate why this is not possible.

29.14. Construct a 2_{III}^{4-1} design using the defining relation $0 = 123$. Is there an alternative eight-run design of higher resolution?

*29.15. Obtain the complete defining relation and the confounding scheme for the eight-run, five-factor design that is fractionated on the basis of the relation $0 = 123 = 245$. What is the resolution of this design? Is there an alternative design with higher resolution?

29.16. The following design matrix was used in an eight-run, five-factor experiment:

X_1	X_2	X_3	X_4	X_5
−1	−1	−1	−1	−1
1	−1	−1	−1	1
−1	1	−1	1	1
1	1	−1	1	−1
−1	−1	1	1	1
1	−1	1	1	−1
−1	1	1	−1	−1
1	1	1	−1	1

Obtain the defining relation and the complete confounding scheme for this design. What is the resolution of this design? Can an alternative five-factor, eight-run design with higher resolution be constructed?

29.17. Construct a 2^{6-3} fractional factorial design of highest resolution using Table 29.6. What is the defining relation for this design? What is its resolution?

*29.18. **Peanut solids.** A food scientist conducted a single replicate of a 2^{7-3} fractional factorial design in an effort to identify factors that affect the extraction of food solids from peanuts using water. Factors of interest were the pH level of the water ($X_1 = -1$: 6.95; $X_1 = 1$: 8.00), water temperature ($X_2 = -1$: 20°C; $X_2 = 1$: 60°C), extraction time ($X_3 = -1$: 15 minutes; $X_3 = 1$: 40 minutes), water-to-peanuts ratio ($X_4 = -1$: 5; $X_4 = 1$: 9), agitation speed ($X_5 = -1$: 5,000 rpm; $X_5 = 1$: 10,000 rpm), hydrolysis ($X_6 = -1$: unhydrolyzed; $X_6 = 1$: hydrolyzed), and presoaking level ($X_7 = -1$: dry; $X_7 = 1$: soaked). The experimental units were 16 randomly selected batches of peanuts. The response (Y) is the percentage of the total solids removed from each batch. The defining relation used to construct the 2^{7-3} fractional design (excluding generalized interactions) is $0 = 1235 = 2346 = 1247$. The design matrix for the experiment and the observed percentage extractions (Y) follow.

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7
10.82	−1	−1	−1	−1	−1	−1	−1
10.59	1	−1	−1	−1	1	−1	1
8.19	−1	1	−1	−1	1	1	1
...	...	...	...	...	...	...	...
5.12	1	−1	1	1	−1	−1	−1
5.60	−1	1	1	1	−1	1	−1
5.73	1	1	1	1	1	1	1

Adapted from I. Y. S. Rustom et al., "A Study of Factors Affecting Extraction of Peanut (*Arachis hypogaea* L.) Solids with Water," *Food Chemistry* 42 (1991), pp. 153–65.

a. Obtain the generalized interactions and the complete defining relation. What is the resolution of the design? Could a design of higher resolution have been used here?

b. Using the defining relation in part (a), determine the confounding pattern for all main effects and two-factor interactions.

c. State the regression model in the form (29.2a). Remember that confounded effects must not be included in your model. Fit this model and obtain the estimated factor effect coefficients. Prepare a dot plot of the estimated factor effect coefficients. Which effects appear to be active?

d. Obtain a normal probability plot of the estimated factor effect coefficients. Which effects appear to be active? Do the estimated effects appear to be normally distributed? How do your results compare with those in part (c)?

e. Test whether all two-factor interaction effects can be dropped from the model; use $\alpha = .01$. State the alternatives, decision rule, and conclusion.

*29.19. Refer to **Peanut Solids** Problem 29.18. The regression model was revised by dropping all interaction effects.

a. Fit the revised model and prepare a plot of the residuals against the fitted values. Do the standard regression assumptions appear to be satisfied?

b. Cases 3 and 14 have fairly large absolute residuals. Conduct the Bonferroni outlier test for each of these cases; use $\alpha = .05$ for each test. What do you conclude?

c. Obtain a normal probability plot of the residuals. Also conduct the correlation test for normality; use $\alpha = .025$. Does the assumption of normality appear to be reasonable here?

d. Using the P-values of the estimated factor effect coefficients, test for the significance of each effect; use $\alpha = .02$ for each test. Which effects are active?

e. Summarize the results of the experiment with an appropriate set of plots of main effects. Interpret the results. How should maximum food solids extraction be achieved?

29.20. **Fiber optics.** A chemist conducted a screening experiment to identify factors that affect the viscosity of a gel used in the manufacture of fiber optic cabling. To minimize the loss of telephone signal, the inner glass fibers must be allowed to move freely within the cabling for a range of temperatures. A lubricant (gel) is used to promote this movement. The viscosity of the gel must be sufficiently low to allow such movement; yet it must not be so low as to lead to dripping (leakage) from the ends. A single replicate of a 2^{9-5} fractional factorial design was conducted. The factors of interest were silica particle size ($X_1 = -1$: 200; $X_1 = 1$: 380), silica weight ($X_2 = -1$: low; $X_2 = 1$: high), oil ratio ($X_3 = -1$: low; $X_3 = 1$: high), oil temperature ($X_4 = -1$: low; $X_4 = 1$: high), stabilizer level ($X_5 = -1$: low; $X_5 = 1$: high), premix time ($X_6 = -1$: short; $X_6 = 1$: long), postmix time ($X_7 = -1$: short; $X_7 = 1$: long), postmix vacuum ($X_8 = -1$: no; $X_8 = 1$: yes), and filter mesh size ($X_9 = -1$: small; $X_9 = 1$: large). The response of interest is gel viscosity (Y); management feels that an optimal (target) gel viscosity is 74.5. The design matrix for the experiment and the observed viscosities (Y) follow.

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
101.2	1	1	1	1	−1	1	−1	−1	1
92.9	−1	−1	−1	−1	−1	−1	−1	−1	1
129.9	1	1	1	1	1	−1	−1	−1	−1
...	...	...	...	...	...	...	...	...	...
73.4	−1	−1	−1	−1	1	1	−1	−1	−1
31.6	1	1	−1	−1	−1	−1	1	−1	−1
121.6	1	−1	1	−1	1	−1	−1	1	1

Adapted from T. L. Reed, "Quality Improvement of Silica-Based Polysiloxane Gel Used in Fiber Optic Cabling by Process Optimization via Taguchi Methods," *Fifth Symposium on Taguchi Methods*, Detroit: ASI Press (1987), pp. 555–71.

a. State the regression model containing only factor main effects in the form (29.2a). Fit this model and obtain the estimated factor effect coefficients. Does it appear from the magnitudes of the estimated coefficients that some factors may be active here?

b. Prepare a Pareto plot of the estimated factor effect coefficients. Which effects appear to be active?

c. Obtain a normal probability plot of the estimated factor effect coefficients. Which effects appear to be active? Do the estimated factor effects appear to be normally distributed? How do your results compare with those in part (b)?

d. Using the P-values of the estimated factor effect coefficients, test for the significance of each effect term; use $\alpha = .10$ for each test. Which effects are active?

29.21. Refer to **Fiber optics** Problem 29.20. The regression model was revised to include only the main effects for factors 1, 5, and 7.

a. Fit the revised regression model and prepare a plot of the residuals against the fitted values. Do the standard regression assumptions appear to be satisfied?

b. Obtain a normal probability plot of the residuals. Also conduct the correlation test for normality; use $\alpha = .05$. Does the assumption of normality appear to be reasonable here?

c. Conduct a lack of fit test for the revised regression model; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What does your conclusion suggest about the possible presence of interactions?

29.22. Refer to **Fiber optics** Problems 29.20 and 29.21. Since the experimental design consists of two complete replicates of a 2^3 factorial in the three active factors 1, 5, and 7, consider now a revised model containing the main effects of factors 1, 5, and 7 and all interactions among these three factors.

a. State the revised regression model and fit it. Using the P-values of the estimated factor effect coefficients, test for the signficance of each factor effect; use $\alpha = .01$ for each test. Which effects are active?

b. Obtain a normal probability plot of the residuals. Compare this plot to that obtained in Problem 29.21b. What do you conclude?

c. Summarize the experimental results with an appropriate set of plots of the main effects and interactions. Interpret the results.

d. How might you proceed to determine the levels of factors 1, 5, and 7 so that the expected viscosity of the resulting gel would be on target at 74.5?

29.23. **Windshield molding manufacture.** An experimental study was undertaken in an effort to reduce the occurrence of dents in a windshield molding manufacturing process. The dents are caused by pieces of metal or plastic that are carried into the dies during stamping and forming operations. Four factors were identified for use in an eight-run experiment: poly-film thickness—used to protect the metal strip during manufacturing to reduce surface blemishes ($X_1 = -1$: .00175; $X_1 = 1$: .0025), oil mixture ratio for surface lubrication ($X_2 = -1$: .05; $X_2 = 1$: .10), operator glove type ($X_3 = -1$: cotton; $X_3 = 1$: nylon), underside oil coating ($X_4 = -1$: no coating; $X_4 = 1$: coating). During each run of the experiment, 1,000 moldings were fabricated; the response (Y) is the number of defect-free moldings produced. The design matrix for the experiment and the observed numbers of defect-free moldings produced (Y) follow.

Y	X_1	X_2	X_3	X_4
338	1	−1	−1	−1
826	1	−1	1	1
350	1	1	−1	−1
647	1	1	1	1
917	−1	−1	−1	1
977	−1	−1	1	−1
953	−1	1	−1	1
972	−1	1	1	−1

Adapted from G. Adel, "Minimize Slugging by Optimizing Controllable Factors on Topaz Windshield Molding," *Fifth Symposium on Taguchi Methods*, Detroit: ASI Press (1987), pp. 519–26.

a. Determine the defining relation and the complete confounding scheme used in the experiment. Could a design of higher resolution have been used?

b. State the regression model in the form (29.2a). Remember that confounded factor effects must not be included in your model. Fit this model and obtain the estimated factor effect coefficients.

c. Prepare a dot plot of the estimated factor effect coefficients. Which effects appear to be active?

d. Obtain a normal probability plot of the estimated factor effect coefficients. Which effects appear to be active? How do your results compare with those in part (c)? Do the estimated factor effects appear to be normally distributed?

29.24. Refer to **Windshield molding manufacture** Problem 29.23. The regression model was revised to include only the main effects for the four factors.

a. Fit the revised regression model. Using the P-values for the estimated factor effect coefficients, test for the significance of each effect; use $\alpha = .05$ in each case. Which effects are active?

b. Summarize the results of the experiment with an appropriate set of plots of main effects. Interpret the results. Identify the settings of the experimental factors within the operating range that lead to the maximum number of defect-free moldings.

29.25. Construct a 2_{III}^{5-2} design in two blocks of size four such that main effects are not confounded with the block effect.

*29.26. **Team effectiveness.** A researcher employed a single replicate of a 2^6 full factorial design, with eight blocks containing eight treatments each, to study the effects of team member's ability level and motivation level on the performance of three-person military teams consisting of an operator, a loader, and a mover. The factors studied were operator's ability (X_1), operator's motivation (X_2), loader's ability (X_3), loader's motivation (X_4), mover's ability (X_5), and mover's motivation (X_6). All factors are quantitative and are coded with $X_i = -1$ referring to the low level of the factor and $X_i = 1$ referring to its high level. The 64 teams were formed by assigning persons to teams in accordance with the 2^6 full factorial design.

The team ratings (Y) were assigned by unit commanders following two months of military activity. Because unit commanders could observe at most 10 teams, and because it was expected that some scoring biases might result, the teams were assigned to commanders in blocks of size eight. Levels of the interaction terms X_{135}, X_{146}, and X_{245} were used to determine the blocks. The observed team ratings, the design matrix, and the blocking arrangement follow.

Y	Block	X_1	X_2	X_3	X_4	X_5	X_6
43	1	−1	−1	−1	−1	−1	−1
61	1	1	1	1	1	−1	−1
60	1	−1	1	1	−1	1	−1
...	...	...	...	...	...	...	...
66	8	1	−1	−1	1	−1	1
64	8	−1	−1	−1	−1	1	1
91	8	1	1	1	1	1	1

Adapted in part from A. E. Tziner, "Effects of Team Composition on Ranked Team Effectiveness," *Small Group Behavior* 19 (1988), pp. 363–78.

a. Obtain a scatter plot of team ratings against block number. Does it appear that blocking was effective here?

b. Identify the complete confounding scheme for blocks. Are any main effects confounded with blocks? Any two-factor interactions?

c. State the regression model in the form (29.2a). Fit this model and obtain the estimated factor effect coefficients. Prepare a dot plot of the estimated factor effect coefficients. Which effects appear to be active?

d. Obtain a normal probability plot of the estimated factor effect coefficients. Which effects appear to be active? How do your findings compare with those in part (c)? Do the estimated factor effects appear to be normally distributed?

*29.27. Refer to **Team effectiveness** Problem 29.26. The regression model was revised to include only the factor main effects, two-factor interactions, and block main effects.

a. Fit the revised model and prepare a plot of the residuals against the fitted values. Do the standard regression assumptions appear to be satisfied?

b. Obtain a normal probability plot of the residuals. Also conduct the correlation test for normality; use $\alpha = .05$. Does the assumption of normality appear to be reasonable here?

c. Using the P-values for the estimated factor effect coefficients, test for the significance of each factor effect; use $\alpha = .01$ for each test. Which effects are active?

*29.28. Refer to **Team effectiveness** Problems 29.26 and 29.27. The finally revised regression model consists of all block main effects and all factor main effects only.

a. Fit the finally revised regression model.

b. Summarize the results of the experiment with an appropriate set of plots of the factor main effects. Interpret the results. How is maximum team effectiveness achieved?

c. Obtain a 95 percent prediction interval for the team performance for a single new team formed as described in part (b); assume that the rater (block) effect is zero in making your prediction.

29.29. **Whipped topping.** Food scientists had developed a prototype soybean-based whipped topping, but the product suffered in that the volume of the whipped product did not meet expectations. In an effort to maximize the topping volume, a 2^{5-1} fractional factorial design of highest resolution was used in an experiment in two blocks of size eight each, with three center-point replicates in each block. The design confounded the block effect with the 45 interaction. The factors studied were soybean solids level (X_1), fat level (X_2), emulsifier level (X_3), and the levels of two stabilizers: methocel (X_4), and avicel (X_5). All factors are quantitative and are coded with $X_i = -1$ referring to the low level of the factor and $X_i = 1$ referring to its high level. The response (Y) is the percent increase in volume of the product due to whipping; large increases are desirable. The observed responses, the design matrix, and the blocking

arrangement follow.

Y	Block	X_1	X_2	X_3	X_4	X_5
124	1	−1	−1	−1	−1	1
144	1	1	1	−1	−1	1
144	1	1	−1	1	−1	1
...	...	...	...	...	...	
121	2	0	0	0	0	0
127	2	0	0	0	0	0
115	2	0	0	0	0	0

a. What is the defining relation for this design? What is the resolution, ignoring blocks?

b. State the regression model in the form (29.2a). Remember that confounded factor effects must not be included in your model. Fit this regression model and obtain the estimated factor effect coefficients. Prepare a dot plot of the estimated factor effect coefficients. Which effects appear to be active?

c. Obtain a normal probability plot of the estimated factor effect coefficients. Which effects appear to be active? How do your results compare with those in part (b)? Do the estimated factor effects appear to be normally distributed?

d. Test for the presence of block effects; use $\alpha = .05$. State the alternatives, decision rule, and conclusion.

e. Fit a revised regression model, omitting the block effect term. Obtain a pure error estimate of the error variance using the six center-point replicates and (29.17) and conduct a test for lack of fit; use $\alpha = .05$. State the decision rule and conclusion. Does your test indicate the presence of curvature?

f. Using the P-values for the estimated factor effect coefficients obtained in part (e) based on the pure error estimate $MSPE$, test for the significance of the factor effects; use $\alpha = .025$ for each test. Which factors are active?

29.30. Refer to **Whipped topping** Problem 29.29. The model has been finally revised to include only the main effects for factors 1, 2, and 5 and the 12 interaction term.

a. Fit the revised model and prepare a plot of the residuals against the fitted values. Do the standard regression assumptions appear to be satisfied?

b. Obtain a normal probability plot of the residuals. Also conduct the correlation test for normality; use $\alpha = .05$. Does the assumption of normality appear to be reasonable here?

c. Summarize the results of the experiment with an appropriate set of plots of the main effects and interactions. Interpret the results. How is maximum whippability achieved?

d. Obtain a 95 percent confidence interval for the expected percent volume increase for the whipped topping product when formulated as recommended in part (c).

29.31. Refer to **Computer monitors** Problem 29.9. Suppose two more replicates were conducted for the 2^4 full factorial design. Ignoring the center points, the design matrix for the new experiment with three replicate responses Y_{i1}, Y_{i2}, and Y_{i3} follows. Assume that the target failure rate is $T = 0$.

i	X_1	X_2	X_3	X_4	Y_{i1}	Y_{i2}	Y_{i3}
1	−1	−1	−1	−1	3.88	3.10	5.30
2	1	−1	−1	−1	3.17	2.75	4.90
...	...	...	...	...	...	...	...
16	1	1	1	1	3.11	1.82	3.95

a. Obtain the sample variances and the logarithms of the sample variances for each of the control-factor-level combinations. Does the variance appear to be constant?

b. Fit the dispersion model (29.41) using the logarithm of the sample variances obtained in part (a). Prepare a Pareto plot of the estimated factor effect coefficients. Which dispersion effects appear to be active?

c. Using the subset dispersion model based on the estimates of the active dispersion effects, provide estimates of the variance of the response for each control-factor-level combination. Are your estimates consistent with the sample variances obtained in part (a)?

d. Fit the location model (29.39) using weighted least squares. Obtain a normal probability plot of the estimated control-factor-effect coefficients. Which effects appear to be active? Use $\alpha = .05$.

e. Using the subset dispersion and location models based on the active dispersion and location effects identified in parts (b) and (d), determine the control factor settings that minimize failure rate with minimum variance.

f. Give 95 percent confidence limits for the predicted variance for the optimal settings identified in part (e). How would these limits be used in a confirmation run?

g. Estimate the mean squared error in (29.38) for the optimal control-factor-level settings determined in part (e).

*29.32. **Leaf springs.** An engineer conducted an experiment to identify factors that affect the height of an unloaded spring to improve a heat treatment process on truck leaf springs. The target value of the height (Y) is $T = 8$ inches. The heat treatment forms the camber (curvature) in leaf springs, and was conducted by heating in a high temperature furnace, processing by a forming machine, and quenching in an oil bath. The factors of interest were furnace temperature ($X_1 = -1$: 1840°F; $X_1 = 1$: 1880°F), heating time ($X_2 = -1$: 23 minutes; $X_2 = 1$: 25 minutes), transfer time ($X_3 = -1$: short; $X_3 = 1$: long), and hold-down time ($X_4 = -1$: short; $X_4 = 1$: long). The defining relation used to construct the 2^{4-1} design is $0 = 1234$. The design matrix for the experiment and the observed heights with 6 replicates (Y) follow.

i	X_1	X_2	X_3	X_4	Y_{i1}	Y_{i2}	$\cdots$	Y_{i6}
1	−1	−1	−1	−1	7.56	7.62	$\cdots$	7.25
2	1	−1	−1	1	7.56	7.81	$\cdots$	7.59
3	−1	1	−1	1	7.84	7.70	$\cdots$	7.20
4	1	1	−1	−1	7.69	8.09	$\cdots$	7.20
5	−1	−1	1	1	7.50	7.56	$\cdots$	7.50
6	1	−1	1	−1	7.59	7.56	$\cdots$	7.56
7	−1	1	1	−1	7.78	7.83	$\cdots$	7.12
8	1	1	1	1	8.15	8.10	$\cdots$	7.25

Adapted in part from J. J. Pignatiello and J. S. Ramburg, "Discussion of 'Off-Line Quality Control, Parameter Design, and the Taguchi Method' by Kackar, R. N.," *Journal of Quality Technology*, 17, pp. 198–206.

a. Obtain the sample variances and the logarithms of the sample variances for each of the control-factor-level combinations. Does the variance appear to be constant?

b. Fit the dispersion model (29.41) using the logarithms of the sample variances obtained in part (a). Prepare a Pareto plot of the estimated factor effect coefficients. Which dispersion effects appear to be active?

c. Using the subset dispersion model based on the estimates of the active dispersion effects, provide estimates of the variance of the response for each control-factor-level combination. Are your estimates consistent with the sample variances obtained in part (a)?

d. Fit the location model (29.39) using weighted least squares. Obtain a normal probability plot of the estimated control-factor-effect coefficients. Which effects appear to be active? Use $\alpha = .05$.

e. Using the subset dispersion and location models based on the active dispersion and location effects identified in parts (b) and (d), determine the control factor settings that lead to a predicted mean height near $T = 8$ with minimal variance.

f. Give simultaneous 95 percent confidence limits for the predicted variance for the optimal settings identified in part (e). How would these limits be used in a confirmation run?

g. Estimate the mean squared error in (29.38) for the optimal control-factor-level settings determined in part (e).

Exercises

29.33. Show that (29.14) holds for balanced two-level experiments; use (2.51) and the additivity of the extra sums of squares in this situation.

29.34. Suppose that the true (full) regression model in matrix form is:

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$$

However, the analyst assumes that the (reduced) model:

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}$$

is correct and uses it for purposes of estimation. For example, the $\mathbf{X}$ matrix for the reduced model ($\mathbf{X}_1$) might include only an intercept column and columns for first-order terms, while the true model involves first-order terms ($\mathbf{X}_1$) and some two-factor interaction terms ($\mathbf{X}_2$).

a. Show that:

$$\mathbf{E}\{\hat{\boldsymbol{\beta}}_1\} = \boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2$$

where $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$ is called the alias matrix.

b. Let $\mathbf{X}_1$ be the $\mathbf{X}$ matrix (based on the intercept and first-order terms only) for the 2_{III}^{3-1} design constructed from the defining relation $0 = 123$. Let $\mathbf{X}_2$ consist of the columns X_{12}, X_{13}, and X_{23}, corresponding to the omitted two-factor interaction effects β_{12}, β_{13}, and β_{23}. Use the result in part (a) and $\mathbf{b} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y} = \mathbf{X}_1'\mathbf{Y}/8$ to show that $E\{b_1\} = \beta_1 + \beta_{23}$, $E\{b_2\} = \beta_2 + \beta_{13}$, and $E\{b_3\} = \beta_3 + \beta_{12}$. Thus, for this design we have: $1 = 23$, $2 = 13$, and $3 = 12$.

Chapter 30

Response Surface Methodology

Chapter 29 was devoted to a discussion of the design of two-level factorial experiments. With these designs, main effects and two-factor interactions can often be studied with relatively few experimental trials. One limitation of two-level designs for factorial studies where the factors are quantitative is that they cannot identify curvatures in the response surface. Modeling curvature effects can be very important when the objective of the experiment is to identify the combination of levels of the quantitative factors that leads to an optimum response. Response surface experiments can be used for this purpose. In this chapter, we discuss the design and analysis of response surface experiments for studies where the factors are quantitative. Response surface designs are generally used in the latter stages of an investigation, when five or fewer factors are under investigation.

30.1 Response Surface Experiments

When a factorial study involves quantitative factors and the shape of the response surface is of interest, the response surface is usually approximated by a second-order regression model. The rationale is that the main effects and second-order effects will generally capture the essence of the response function since third-order and higher effects are usually unimportant.

The second-order response function for three quantitative factors was given in (8.10). We shall generalize it now for k quantitative factors. We continue to use the special coding employed in Chapter 29 for the level X_j of the jth quantitative factor:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + \beta_{11} X_1^2 + \cdots + \beta_{kk} X_k^2 + \beta_{12} X_1 X_2 + \cdots + \beta_{k-1,k} X_{k-1} X_k \tag{30.1}$$

where the level X_j of the jth factor is coded as follows:

$$X_j = \frac{\text{Actual Level} - \dfrac{\text{High Level} + \text{Low Level}}{2}}{\dfrac{\text{High Level} - \text{Low Level}}{2}} \tag{30.2}$$

This coding scheme results in a coded value of -1 for the low level of factor j, a coded value of 1 for the high level, a coded value of 0 for the midlevel, and so on. For instance, if the temperature levels of factor j in a study range from 75° to 85°, the following coded values X_j will be used:

Temperature Level	Coded Value X_j
75	-1
78	$-.4$
80	0
85	1

Occasionally, the experimental design will be supplemented with treatments consisting of factor levels outside the original range. This will result in coded values below -1 or above 1. For instance, if a supplemental treatment in our example involves factor j at temperature level 70°, the coded value will be $X_j = (70 - 80)/5 = -2$.

As before, the coefficients $\beta_1, \ldots, \beta_k$ in regression model (30.1) are the linear main effect coefficients, the coefficients $\beta_{11}, \ldots, \beta_{kk}$ are the quadratic main effect coefficients, and the coefficients $\beta_{12}, \beta_{13}, \ldots, \beta_{k-1,k}$ are the interaction effect coefficients. Notice that model (30.1) involves $p = 1 + k + k + k(k-1)/2 = (k+1)(k+2)/2$ regression parameters.

When designing a response surface study, a minimal requirement is that the design must be capable of providing estimates of the $p = (k+1)(k+2)/2$ parameters in model (30.1). Any design of resolution V or higher for a two-level factorial study will provide estimates of linear main effects and all two-factor interaction effects that are confounded only with higher-order effects. However, at least three levels of each factor must be present to obtain estimates of the k quadratic main effects.

One type of design that provides estimates of all parameters in regression model (30.1) is the full factorial design with each factor at three levels. Full factorial designs with each factor at three levels are referred to as 3^k designs, where k denotes the number of factors in the study. A number of practical limitations are associated with 3^k designs. The first is expense. The number of treatments required by a 3^k design grows rapidly with the number of factors. For four factors, for instance, a three-level full factorial design consists of $3^4 = 81$ treatments. A second disadvantage is that each factor appears at exactly three levels so that it will not be possible to test for the presence of cubic or higher-order main effects.

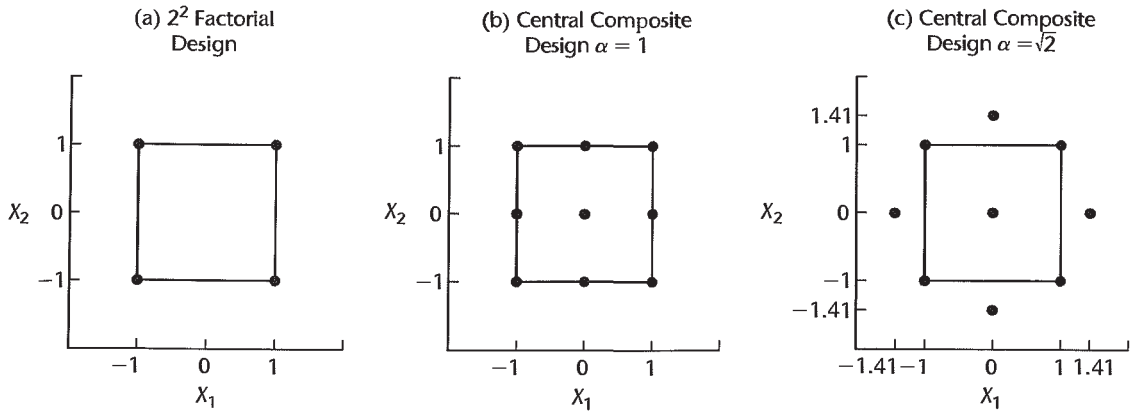
In Sections 30.2 and 30.3, we shall discuss a variety of response surface designs that have been developed for estimation of response surfaces based on second-order model (30.1) that overcome the limitations of 3^k designs. Central composite designs, discussed in the next section, are general purpose designs that are widely used in practice. Optimal response surface designs, discussed in Section 30.3, are designs that meet an optimality criterion specified by the experimenter.

30.2 Central Composite Response Surface Designs

Structure of Central Composite Designs

Central composite designs are two-level full or fractional factorial designs that have been augmented with a small number of carefully chosen treatments to permit estimation of the

FIGURE 30.1 Two Central Composite Designs for Two Factors.



second-order response surface model (30.1). Consider first the 2^2 factorial design pictured in terms of its coded factor levels in Figure 30.1a. If we add a single center point and four star points (also called *axial* points), as shown in Figure 30.1b, the resulting design is a central composite design. A star point is one in which all factors but one are set at their mid-levels. In terms of the coded values, the coordinates of the four star points in Figure 30.1b are $(-1, 0)$, $(1, 0)$, $(0, -1)$, and $(0, 1)$. As shown in Figure 30.1b, the four star points are located at the centers of each of the four edges of the experimental region. Notice that the central composite design in Figure 30.1b is in fact a 3^2 factorial design, where both factors are at three levels and all factor level combinations are included.

The distance from a star point to the center point in coded units is typically denoted by α. In Figure 30.1b, the star points are one coded unit from the center; hence for this design $\alpha = 1$. It is sometimes possible to place the star points beyond the experimental region defined by the original upper and lower limits of the factors. Figure 30.1c presents a central composite design where the star points are located at a distance $\alpha = \sqrt{2} = 1.414$ from the center. As may be seen from Figure 30.1c, each factor is run at five distinct levels when α is larger than 1.0, whereas use of $\alpha = 1.0$ yields just three distinct levels for each factor, as shown in Figure 30.1b. One advantage of setting α greater than 1.0, therefore, is that tests for cubic and quadratic curvature effects can then be conducted.

To summarize, central composite designs consist of three components:

1. 2^{k-f} *corner points*. At the base of any central composite design is a two-level full factorial design or a fractional factorial design of resolution V or higher. This component provides for the estimation of linear main effects and all two-factor interaction effects. Corner points have coded coordinates of the form $(\pm 1, \pm 1, \ldots, \pm 1)$.

2. $2k$ *star points*. These factor level combinations permit the estimation of all quadratic main effects. In addition, when $\alpha > 1.0$, significance tests for higher-order curvature effects can be conducted. Star points have coordinates $(\pm\alpha, 0, \ldots, 0)$, $(0, \pm\alpha, 0, \ldots, 0)$, etc.

3. n_0 *center points*. If $n_0 > 1$, a pure error estimate of σ^2 is available and a lack of fit test is possible. The coded coordinates of the center point replicates are $(0, 0, \ldots, 0)$.

TABLE 30.1 Three-Factor Central Composite Designs with $n_0 = 4$ Replications at Center Point.

Experimental Trial	Factor Level Settings		
	X_1	X_2	X_3
1	−1	−1	−1
2	1	−1	−1
3	−1	1	−1
4	1	1	−1
5	−1	−1	1
6	1	−1	1
7	−1	1	1
8	1	1	1
9	$-\alpha$	0	0
10	α	0	0
11	0	$-\alpha$	0
12	0	α	0
13	0	0	$-\alpha$
14	0	0	α
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0

Table 30.1 presents the coded factor level settings for central composite designs for three factors, with $n_0 = 4$ replications at the center point.

Commonly Used Central Composite Designs

As we have seen, the term "central composite design" refers to a family of experimental designs. Within that family, numerous designs exist, depending on the choice of the base corner points, α, and the extent of replications. Not only may there be n_0 replications at the center point but there may also be replications at the corner and star points. We shall let n_c and n_s denote, respectively, the number of replications at each corner point and star point. The number of experimental trials at the corner points then is:

$$2^{k-f} n_c \tag{30.3a}$$

where k is the number of factors and f is the level of fractionation in the two-level factorial design selected. Similarly, the number of replications at the star points is:

$$2kn_s \tag{30.3b}$$

Thus, the total number of experimental trials planned, denoted by n_T as usual, is:

$$n_T = 2^{k-f} n_c + 2kn_s + n_0 \tag{30.3c}$$

The characteristics of any particular central composite design therefore depend on the choices of k, f, α, n_0, n_s, and n_c.

A list of widely used central composite designs is given in Table 30.2 for studies involving two to eight factors. (The meaning of the term "rotatability" in Table 30.2 will be explained shortly.) The base fractional factorial designs for five to eight factors are the smallest such

TABLE 30.2 Some Useful Central Composite Designs.

	Number of Factors						
Design Characteristic	2	3	4	5	6	7	8
Base factorial design	2^2	2^3	2^4	2_{V}^{5-1}	2_{V}^{6-1}	2_{V}^{7-1}	2_{V}^{8-2}
Star points	4	6	8	10	12	14	16
Center point	1	1	1	1	1	1	1
α for rotatability ($n_c = n_s = 1$)	1.4142	1.6818	2.0000	2.0000	2.3784	2.8284	3.3636
Total number of trials ($n_c = n_s = 1, n_0 = 4$)	12	18	28	30	48	82	84

designs that will provide resolution $R = \text{V}$. Table 30.2 also shows the total number of experimental trials required when a single replication at the corner and star points of the design (i.e., $n_c = n_s = 1$) and $n_0 = 4$ replications at the center point are sufficient. When the error variance σ^2 is large relative to the factor effects, larger numbers of replications at each treatment will be needed.

Rotatable Central Composite Designs

When choosing a particular central composite design, a criterion that is often considered is that of rotatability. The rotatability criterion is concerned with the precision of the estimator $\hat{Y}_h$ since a main purpose of response surface designs is to estimate the response surface, i.e., to estimate the mean response $E\{Y_h\}$ in (30.1) at different locations $\mathbf{X}_h$, the vector of the given levels of the k factors. Rotatable designs have the property that the variance of the fitted value at $\mathbf{X}_h$, $\sigma^2\{\hat{Y}_h\}$, is the same for any point $\mathbf{X}_h$ that is a given distance from the center point, regardless of the direction. The property of equal precision at any given distance from the center point is desirable because it is not usually known in advance which direction from the center point will be of later interest. A rotatable design provides assurance that the precision of the fitted values is not affected by the direction, only by the distance from the center point.

We can examine whether a central composite design is rotatable by considering the variance of $\hat{Y}_h$ as a function of $\mathbf{X}_h$. The variance was given in (6.57):

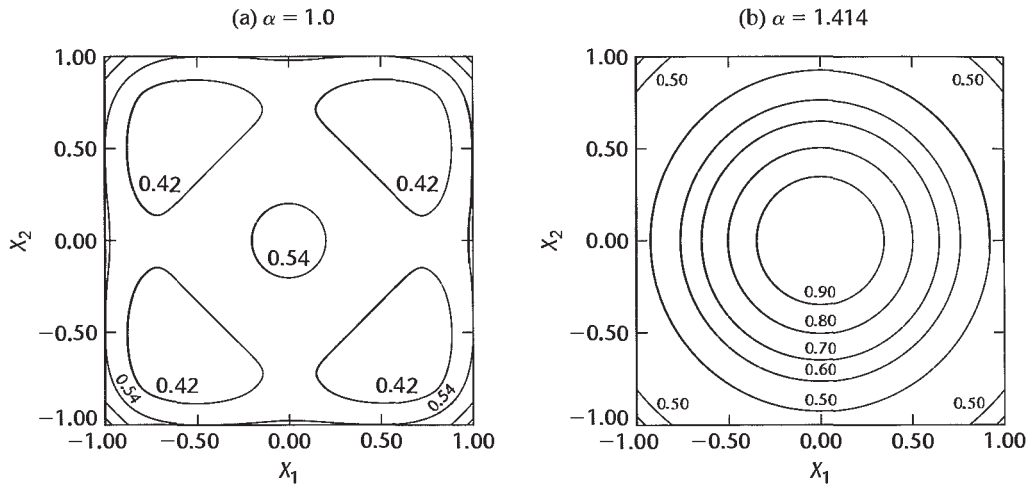
$$\sigma^2\{\hat{Y}_h\} = \sigma^2\mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h = \sigma^2 V_h \tag{30.4}$$

where:

$$V_h = \mathbf{X}_h'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_h \tag{30.4a}$$

V_h is sometimes called the *variance function*. Note that V_h is a function solely of the coded values of the factor levels for the treatments in the design and of the point $\mathbf{X}_h$ where the mean response is to be estimated. Also note that the variance of $\hat{Y}_h$ is a constant multiple of V_h, the constant being the error variance σ^2. Hence, the variance function provides complete information of how the variance $\sigma^2\{\hat{Y}_h\}$ behaves for different points $\mathbf{X}_h$. Figure 30.2 presents contour plots of the variance functions for the two central composite designs in Figure 30.1. For both of these designs, $n_c = n_s = n_0 = 1$, and both use a 2^2 factorial design as the

FIGURE 30.2 Contours of Variance Functions for Two-Factor Central Composite Designs.



base design. They differ only with respect to α. Notice in Figure 30.2b that the contours of the variance function for the central composite design with $\alpha = \sqrt{2}$ are circular, indicating equal precision at a given distance from the center point. Hence, this is a rotatable design. On the other hand, the contours of the variance function in Figure 30.2a are not circular, indicating that the design with $\alpha = 1$ is not rotatable.

It can be shown that a central composite design is rotatable if:

$$\alpha = \left[\frac{2^{k-f}(n_c)}{n_s}\right]^{1/4} \tag{30.5}$$

For the example in Figure 30.2b, we have $n_c = n_s = 1$, $k = 2$, and $f = 0$. Hence, the choice of:

$$\alpha = \left[\frac{2^{2-0}(1)}{1}\right]^{1/4} = \sqrt{2}$$

leads to a rotatable design. Values of α that lead to rotatable designs when $n_c = n_s = 1$ are provided in Table 30.2.

While rotatability is a desirable property of a central composite design, it should not be the sole basis for making the choice of α. For example, in many instances, it may be physically difficult or impossible to extend the star points beyond the experimental region defined by the upper and lower limits of each factor. In such cases, α must not exceed 1.0. Also, a design with $\alpha = 1$ is sometimes easy to implement because only three levels are involved for each factor. In these cases, the resulting lack of rotatability may not be considered a serious disadvantage.

Example

The levels of four ingredients of a prototype solid chocolate bar developed by food scientists at Fisher Company were to be fine-tuned prior to national distribution. The factors and

associated ranges were as follows:

Factor	Low Level	High Level
Cocoa butter	8.0	10.0
Added milk solids	2.0	3.0
Flavoring	2.5	3.5
Sugar	12.5	18.5

The response of interest was the overall consumer acceptability as measured on a 10-point scale. The objective of the experiment was to determine the levels of cocoa butter, added milk solids, flavoring, and sugar that lead to highest acceptability. To carry out the experiment, chocolate bars were to be made with different factor level combinations for the ingredients, and each type of chocolate bar was then to be subjected to a small consumer test. The firm's marketing research department determined that each consumer test would cost about $2,500. Because the total cost of the study was not to exceed $75,000, 30 or fewer consumer tests could be performed. From Table 30.2, we see that the total number of trials for a central composite four-factor design with $n_0 = 4$ replications at the center point is 28, and that this design is rotatable when $\alpha = 2$. The selected design in coded units is shown in Table 30.3.

Comments

1. A central composite design with $\alpha = 1$ is often called a *face-centered design*. For $k = 3$ factors, for instance, this design locates the star points at the center of each of the six faces of the base design cube.

2. When it is not possible to extend the star points beyond the factorial region defined by the original ranges of the factors, a *rotatable inscribed central composite design* can often be used. In such an inscribed design, the coded factor level settings are rescaled by the factor $1/\alpha$ so that all coded factor levels fall between -1 and 1. To illustrate the rescaling, we know from Table 30.2 that a two-factor central composite rotatable design requires the choice of $\alpha = 1.414$ when $n_c = n_s = 1$. To obtain an inscribed two-factor, rotatable central composite design, each coded factor level is multiplied by $1/1.414$. The original rotatable design, with $n_0 = n_c = n_s = 1$, and the corresponding inscribed design are shown in Table 30.4. The inscribed design has the appropriate value of α (1.0), and no factor levels are outside the original ranges for each factor. Note that the actual factor levels now need to be rescaled as well. Consequently, the corner points of the design will no longer be at the limits of the ranges for the factor levels. When this is undesirable, an inscribed design will not be appropriate. ■

Other Criteria for Choosing a Central Composite Design

Other criteria for the choice of a central composite response surface design, besides rotatability, have been proposed. Two of these are *orthogonality* and *uniform precision*. An unblocked central composite design is orthogonal if the estimated factor effect coefficients are all uncorrelated. A proper choice of n_0, the number of center point replicates, will lead to an orthogonal central composite design. For example, some orthogonal central composite designs for two to five factors are as follows for $n_s = n_c = 1$ replicate at each star and

TABLE 30.3 Three-Factor Central Composite Design with $\alpha = 2.0$ —Fisher Company Example.

Experimental Trial	Factor Level Settings			
	X_1	X_2	X_3	X_4
1	−1	−1	−1	−1
2	1	−1	−1	−1
3	−1	1	−1	−1
4	1	1	−1	−1
5	−1	−1	1	−1
6	1	−1	1	−1
7	−1	1	1	−1
8	1	1	1	−1
9	−1	−1	−1	1
10	1	−1	−1	1
11	−1	1	−1	1
12	1	1	−1	1
13	−1	−1	1	1
14	1	−1	1	1
15	−1	1	1	1
16	1	1	1	1
17	−2.0	0	0	0
18	2.0	0	0	0
19	0	−2.0	0	0
20	0	2.0	0	0
21	0	0	−2.0	0
22	0	0	2.0	0
23	0	0	0	−2.0
24	0	0	0	2.0
25	0	0	0	0
26	0	0	0	0
27	0	0	0	0
28	0	0	0	0

TABLE 30.4 Two-Factor Inscribed Central Composite Design with $n_0 = n_c = n_s = 1, \alpha = 1.414$.

Experimental Trial	Central Composite Design		Inscribed Central Composite Design	
	X_1	X_2	X_1	X_2
1	−1	−1	−.707	−.707
2	1	−1	.707	−.707
3	−1	1	−.707	.707
4	1	1	.707	.707
5	−1.414	0	−1	0
6	1.414	0	1	0
7	0	−1.414	0	−1
8	0	1.414	0	1
9	0	0	0	0

corner point:

	Number of Factors			
Design Characteristic	2	3	4	5
Base factorial design	2^2	2^3	2^4	2^5
n_0	8	9	12	17

Notice that for each of these designs, the required number of replications at the center point is quite large. While orthogonality is desirable because it simplifies the analysis of the results, at times it will be difficult to justify large expenditures for replications at the center point. Lack of orthogonality is not a serious disadvantage in practice today because the analysis of the experimental results is easily handled by using a computer regression package.

A uniform precision central composite design is a rotatable design for which the precision of the estimated mean response is the same at the center point as it is one unit from the center point (in any direction). Uniform precision designs are obtained by appropriate choices of α and n_0. The following are the required values of α and n_0 for studies with two to five factors:

	Number of Factors			
Design Characteristic	2	3	4	5
Base factorial design	2^2	2^3	2^4	2^5
α for rotatability	1.414	1.682	2.000	2.378
n_0	5	6	7	10

Like for the orthogonal designs above, the number of center point replications required for uniform precision may be too large. Uniform precision is therefore often used only as a secondary criterion for determining the number of replications at the center point.

Blocking Central Composite Designs

One useful characteristic of central composite response surface designs is that they can be blocked easily. The corner points of the central composite design, which constitute a 2^{k-f} factorial design, can be blocked by the methods described in Section 29.5. As noted there, one or more center point replications can be allocated to each of these blocks. Any remaining center point replications and all star points will constitute a final, separate block. Thus, if the base 2^{k-f} factorial design is run in b blocks, the central composite design is run in $b+1$ blocks. The resulting blocking arrangement is then as follows:

$$\begin{array}{ll} \text{Blocks 1 to } b\text{:} & 2^{k-f} \text{ base factorial design in } b \text{ blocks, with } n_0^* \\ & \text{center point replications in each block} \\ \text{Block } b+1\text{:} & 2k \text{ star points, with } n_0 - bn_0^* \text{ center point} \\ & \text{replications added} \end{array} \tag{30.6}$$

Augmenting Two-Level Studies. The blocking arrangement just described can also be used to facilitate the implementation of a central composite design in two stages, which is often desirable. In the first stage, a two-level study with some center point replications is conducted in one or more blocks. If the test for lack of fit suggests the presence of curvature, or if a better approximation of the response surface is desired, the initial two-level study is augmented with star points and additional center point replications. These additional experimental trials constitute an additional block.

Comment

Blocking arrangement (30.6) ensures that estimated block effects will be uncorrelated with estimated linear main effects and two-factor interactions, but the estimated block effects may be correlated with the estimated quadratic main effects. A central composite design that is *orthogonally blocked* will also provide that the estimated block effects are uncorrelated with the estimated quadratic main effects. It is not always possible to achieve both rotatability and orthogonal blocking. Often, however, orthogonal blocking and approximate rotatability can be achieved by suitable choices of the locations of the star points and by the allocation of the center point replications to the blocks. Reference 30.1 provides further information on orthogonally blocked central composite designs. ■

Additional General-Purpose Response Surface Designs

While central composite designs are the most widely used general-purpose response surface designs, other general-purpose designs are available. One important class of alternative designs is the Box-Behnken family of designs. Box-Behnken designs differ from central composite designs in two ways. First, only three levels for each factor are employed. Second, Box-Behnken designs have no corner points. Box-Behnken designs are sometimes preferred to central composite designs when physical or economic constraints prevent the use of the corner points—where all factor levels are at an extreme. A listing of Box-Behnken designs and their blocking arrangements is provided in Reference 30.2.

30.3 Optimal Response Surface Designs

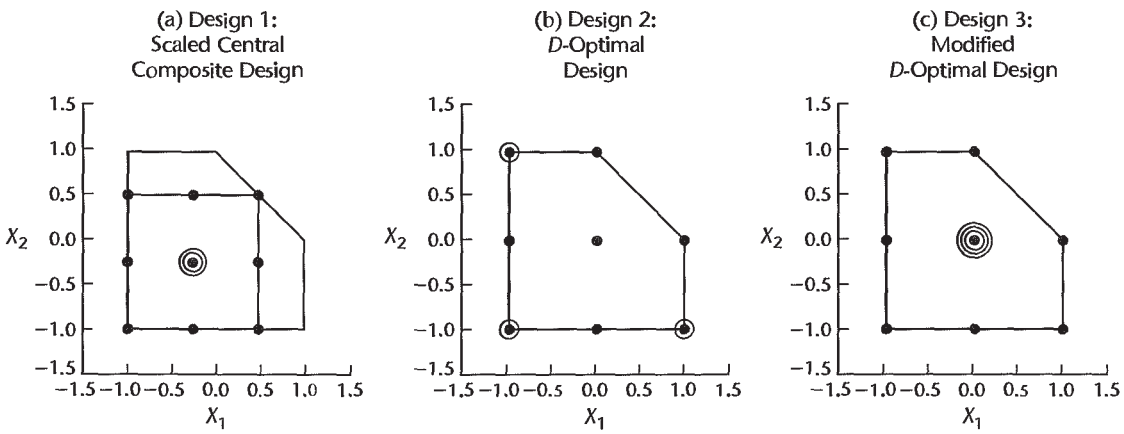
Purpose of Optimal Designs

Central composite response surface designs have been developed for fairly standard experimental situations where the response surface of interest can be reasonably approximated by the second-order polynomial response function (30.1) and the experimental region is defined by the upper and lower limits of the factor levels. Also, since central composite designs are general purpose designs, they are not oriented to provide either optimum precision of the regression parameters or optimum precision for estimating mean responses for particular circumstances.

Optimal designs are useful when optimization of the precision is of key importance and/or when nonstandard experimental situations are encountered. We consider now three main types of nonstandard experimental conditions where central composite designs may not be feasible—irregular experimental regions, nonstandard models, and nonstandard sample sizes.

Irregular Experimental Regions. Irregular experimental regions are quite common in industrial studies. One simple example, described in Reference 30.3, involved the

FIGURE 30.3 Operating Region and Three Alternative Designs with $n_T = 11$—Rutgers Experimental Station Example.



application of two fertilizers at the Rutgers Experimental Station to determine the levels of the fertilizers that would optimize the yield of a particular crop. It was known in advance of the experiment that a toxic level of the chemicals would result if both of the fertilizers were applied simultaneously at their high levels. The investigators determined that the sum of the two fertilizers (in coded units) should not exceed 1.0:

$$X_1 + X_2 \leq 1.0 \tag{30.7}$$

This constraint leads to the irregular experimental region shown in Figure 30.3a. Also shown in Figure 30.3a is a face-centered central composite design with three replications at the center point. Notice that the ranges of the two factors must be considerably reduced to accommodate the standard central composite design here. Figure 30.3 also contains two other designs for this experimental study that we shall discuss shortly.

Nonstandard Models. Nonstandard models can arise for a variety of reasons. For example, the investigator may know that the response function for a two-factor study is approximately linear in X_1 for constant X_2 and approximately quadratic in X_2. An appropriate regression function then would be:

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \beta_{22} X_2^2$$

Nonstandard models also arise in response surface experiments when both qualitative and quantitative factors are present. In the above example, if the first factor were a qualitative factor with two levels, a response function of the following form would be appropriate:

$$E\{Y\} = \beta_0 + \beta_1 I_1 + \beta_2 X_2 + \beta_{12} I_1 X_2 + \beta_{22} X_2^2$$

where:

$$I_1 = \begin{cases} -1 & \text{if factor 1 at level 1} \\ 1 & \text{if factor 1 at level 2} \end{cases}$$

Nonstandard Sample Sizes. In the chocolate bar optimization study of Section 30.2, budgetary considerations required that the number of runs in the experiment not exceed 30. From Table 30.2, we found that a four-factor central composite design with four replications at the center point was feasible since it would require $n_T = 28$ experimental trials. Suppose now that the budget for the experiment were only \$50,000. At \$2,500 per market test, the maximum number of trials now would be 20, and the selected central composite design would no longer be feasible, even with no replications at the center point.

It is possible, nonetheless, to construct experimental designs that will provide estimates of all of the parameters in the full second-order response function (30.1) in fewer than 20 runs since there are only 15 parameters in this model when $k = 4$. Optimal design techniques can be used here to construct a potentially useful second-order design for any feasible experimental size between 15 and 20 trials.

Optimal Design Approach

In order to construct an optimal experimental design, the investigator must first specify the following:

1. The number of experimental trials, n_T.
2. The response function of interest.
3. A *candidate list, C*, of feasible treatments.
4. A statistical *design criterion* for the selection of the treatments from the candidate list C and for the allocation of the n_T trials to the selected treatments.

Once these specifications have been made, numerical computer search procedures are usually employed to find the experimental design that meets optimally the design criterion.

Example

To illustrate the optimal design approach, consider again the Rutgers Experimental Station example. The feasible experimental region is shown in Figure 30.4a. Suppose that no more than $n_T = 11$ experimental trials can be made, and that the response function is the

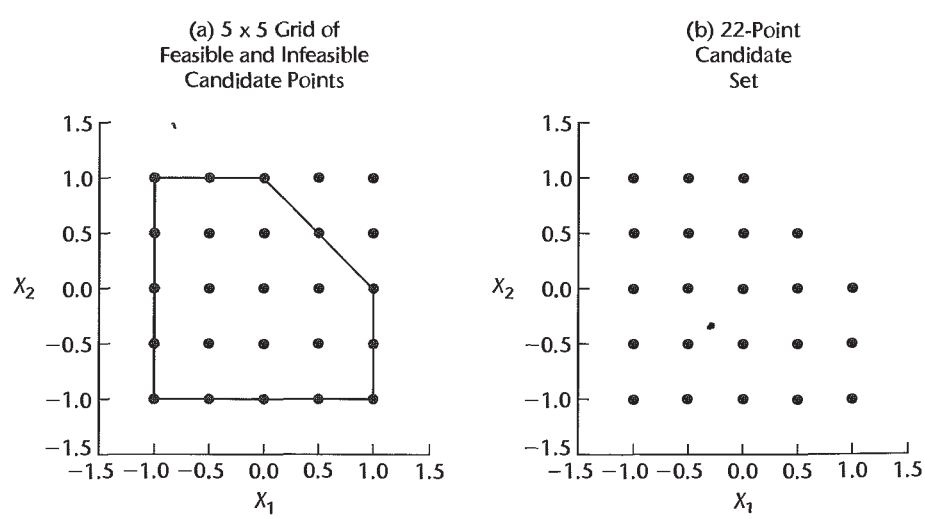


FIGURE 30.4 Candidate Set of Treatments—Rutgers Experimental Station Example.

second-order one in (30.1):

$$E\{Y\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 \tag{30.8}$$

To obtain an optimal design, it is still necessary to specify a candidate list of treatments and a criterion for design selection. Often, the candidate list of treatments is obtained from a grid of regularly spaced points in the feasible experimental region. Figure 30.4a shows a 5×5 grid of treatment points over the unconstrained region. Of the 25 grid points, three fall in the infeasible region because the sum $X_1 + X_2$ for these points exceeds the constraint in (30.7). These three infeasible grid points therefore need to be deleted, resulting in the 22-point candidate set shown in Figure 30.4b.

Finally, a statistical criterion for the selection of the experimental design with n_T trials must be provided. We shall now discuss two such criteria that are widely employed.

Design Criteria for Optimal Design Selection

***D* Criterion.** When precise estimation of model parameters is of primary interest, the D (determinant) criterion provides a useful measure of the precision of an experiment. This criterion is based on the joint confidence region for the parameters in the normal error regression model. This joint confidence region is given by the set of coefficient vectors $\boldsymbol{\beta}$ that satisfy the inequality:

$$\frac{(\mathbf{b} - \boldsymbol{\beta})'\mathbf{X}'\mathbf{X}(\mathbf{b} - \boldsymbol{\beta})}{pMSE} \leq F(1 - \alpha; p, n - p) \tag{30.9}$$

For simple linear regression, where the unknown parameters are β_0 and β_1, the boundary of this region is an ellipse. For models with three or more parameters, the boundary of the confidence region is an ellipsoid. One measure of the precision of the parameter estimates is the area or volume (for three or more parameters) of the confidence region. A small confidence region area or volume implies high precision. When the objective of the experiment is to estimate the vector $\boldsymbol{\beta}$ precisely, the confidence ellipse or ellipsoid for $\boldsymbol{\beta}$ should therefore be small. It can be shown that minimizing the volume of the confidence region (30.9) is equivalent to minimizing:

$$D = |(\mathbf{X}'\mathbf{X})^{-1}| \tag{30.10}$$

where $|(\mathbf{X}'\mathbf{X})^{-1}|$ denotes the determinant of $(\mathbf{X}'\mathbf{X})^{-1}$. Hence, the smaller is the determinant $|(\mathbf{X}'\mathbf{X})^{-1}|$, the smaller is the volume of the confidence region. A design that minimizes $|(\mathbf{X}'\mathbf{X})^{-1}|$ is said to be *D-optimal.*

Example

We illustrate the use of the D criterion for the Rutgers Experimental Station example by considering the three experimental designs in Figure 30.3. The design in Figure 30.3a, as we noted earlier, is a scaled central composite design with three replications at the center point, requiring $n_T = 11$ trials. The designs in Figures 30.3b and 30.3c also require $n_T = 11$ trials but involve a different set of treatments than the scaled central composite design. The designs in Figures 30.3b and 30.3c utilize the same set of treatments but differ as to which treatments receive more than one replication. Calculation of the determinant $|(\mathbf{X}'\mathbf{X})^{-1}|$ will be done ordinarily by use of a computer or a programmable calculator. We find that the

values of the determinant criterion for the three designs under consideration are:

$$\text{Design 1:} \quad D = |(\mathbf{X}'\mathbf{X})^{-1}| = .009117$$

$$\text{Design 2:} \quad D = |(\mathbf{X}'\mathbf{X})^{-1}| = .000161$$

$$\text{Design 3:} \quad D = |(\mathbf{X}'\mathbf{X})^{-1}| = .000347$$

Since design 2 yields the smallest value of D among the three proposed designs, design 2 is preferred to designs 1 and 3 on the basis of the determinant criterion.

Relative Efficiency of Two Designs. A measure of the relative efficiency of design 1 relative to design 2 according to the D criterion is the following, where $\mathbf{X}_1$ and $\mathbf{X}_2$ are the $\mathbf{X}$ matrices for the two designs:

$$E_D = \left(\frac{|(\mathbf{X}_2'\mathbf{X}_2)^{-1}|}{|(\mathbf{X}_1'\mathbf{X}_1)^{-1}|}\right)^{1/p} \tag{30.11}$$

For the Rutgers Experimental Station example, the relative efficiency of design 1 compared to design 2 is:

$$E_D = \left(\frac{.000161}{.009117}\right)^{1/6} = .51$$

The relative efficiency measure states that design 1 is only 51 percent as efficient as design 2. This means that design 1 would need to be replicated $1/.51 = 1.96$ times in order to achieve as small a confidence region for the regression parameters as design 2.

***V* Criterion.** The objective of response surface experiments often is the estimation of the mean response $E\{Y_h\}$ at different combinations of factor level settings, denoted by $\mathbf{X}_h$. The estimation of these mean responses often is used to identify the factor settings $\mathbf{X}_h$ for which the mean response $E\{Y_h\}$ is either maximized or minimized. The V criterion considers the variances $\sigma^2\{\hat{Y}_h\}$ at factor level combinations $\mathbf{X}_h$ of interest and employs the average of these variances as the criterion. Let P denote the set of n_P factor level combinations ($\mathbf{X}_h$ vectors) at which the experimenter wishes to estimate the mean response. Often, the estimation set P is the same as the candidate set C. At other times, the two sets do not coincide, as when P contains points outside of the experimental region because the investigator anticipates the need for estimating mean responses in a region where experimentation is costly. Using (30.4) to express the variance $\sigma^2\{\hat{Y}_h\}$ in terms of the variance function V_h, we can state the average of the variances of $\hat{Y}_h$ for the estimation set P as follows:

$$\frac{\sum \sigma^2\{\hat{Y}_h\}}{n_P} = \frac{\sum \sigma^2 V_h}{n_P} = \sigma^2 \overline{V} \tag{30.12}$$

where:

$$\overline{V} = \frac{\sum V_h}{n_P} \tag{30.12a}$$

A design that minimizes $\overline{V}$ in (30.12a) is called a *V-optimal design.*

Example

In the Rutgers Experimental Station example, the estimation set P is to consist of the 22 candidate treatments in Figure 30.4b. The variance function V_h was evaluated first for

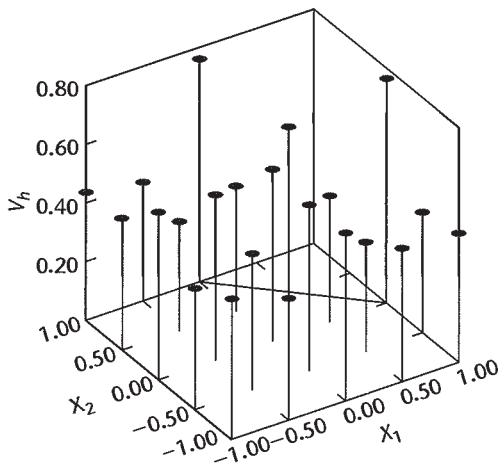


FIGURE 30.5 Design 2 Variance Function V_h Evaluated at Points in Estimation Set—Rutgers Experimental Station Example.

design 2 for the 22 treatments in the estimation set. The results are shown in Figure 30.5. Note that the treatments at (1, 0) and (0, 1), the two vertices of the operating region with no replications, have large V_h values. Consequently, with design 2 the mean responses for these two treatments will not be estimated as precisely as for the other treatments. The mean of the 22 V_h values for design 2 is $\overline{V} = .500$. In the same fashion, we find $\overline{V}$ for the other two designs. The comparative results for the three designs are:

Design	$\overline{V}$
1	1.192
2	.500
3	.486

Hence, according to the V criterion design 3 is slightly preferred over design 2, and both of these designs are substantially better than design 1.

Relative Efficiency of Two Designs. A measure of the relative efficiency of design 1 relative to design 2 according to the V criterion is the following, where $\overline{V}_1$ and $\overline{V}_2$ denote the averages of the V_h values for the two designs:

$$E_V = \frac{\overline{V}_2}{\overline{V}_1} \tag{30.13}$$

For the Rutgers Experimental Station example, the relative efficiency of design 1 relative to design 3 is:

$$E_V = \frac{.486}{1.192} = .408$$

Design 1 is only 41 percent as efficient as design 3 according to the V criterion, implying that it would require $1/.408 = 2.45$ replications of design 1 to achieve the same average precision as with design 3.

Comment

Other criteria that have been proposed for identifying a design as optimal involve minimizing the average variance of the estimated regression coefficients (A-optimality) and minimizing the maximum variance of $\hat{Y}_h$ over the estimation set (G-optimality when the estimation set P is the same as the candidate set C). ■

Construction of Optimal Response Surface Designs

On occasion, the optimal design for a given criterion is known or can be found analytically. Usually, however, a computer search is required to find the optimal design. Many statistical software packages provide capabilities for finding optimal designs. To reduce the amount of computing required, these packages do not evaluate all possible designs. Instead, fast, special-purpose computer search procedures, called *exchange algorithms,* are used to find designs that are either optimal or nearly optimal. These algorithms begin the search with a starting design, sometimes randomly chosen. They then alternately add new points to the design and subtract points from the design in ways that lead to improvements in the design criterion. Since these algorithms do not evaluate every possible design, they cannot guarantee that an optimal design has been found. To increase the likelihood that a best or near-best design is found, some software packages provide capabilities for repeated attempts, beginning the search from different, randomly selected starting designs. A discussion of these search procedures is given in Reference 30.4.

Example

IC Technologies is a manufacturer of dashboard displays used in the automotive industry. An important component of the manufacturing process involves the bonding of a computer chip to a glass surface with adhesive. Management wished to determine which of two types of adhesive, provided by two different suppliers, was superior. Identification of the optimum processing temperature was also of interest. The response of interest was bonding strength—the amount of force required to break the chip free of the surface. The factors and associated levels were as follows:

Factors	Levels		
Adhesive	Type 1	Type 2	
Process temperature	210	240	270

Notice that adhesive is a qualitative factor that can assume only two levels. Process temperature is a quantitative factor that has a range from 210 to 270. The process engineers wished to limit the number of temperature levels to the limits of the range (210, 270) and to the middle (240). The candidate set of treatments is therefore given by the six factor-level combinations shown in Figure 30.6a.

Since adhesive type is a qualitative factor with two levels and a quadratic (second-order) temperature effect was expected, the response function chosen was the following:

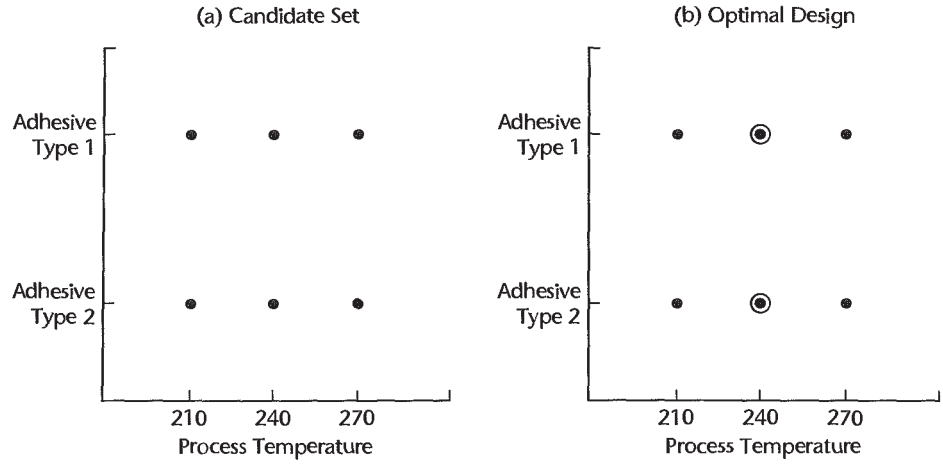
$$E\{Y\} = \beta_0 + \beta_1 I_1 + \beta_2 X_2 + \beta_{22} X_2^2 + \beta_{12} I_1 X_2$$

where:

$$I_1 = \begin{cases} -1 & \text{if adhesive is type 1} \\ 1 & \text{if adhesive is type 2} \end{cases}$$

X_2 = coded temperature

FIGURE 30.6 SAS Optimal Design Construction—IC Technologies Example.



Management determined that at most eight experimental trials could be handled and specified that the V criterion be employed. The estimation set of interest consisted of 21 equally spaced points spanning the process temperature range for each of the two types of adhesive. Note that the 42-point estimation set P here is not the same as the candidate set C. The JMP Custom Design option was used to obtain the V-optimal design for $n_T = 8$ shown in Figure 30.6b. Notice that the medium level of temperature (240) is replicated twice for each adhesive type. Thus, two degrees of freedom will be available for a pure error estimate of the error variance and a lack of fit test will be possible.

Some Final Cautions

Caution in using optimal designs is important because these designs are best for particular choices of sample size, design space, response function, and design criterion. For example, designs that are optimal according to one criterion may be far from optimal according to another criterion. Also, optimal designs are highly sensitive to the choice of response function. A design that is optimal for a second-order response function is generally not optimal if a first-order response function is the true function. Consequently, the experimenter needs to consider whether the optimal design will be far from being optimal if the assumed response function is incorrect, and whether the optimal design will provide sufficient information about the true response function if the assumed one is incorrect.

Another reason for caution in choosing optimal designs is that they are constructed on the basis of a single design criterion. Frequently, an experimenter has a number of potentially conflicting objectives. It is therefore important that any candidate design be evaluated for its ability to satisfy each of these goals. Small modifications to computer-generated designs—such as the addition of replications at the center point—can be useful for increasing the overall utility of a design even if it is then no longer an optimal design according to a given criterion. It is often useful to construct optimal designs for a range of sample sizes and a variety of response functions and criteria. A final design can then be chosen on the basis of its ability to reasonably meet the different objectives over the range of response functions and criteria.

A thorough discussion of optimal designs is presented in Reference 30.5.

30.4 Analysis of Response Surface Experiments

The analysis of second-order response surface designs frequently involves three phases:

1. Estimation of response function
2. Model interpretation and visualization
3. Identification of optimum operating conditions

In phase 1, standard regression tools are used to estimate the response function and obtain a good regression fit. The fitted surface is then explored graphically in phase 2. Finally, in phase 3, factor level combinations that lead to an optimum response are identified. Fitting of polynomial regression models was already discussed in Chapter 8. Here, we shall focus on the visualization of the fitted model and the identification of optimum operating conditions.

Model Interpretation and Visualization

Three-dimensional plots of the response surface, contour plots, and conditional effects plots are the primary visual tools for interpreting and communicating the results of response surface experiments. Generally, three kinds of fitted surfaces arise in practice.

1. A mound-shaped surface, which is characterized by contours that are ellipses or circles. Figure 30.7 presents a three-dimensional response surface plot and a contour plot of the fitted response function:

$$\hat{Y} = 65 + 3X_1 + 4X_2 - 10X_1^2 - 15X_2^2 + 15X_1X_2$$

The contour plot in Figure 30.7b shows that the estimated mean response increases from a minimum of $\hat{Y} = 34$ in the lower right corner ($X_1 = 1$, $X_2 = -1$) to a maximum in the center of the region bounded by the $\hat{Y} = 66$ contour.

FIGURE 30.7 Two-Factor Response Surface and Contour Plot—Mound-Shaped Surface.

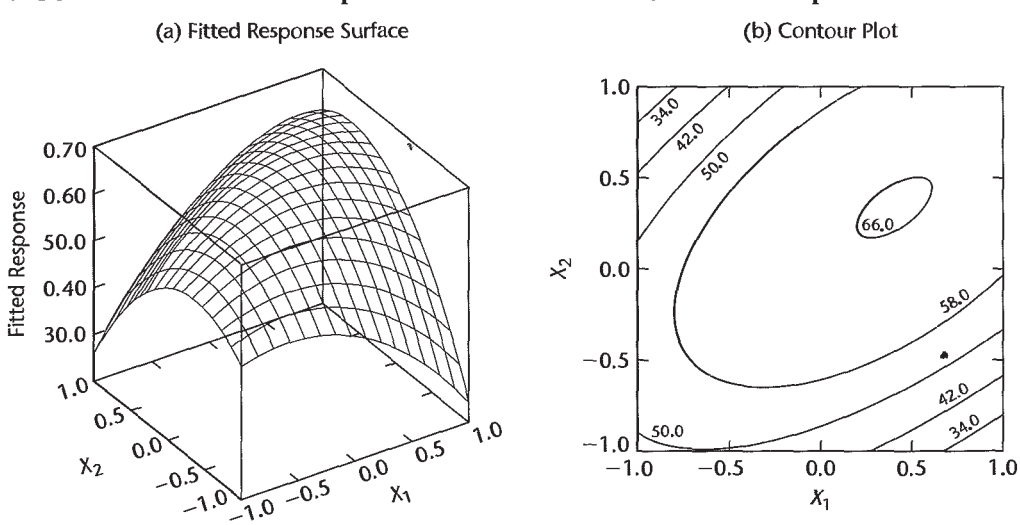
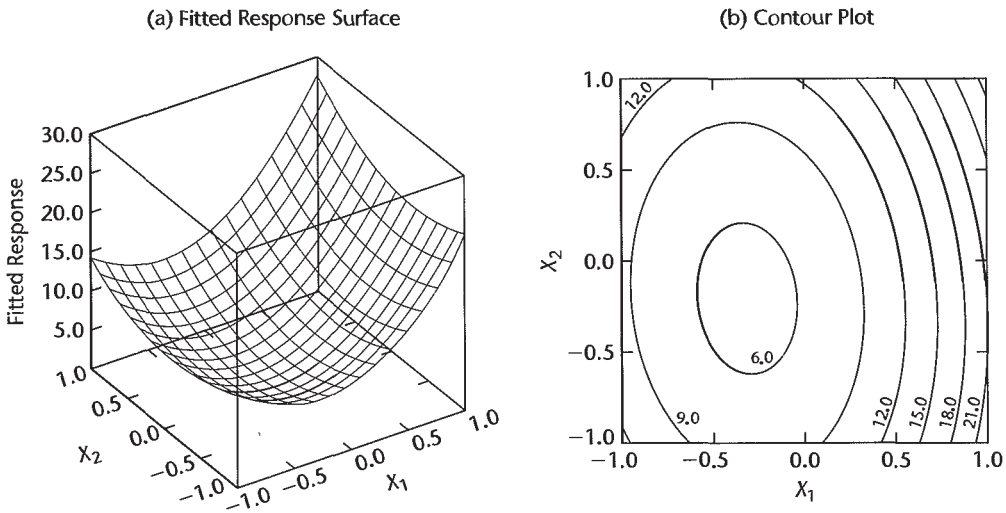


FIGURE 30.8 Two-Factor Response Surface and Contour Plot—Bowl-Shaped Surface.



2. A bowl-shaped surface, which also has elliptical or circular contours; however, the response function decreases in the direction of the smallest ellipse. Figure 30.8 presents the response surface and a contour plot of the fitted response function:

$$\hat{Y} = 6.5 + 6X_1 + 2X_2 + 9X_1^2 + 4X_2^2 + X_1X_2$$

From the contour plot in Figure 30.8b, we see that the surface decreases from a maximum in the upper right corner ($X_1 = 1, X_2 = 1$) to a minimum in the center of the region bounded by the $\hat{Y} = 6$ contour.

3. A response surface with a *saddle* or a *minimax*. Figure 30.9 presents the response surface and a contour plot of the fitted response function:

$$\hat{Y} = 65 + 3X_1 + 4X_2 - 10X_1^2 - 15X_2^2 + 35X_1X_2$$

From the contour plot in Figure 30.9b, notice that the mean response increases from the upper left corner to a maximum in the center of the region and then decreases as we approach the lower right corner. The opposite occurs when moving from the upper right corner to the lower left corner.

Conditional effects plots, or interaction plots, can also provide useful insights. Figure 30.10 presents a conditional effects plot for the saddle-shaped surface in Figure 30.9 at $X_2 = -1, 0, 1$:

$$X_2 = -1: \quad \hat{Y} = 46 - 32X_1 - 10X_1^2$$

$$X_2 = \ \ 0: \quad \hat{Y} = 65 + 3X_1 - 10X_1^2$$

$$X_2 = \ \ 1: \quad \hat{Y} = 54 + 38X_1 - 10X_1^2$$

FIGURE 30.9 Two-Factor Response Surface and Contour Plot—Saddle-Shaped Surface.

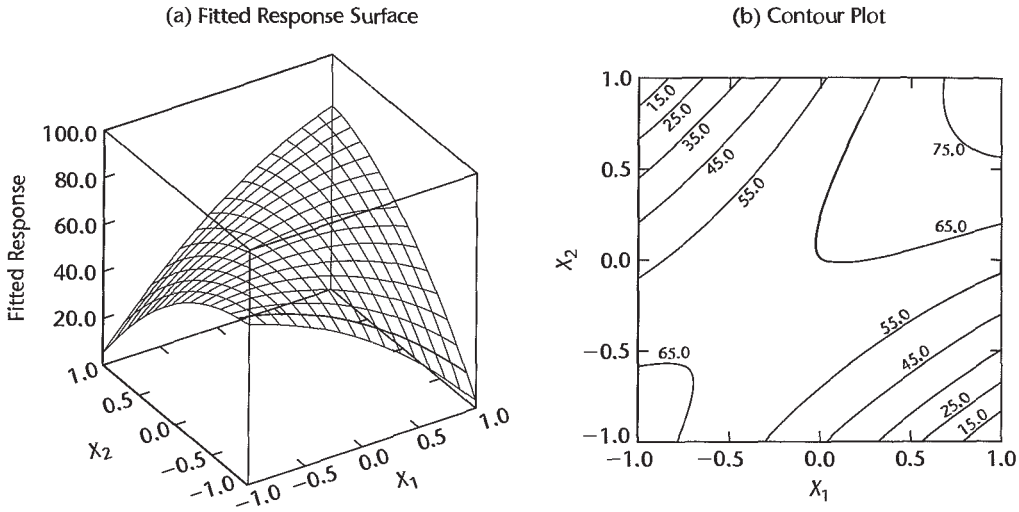
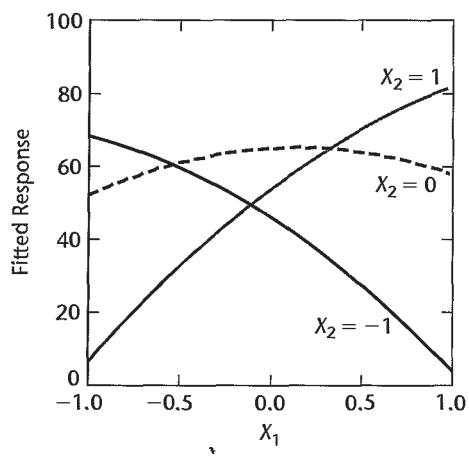


FIGURE 30.10 Conditional Effects Plots for Saddle-Shaped Surface in Figure 30.9.



Notice that at low X_2 the mean response is decreasing in X_1, whereas at high X_2 the mean response is increasing in X_1. Thus, the presence of interaction effects is clearly indicated by the plot. Absence of interaction effects would be indicated, as usual, by parallel curves.

Response Surface Optimum Conditions

Response surfaces are frequently fitted for the purpose of finding the combination of factor levels that leads to an optimum response. Usually, either a maximum response (e.g., maximum yield) or a minimum response (e.g., minimum waste) is sought. Mound-shaped response surfaces, such as in Figure 30.7, have a unique maximum, while bowl-shaped response surfaces, such as in Figure 30.8, have a unique minimum. Occasionally, more

complex response surfaces are encountered that have saddle points, such as in Figure 30.9, or a number of local maximum or minimum points.

For a second-order fitted response surface, the point where a maximum, a minimum, or a saddle point occurs, denoted by the vector $\mathbf{X}_s$, is:

$$\mathbf{X}_s = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b}^* \tag{30.14}$$

where:

$$\underset{k\times k}{\mathbf{B}} = \begin{bmatrix} b_{11} & b_{12}/2 & \cdots & b_{1k}/2 \\ b_{12}/2 & b_{22} & \cdots & b_{2k}/2 \\ \vdots & \vdots & & \vdots \\ b_{1k}/2 & b_{2k}/2 & \cdots & b_{kk} \end{bmatrix} \qquad \mathbf{b}^* = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} \tag{30.14a}$$

To determine whether the point $\mathbf{X}_s$ corresponds to a maximum, a minimum, or a saddle point, the nature of the response surface must be known. If a contour plotting capability is available and there are just two or three factors, the nature of the surface can usually be determined by examining the contours in the vicinity of $\mathbf{X}_s$. Otherwise, characteristics of the matrix $\mathbf{B}$ called *eigenvalues* can be used to determine whether the point at $\mathbf{X}_s$ is a maximum, a minimum, or a saddle point. Many computer packages for response surface analysis provide these eigenvalues. If the eigenvalues are all positive, the point is a minimum. If the eigenvalues are all negative, the point is a maximum. Finally, if some eigenvalues are positive and some negative, the point is a saddle point.

Example

Consider again the mound-shaped response surface in Figure 30.7:

$$\hat{Y} = 65 + 3X_1 + 4X_2 - 10X_1^2 - 15X_2^2 + 15X_1X_2$$

We know that this surface has a maximum and wish to locate it. We require the matrix $\mathbf{B}$ and the vector $\mathbf{b}^*$. Using (30.14a), we obtain:

$$\mathbf{B} = \begin{bmatrix} -10 & 15/2 \\ 15/2 & -15 \end{bmatrix} \qquad \mathbf{b}^* = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

Using (30.14), we find the point where the response surface is at the maximum:

$$\mathbf{X}_s = -\frac{1}{2}\begin{bmatrix} -10 & 15/2 \\ 15/2 & -15 \end{bmatrix}^{-1}\begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$$= -\frac{1}{2}\begin{bmatrix} -.1600 & -.0800 \\ -.0800 & -.1067 \end{bmatrix}\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} .40 \\ .25 \end{bmatrix}$$

The maximum response on the fitted surface, at $X_1 = .40$ and $X_2 = .25$, is:

$$\hat{Y} = 65 + 3(.40) + 4(.25) - 10(.40)^2 - 15(.25)^2 + 15(.40)(.25) = 66.16$$

Comments

1. When the maximum or minimum point for the response surface falls well outside the operating region, it may not be feasible to operate at this point and the investigator must then search for the factor level combination that optimizes the mean response within the operating region. For problems involving just two or three predictors, this point can usually be pinpointed using contour plots and

conditional effects plots. For problems involving four or more factors, constrained nonlinear programming methods can be used to identify the optimum factor level combination. Many statistical software packages that provide capabilities for the design of experiments include this feature. Alternatively, a grid of points (such as those used to identify candidate points for optimal design construction) can be constructed and the estimated mean response for each gridpoint is then obtained. If the grid is sufficiently dense, the gridpoint that leads to the maximum (minimum) estimated mean response will closely approximate the optimum point.

When it is feasible to operate outside the experimental region and the optimum point falls well outside this region, it is often necessary to extend the experiment because of uncertainty about the shape of the response surface outside the region of experimentation.

2. In most experiments, more than a single response variable is of interest. For example, in food processing experiments, response variables such as taste, texture, aftertaste, mouthfeel, shelf life, and cost are all frequently of interest. As discussed in Section 29.6, another variable of interest in many studies is the variance of the response variable. In the IC Technologies example, for instance, the manufacturer is concerned not only that the mean bonding strength be adequately high but also that the process variability be small so that almost all components will be bonded with sufficient strength. To analyze experiments with multiple responses, a response surface must be fitted to each response variable. Unfortunately, it is rare that a single factor level combination can be found that simultaneously optimizes all fitted response surfaces. In fact, often the conditions that lead to an optimum value of one response variable (such as texture) lead to a poor response for another variable (such as taste). The investigator must then search for conditions that lead to acceptable responses for all response variables. ■

Example

Dorle Exterior Trim manufactures polyurethane bumpers for automobiles and light trucks. During the initial production stages of a new model, blemishes appeared on the surface of the bumpers. These blemishes, resulting from a high degree of surface porosity, were so extensive that none of the bumpers could be shipped. A response surface experiment was quickly conducted to investigate the effects of three key process variables on porosity and to identify the optimum operating levels for the active process variables. The three factors were chemical temperature, mold temperature, and curing time. The operating ranges for these factors were:

Factor	Low Level	High Level
Chemical temperature	405	425
Mold temperature	100	240
Curing time	20	40

A three-factor central composite response surface design with $\alpha = 1$ and $n_0 = 3$ replications at the center point was chosen. Porosity counts were obtained from visual inspections of the surface of the bumpers.

The analyst first obtained an initial fit of the three-factor second-order response function (30.1). Residual analysis did not reveal any departures from the standard regression assumptions. The fit suggested that the third factor, curing time, was unrelated to porosity. All P-values for terms involving curing time were greater than or equal to .600. A test of H_0: $\beta_3 = \beta_{33} = \beta_{13} = \beta_{23} = 0$ by the general linear test statistic (2.70) resulted in the test statistic $F^* = .179$ and the P-value .943. The analyst therefore concluded H_0, that curing time is unrelated to porosity.

FIGURE 30.11 SAS PROC RSREG Regression Output—Dorle Exterior Trim Example.

Regression	Degrees of Freedom	Type I Sum of Squares	R-Square	F-Ratio	Prob > F
Linear	2	5075.300000	0.6894	87.308	0.0000
Quadratic	2	1854.363485	0.2519	31.900	0.0000
Crossproduct	1	112.500000	0.0153	3.871	0.0749
Total Regress	5	7042.163485	0.9566	48.457	0.0000

Residual	Degrees of Freedom	Sum of Squares	Mean Square
Total Error	11	319.718868	29.065352

Parameter	Degrees of Freedom	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEPT	1	16.301887	2.221627	7.338	0.0000
X1	1	−22.300000	1.704856	−13.080	0.0000
X2	1	3.200000	1.704856	1.877	0.0873
X1*X1	1	12.443396	3.097911	4.017	0.0020
X2*X1	1	3.750000	1.906087	1.967	0.0749
X2*X2	1	11.943396	3.097911	3.855	0.0027

	Critical Value	
Factor	Coded	Uncoded
X1	0.938443	0.938443
X2	−0.281292	−0.281292
Predicted value at stationary point		5.388176

	Eigenvectors	
Eigenvalues	X1	X2
14.084989	0.752384	0.658725
10.301803	−0.658725	0.752384

The SAS PROC RSREG output for the fit of the second-order response surface model with only chemical temperature and mold temperature as the explanatory variables is shown in Figure 30.11. The fitted response surface is:

$$\hat{Y} = 16.30 - 22.30X_1 + 3.20X_2 + 12.44X_1^2 + 11.94X_2^2 + 3.75X_1X_2$$

Notice that the P-values for all estimated coefficients are less than .1, and that R^2 is .957. A lack of fit test was conducted with $\alpha = .01$. The results ($F^* = 3.10$; P-value $= .089$) supported the appropriateness of the model fitted.

A response surface plot and a contour plot for the fitted response function are shown in Figure 30.12. The X_1 scale has been reversed in these plots to provide a better view of the response surface. Notice that the surface is bowl-shaped. Since a main objective of the experiment was to find the levels of the process variables that minimize the porosity on the bumper surface, the analyst next determined the optimum levels of X_1 and X_2 by means of (30.14). Substituting into this formula, the analyst obtained:

$$\mathbf{X}_s = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b}^* = -\frac{1}{2}\begin{bmatrix} 12.44 & 1.875 \\ 1.875 & 11.94 \end{bmatrix}^{-1}\begin{bmatrix} -22.30 \\ 3.20 \end{bmatrix} = \begin{bmatrix} .94 \\ -.28 \end{bmatrix}$$

where $f_i(\theta)$ is a known function of the parameter θ and the ε_i are random variables, usually assumed to have expectation $E\{\varepsilon_i\} = 0$.

With the method of least squares, for the given sample observations, the sum of squares:

$$Q = \sum_{i=1}^{n} [Y_i - f_i(\theta)]^2 \tag{A.57}$$

is considered as a function of θ. The least squares estimator of θ is obtained by minimizing Q with respect to θ. In many instances, least squares estimators are unbiased and consistent.

A.6 Inferences about Population Mean—Normal Population

We have a random sample of n observations $Y_1, \ldots, Y_n$ from a normal population with mean μ and standard deviation σ. The sample mean and sample standard deviation are:

$$\bar{Y} = \frac{\sum_i Y_i}{n} \tag{A.58a}$$

$$s = \left[\frac{\sum_i (Y_i - \bar{Y})^2}{n-1}\right]^{1/2} \tag{A.58b}$$

and the estimated standard deviation of the sampling distribution of $\bar{Y}$, denoted by $s\{\bar{Y}\}$, is:

$$s\{\bar{Y}\} = \frac{s}{\sqrt{n}} \tag{A.58c}$$

We then have:

$$\frac{\bar{Y} - \mu}{s\{\bar{Y}\}} \text{ is distributed as } t \text{ with } n-1 \text{ degrees of freedom when the random sample is from a normal population.} \tag{A.59}$$

Interval Estimation

The confidence limits for μ with confidence coefficient $1 - \alpha$ are obtained by means of (A.59):

$$\bar{Y} \pm t(1 - \alpha/2; n-1)s\{\bar{Y}\} \tag{A.60}$$

Example 1 Obtain a 95 percent confidence interval for μ when:

$$n = 10 \qquad \bar{Y} = 20 \qquad s = 4$$

We require:

$$s\{\bar{Y}\} = \frac{4}{\sqrt{10}} = 1.265 \qquad t(.975; 9) = 2.262$$

The 95 percent confidence limits therefore are $20 \pm 2.262(1.265)$ and the 95 percent confidence interval for μ is:

$$17.1 \leq \mu \leq 22.9$$

TABLE A.1 **Decision Rules for Tests Concerning Mean μ of Normal Population.**

Alternatives	Decision Rule
	(a)
$H_0: \mu = \mu_0$	If $\lvert t^* \rvert \leq t(1-\alpha/2; n-1)$, conclude H_0
$H_a: \mu \neq \mu_0$	If $\lvert t^* \rvert > t(1-\alpha/2; n-1)$, conclude H_a
	where: $t^* = \dfrac{\bar{Y} - \mu_0}{s\{\bar{Y}\}}$
	(b)
$H_0: \mu \geq \mu_0$	If $t^* \geq t(\alpha; n-1)$, conclude H_0
$H_a: \mu < \mu_0$	If $t^* < t(\alpha; n-1)$, conclude H_a
	(c)
$H_0: \mu \leq \mu_0$	If $t^* \leq t(1-\alpha; n-1)$, conclude H_0
$H_a: \mu > \mu_0$	If $t^* > t(1-\alpha; n-1)$, conclude H_a

Tests

One-sided and two-sided tests concerning the population mean μ are constructed by means of (A.59), based on the test statistic:

$$t^* = \frac{\bar{Y} - \mu_0}{s\{\bar{Y}\}} \tag{A.61}$$

Table A.1 contains the decision rules for three possible cases, with the risk of making a Type I error controlled at α.

Example 2

Choose between the alternatives:

$$H_0: \mu \leq 20$$
$$H_a: \mu > 20$$

when α is to be controlled at .05 and:

$$n = 15 \qquad \bar{Y} = 24 \qquad s = 6$$

We require:

$$s\{\bar{Y}\} = \frac{6}{\sqrt{15}} = 1.549$$
$$t(.95; 14) = 1.761$$

The decision rule is:

$$\text{If } t^* \leq 1.761, \text{ conclude } H_0$$
$$\text{If } t^* > 1.761, \text{ conclude } H_a$$

Since $t^* = (24 - 20)/1.549 = 2.58 > 1.761$, we conclude H_a.

Example 3 Choose between the alternatives:

$$H_0: \mu = 10$$
$$H_a: \mu \neq 10$$

when α is to be controlled at .02 and:

$$n = 25 \qquad \bar{Y} = 5.7 \qquad s = 8$$

We require:

$$s\{\bar{Y}\} = \frac{8}{\sqrt{25}} = 1.6$$
$$t(.99; 24) = 2.492$$

The decision rule is:

$$\text{If } |t^*| \leq 2.492, \text{ conclude } H_0$$
$$\text{If } |t^*| > 2.492, \text{ conclude } H_a$$

where the symbol $|\quad|$ stands for the absolute value. Since $|t^*| = |(5.7 - 10)/1.6| = |-2.69| = 2.69 > 2.492$, we conclude H_a.

***P*-Value for Sample Outcome.** The P-value for a sample outcome is the probability that the sample outcome could have been more extreme than the observed one when $\mu = \mu_0$. Large P-values support H_0 while small P-values support H_a. A test can be carried out by comparing the P-value with the specified α risk. If the P-value equals or is greater than the specified α, H_0 is concluded. If the P-value is less than α, H_a is concluded.

Example 4 In Example 2, $t^* = 2.58$. The P-value for this sample outcome is the probability $P\{t(14) > 2.58\}$. From Table B.2, we find $t(.985; 14) = 2.415$ and $t(.990; 14) = 2.624$. Hence, the P-value is between .010 and .015. The exact P-value can be found from many statistical calculators or statistical computer packages; it is .0109. Thus, for $\alpha = .05$, H_a is concluded.

Example 5 In Example 3, $t^* = -2.69$. We find from Table B.2 that the one-sided P-value, $P\{t(24) < -2.69\}$, is between .005 and .0075. The exact one-sided P-value is .0064. Because the test is two-sided and the t distribution is symmetrical, the two-sided P-value is twice the one-sided value, or $2(.0064) = .013$. Hence, for $\alpha = .02$, we conclude H_a.

Relation between Tests and Confidence Intervals. There is a direct relation between tests and confidence intervals. For example, the two-sided confidence limits (A.60) can be used for testing:

$$H_0: \mu = \mu_0$$
$$H_a: \mu \neq \mu_0$$

If μ_0 is contained within the $1 - \alpha$ confidence interval, then the two-sided decision rule in Table A.1a, with level of significance α, will lead to conclusion H_0, and vice versa. If μ_0 is not contained within the confidence interval, the decision rule will lead to H_a, and vice versa.

There are similar correspondences between one-sided confidence intervals and one-sided decision rules.

A.7 Comparisons of Two Population Means—Normal Populations

Independent Samples

There are two normal populations, with means μ_1 and μ_2, respectively, and with the same standard deviation σ. The means μ_1 and μ_2 are to be compared on the basis of independent samples for each of the two populations:

$$\text{Sample 1: } Y_1, \ldots, Y_{n_1}$$
$$\text{Sample 2: } Z_1, \ldots, Z_{n_2}$$

Estimators of the two population means are the sample means:

$$\bar{Y} = \frac{\sum_i Y_i}{n_1} \tag{A.62a}$$

$$\bar{Z} = \frac{\sum_i Z_i}{n_2} \tag{A.62b}$$

and an estimator of $\mu_1 - \mu_2$ is $\bar{Y} - \bar{Z}$.

An estimator of the common variance σ^2 is:

$$s^2 = \frac{\sum_i (Y_i - \bar{Y})^2 + \sum_i (Z_i - \bar{Z})^2}{n_1 + n_2 - 2} \tag{A.63}$$

and an estimator of $\sigma^2\{\bar{Y} - \bar{Z}\}$, the variance of the sampling distribution of $\bar{Y} - \bar{Z}$, is:

$$s^2\{\bar{Y} - \bar{Z}\} = s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right) \tag{A.64}$$

We have:

$$\frac{(\bar{Y} - \bar{Z}) - (\mu_1 - \mu_2)}{s\{\bar{Y} - \bar{Z}\}} \text{ is distributed as } t \text{ with } n_1 + n_2 - 2 \text{ degrees of freedom when the two independent samples come from normal populations with the same standard deviation.} \tag{A.65}$$

Interval Estimation. The confidence limits for $\mu_1 - \mu_2$ with confidence coefficient $1 - \alpha$ are obtained by means of (A.65):

$$(\bar{Y} - \bar{Z}) \pm t(1 - \alpha/2; n_1 + n_2 - 2)s\{\bar{Y} - \bar{Z}\} \tag{A.66}$$

Example 6 Obtain a 95 percent confidence interval for $\mu_1 - \mu_2$ when:

$$n_1 = 10 \qquad \bar{Y} = 14 \qquad \sum(Y_i - \bar{Y})^2 = 105$$
$$n_2 = 20 \qquad \bar{Z} = 8 \qquad \sum(Z_i - \bar{Z})^2 = 224$$

TABLE A.2 Decision Rules for Tests Concerning Means μ_1 and μ_2 of Two Normal Populations ($\sigma_1 = \sigma_2 = \sigma$)—Independent Samples.

Alternatives	Decision Rule
	(a)
H_0: $\mu_1 = \mu_2$	If $\lvert t^* \rvert \leq t(1-\alpha/2; n_1+n_2-2)$, conclude H_0
H_a: $\mu_1 \neq \mu_2$	If $\lvert t^* \rvert > t(1-\alpha/2; n_1+n_2-2)$, conclude H_a
	where: $t^* = \dfrac{\bar{Y}-\bar{Z}}{s\{\bar{Y}-\bar{Z}\}}$
	(b)
H_0: $\mu_1 \geq \mu_2$	If $t^* \geq t(\alpha; n_1+n_2-2)$, conclude H_0
H_a: $\mu_1 < \mu_2$	If $t^* < t(\alpha; n_1+n_2-2)$, conclude H_a
	(c)
H_0: $\mu_1 \leq \mu_2$	If $t^* \leq t(1-\alpha; n_1+n_2-2)$, conclude H_0
H_a: $\mu_1 > \mu_2$	If $t^* > t(1-\alpha; n_1+n_2-2)$, conclude H_a

We require:

$$s^2 = \frac{105+224}{10+20-2} = 11.75 \qquad s\{\bar{Y}-\bar{Z}\} = 1.328$$

$$s^2\{\bar{Y}-\bar{Z}\} = 11.75\left(\frac{1}{10}+\frac{1}{20}\right) = 1.7625 \qquad t(.975;28) = 2.048$$

Hence, the 95 percent confidence interval for $\mu_1 - \mu_2$ is:

$$3.3 = (14-8) - 2.048(1.328) \leq \mu_1 - \mu_2 \leq (14-8) + 2.048(1.328) = 8.7$$

Tests. One-sided and two-sided tests concerning $\mu_1 - \mu_2$ are constructed by means of (A.65). Table A.2 contains the decision rules for three possible cases, based on the test statistic:

$$t^* = \frac{\bar{Y}-\bar{Z}}{s\{\bar{Y}-\bar{Z}\}} \tag{A.67}$$

with the risk of making a Type I error controlled at α.

Example 7

Choose between the alternatives:

$$H_0: \mu_1 = \mu_2$$
$$H_a: \mu_1 \neq \mu_2$$

when α is to be controlled at .10 and the data are those of Example 6. We require $t(.95; 28) = 1.701$, so that the decision rule is:

$$\text{If } \lvert t^* \rvert \leq 1.701, \text{ conclude } H_0$$
$$\text{If } \lvert t^* \rvert > 1.701, \text{ conclude } H_a$$

Since $\lvert t^* \rvert = \lvert (14-8)/1.328 \rvert = \lvert 4.52 \rvert = 4.52 > 1.701$, we conclude H_a.

The one-sided P-value here is the probability $P\{t(28) > 4.52\}$. We see from Table B.2 that this P-value is less than .0005; the exact one-sided P-value is .00005. Hence, the two-sided P-value is .0001. For $\alpha = .10$, the appropriate conclusion therefore is H_a.

Paired Observations

When the observations in the two samples are paired (e.g., attitude scores Y_i and Z_i for the ith sample employee before and after a year's experience on the job), we use the differences:

$$W_i = Y_i - Z_i \qquad i = 1, \ldots, n \tag{A.68}$$

in the fashion of a sample from a single population. Thus, when the W_i can be treated as observations from a normal population, we have:

$$\frac{\bar{W} - (\mu_1 - \mu_2)}{s\{\bar{W}\}} \text{ is distributed as } t \text{ with } n-1 \text{ degrees of freedom when the differences } W_i \text{ can be considered to be observations from a normal population and:} \tag{A.69}$$

$$\bar{W} = \frac{\sum_i W_i}{n} \qquad s^2\{\bar{W}\} = \left(\frac{\sum_i (W_i - \bar{W})^2}{n-1}\right) \div n$$

A.8 Inferences about Population Variance—Normal Population

When sampling from a normal population, the following holds for the sample variance s^2, where s is defined in (A.58b):

$$\frac{(n-1)s^2}{\sigma^2} \text{ is distributed as } \chi^2 \text{ with } n-1 \text{ degrees of freedom when the random sample is from a normal population.} \tag{A.70}$$

Interval Estimation

The lower confidence limit L and the upper confidence limit U in a confidence interval for the population variance σ^2 with confidence coefficient $1-\alpha$ are obtained by means of (A.70):

$$L = \frac{(n-1)s^2}{\chi^2(1-\alpha/2; n-1)} \qquad U = \frac{(n-1)s^2}{\chi^2(\alpha/2; n-1)} \tag{A.71}$$

Example 8 Obtain a 98 percent confidence interval for σ^2, using the data of Example 1 ($n = 10, s = 4$). We require:

$$s^2 = 16 \qquad \chi^2(.01; 9) = 2.09 \qquad \chi^2(.99; 9) = 21.67$$

The 98 percent confidence interval for σ^2 therefore is:

$$6.6 = \frac{9(16)}{21.67} \leq \sigma^2 \leq \frac{9(16)}{2.09} = 68.9$$

Tests

One-sided and two-sided tests concerning the population variance σ^2 are constructed by means of (A.70). Table A.3 contains the decision rules for three possible cases, with the risk of making a Type I error controlled at α.

TABLE A.3 Decision Rules for Tests Concerning Variance σ^2 of Normal Populations.

Alternatives	Decision Rule
	(a)
H_0: $\sigma^2 = \sigma_0^2$ H_a: $\sigma^2 \neq \sigma_0^2$	If $\chi^2(\alpha/2; n-1) \leq \dfrac{(n-1)s^2}{\sigma_0^2} \leq \chi^2(1-\alpha/2; n-1)$, conclude H_0 Otherwise conclude H_a
	(b)
H_0: $\sigma^2 \geq \sigma_0^2$	If $\dfrac{(n-1)s^2}{\sigma_0^2} \geq \chi^2(\alpha; n-1)$, conclude H_0
H_a: $\sigma^2 < \sigma_0^2$	If $\dfrac{(n-1)s^2}{\sigma_0^2} < \chi^2(\alpha; n-1)$, conclude H_a
	(c)
H_0: $\sigma^2 \leq \sigma_0^2$	If $\dfrac{(n-1)s^2}{\sigma_0^2} \leq \chi^2(1-\alpha; n-1)$, conclude H_0
H_a: $\sigma^2 > \sigma_0^2$	If $\dfrac{(n-1)s^2}{\sigma_0^2} > \chi^2(1-\alpha; n-1)$, conclude H_a

Comment

The inference procedures about the population variance described here are very sensitive to the assumption of a normal population, and the procedures are not robust to departures from normality. ■

A.9 Comparisons of Two Population Variances—Normal Populations

Independent samples are selected from two normal populations, with means and variances μ_1 and σ_1^2 and μ_2 and σ_2^2, respectively. Using the notation of Section A.7, the two sample variances are:

$$s_1^2 = \frac{\sum_i (Y_i - \bar{Y})^2}{n_1 - 1} \tag{A.72a}$$

$$s_2^2 = \frac{\sum_i (Z_i - \bar{Z})^2}{n_2 - 1} \tag{A.72b}$$

We have:

$$\frac{s_1^2}{\sigma_1^2} \div \frac{s_2^2}{\sigma_2^2} \text{ is distributed as } F(n_1 - 1, n_2 - 1) \text{ when the two independent samples come from normal populations.} \tag{A.73}$$

Interval Estimation

The lower and upper confidence limits L and U for σ_1^2/σ_2^2 with confidence coefficient $1-\alpha$ are obtained by means of (A.73):

$$L = \frac{s_1^2}{s_2^2}\left[\frac{1}{F(1-\alpha/2; n_1-1, n_2-1)}\right]$$
$$U = \frac{s_1^2}{s_2^2}\left[\frac{1}{F(\alpha/2; n_1-1, n_2-1)}\right] \tag{A.74}$$

Example 9

Obtain a 90 percent confidence interval for σ_1^2/σ_2^2 when the data are:

$$n_1 = 16 \qquad n_2 = 21 \qquad s_1^2 = 54.2 \qquad s_2^2 = 17.8$$

We require:

$$F(.05; 15, 20) = 1/F(.95; 20, 15) = 1/2.33 = .429$$
$$F(.95; 15, 20) = 2.20$$

The 90 percent confidence interval for σ_1^2/σ_2^2 therefore is:

$$1.4 = \frac{54.2}{17.8}\left(\frac{1}{2.20}\right) \le \frac{\sigma_1^2}{\sigma_2^2} \le \frac{54.2}{17.8}\left(\frac{1}{.429}\right) = 7.1$$

Tests

One-sided and two-sided tests concerning σ_1^2/σ_2^2 are constructed by means of (A.73). Table A.4 contains the decision rules for three possible cases, with the risk of making a Type I error controlled at α.

TABLE A.4 Decision Rules for Tests Concerning Variances σ_1^2 and σ_2^2 of Two Normal Populations—Independent Samples.

Alternatives	Decision Rule
	(a)
H_0: $\sigma_1^2 = \sigma_2^2$	If $F(\alpha/2; n_1-1, n_2-1) \le \frac{s_1^2}{s_2^2}$
H_a: $\sigma_1^2 \ne \sigma_2^2$	$\le F(1-\alpha/2; n_1-1, n_2-1)$, conclude H_0
	Otherwise conclude H_a
	(b)
H_0: $\sigma_1^2 \ge \sigma_2^2$	If $\frac{s_1^2}{s_2^2} \ge F(\alpha; n_1-1, n_2-1)$, conclude H_0
H_a: $\sigma_1^2 < \sigma_2^2$	If $\frac{s_1^2}{s_2^2} < F(\alpha; n_1-1, n_2-1)$, conclude H_a
	(c)
H_0: $\sigma_1^2 \le \sigma_2^2$	If $\frac{s_1^2}{s_2^2} \le F(1-\alpha; n_1-1, n_2-1)$, conclude H_0
H_a: $\sigma_1^2 > \sigma_2^2$	If $\frac{s_1^2}{s_2^2} > F(1-\alpha; n_1-1, n_2-1)$, conclude H_a

Example 10 Choose between the alternatives:

$$H_0\colon \sigma_1^2 = \sigma_2^2 \qquad H_a\colon \sigma_1^2 \neq \sigma_2^2$$

when α is to be controlled at .02 and the data are those of Example 9.

We require:

$$F(.01; 15, 20) = 1/F(.99; 20, 15) = 1/3.37 = .297$$
$$F(.99; 15, 20) = 3.09$$

The decision rule is:

$$\text{If } .297 \leq \frac{s_1^2}{s_2^2} \leq 3.09, \text{ conclude } H_0$$
$$\text{Otherwise conclude } H_a$$

Since $s_1^2/s_2^2 = 54.2/17.8 = 3.04$, we conclude H_0.

Comment

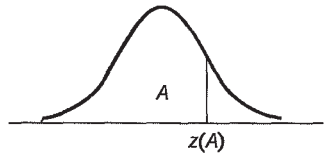
The inference procedures about the ratio of two population variances described here are very sensitive to the assumption of normal populations, and the procedures are not robust to departures from normality. ■

Appendix B

Tables

TABLE B.1 Cumulative Probabilities of the Standard Normal Distribution.

Entry is area A under the standard normal curve from $-\infty$ to $z(A)$



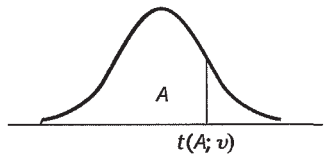
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Selected Percentiles

Cumulative probability A:	.90	.95	.975	.98	.99	.995	.999
$z(A)$:	1.282	1.645	1.960	2.054	2.326	2.576	3.090

TABLE B.2
Percentiles of the t Distribution.

Entry is $t(A;\nu)$ where $P\{t(\nu) \le t(A;\nu)\} = A$



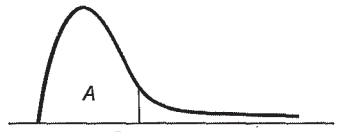
	A						
ν	.60	.70	.80	.85	.90	.95	.975
1	0.325	0.727	1.376	1.963	3.078	6.314	12.706
2	0.289	0.617	1.061	1.386	1.886	2.920	4.303
3	0.277	0.584	0.978	1.250	1.638	2.353	3.182
4	0.271	0.569	0.941	1.190	1.533	2.132	2.776
5	0.267	0.559	0.920	1.156	1.476	2.015	2.571
6	0.265	0.553	0.906	1.134	1.440	1.943	2.447
7	0.263	0.549	0.896	1.119	1.415	1.895	2.365
8	0.262	0.546	0.889	1.108	1.397	1.860	2.306
9	0.261	0.543	0.883	1.100	1.383	1.833	2.262
10	0.260	0.542	0.879	1.093	1.372	1.812	2.228
11	0.260	0.540	0.876	1.088	1.363	1.796	2.201
12	0.259	0.539	0.873	1.083	1.356	1.782	2.179
13	0.259	0.537	0.870	1.079	1.350	1.771	2.160
14	0.258	0.537	0.868	1.076	1.345	1.761	2.145
15	0.258	0.536	0.866	1.074	1.341	1.753	2.131
16	0.258	0.535	0.865	1.071	1.337	1.746	2.120
17	0.257	0.534	0.863	1.069	1.333	1.740	2.110
18	0.257	0.534	0.862	1.067	1.330	1.734	2.101
19	0.257	0.533	0.861	1.066	1.328	1.729	2.093
20	0.257	0.533	0.860	1.064	1.325	1.725	2.086
21	0.257	0.532	0.859	1.063	1.323	1.721	2.080
22	0.256	0.532	0.858	1.061	1.321	1.717	2.074
23	0.256	0.532	0.858	1.060	1.319	1.714	2.069
24	0.256	0.531	0.857	1.059	1.318	1.711	2.064
25	0.256	0.531	0.856	1.058	1.316	1.708	2.060
26	0.256	0.531	0.856	1.058	1.315	1.706	2.056
27	0.256	0.531	0.855	1.057	1.314	1.703	2.052
28	0.256	0.530	0.855	1.056	1.313	1.701	2.048
29	0.256	0.530	0.854	1.055	1.311	1.699	2.045
30	0.256	0.530	0.854	1.055	1.310	1.697	2.042
40	0.255	0.529	0.851	1.050	1.303	1.684	2.021
60	0.254	0.527	0.848	1.045	1.296	1.671	2.000
120	0.254	0.526	0.845	1.041	1.289	1.658	1.980
∞	0.253	0.524	0.842	1.036	1.282	1.645	1.960

TABLE B.2 (*concluded*) **Percentiles of the t Distribution.**

	A						
ν	.98	.985	.99	.9925	.995	.9975	.9995
1	15.895	21.205	31.821	42.434	63.657	127.322	636.590
2	4.849	5.643	6.965	8.073	9.925	14.089	31.598
3	3.482	3.896	4.541	5.047	5.841	7.453	12.924
4	2.999	3.298	3.747	4.088	4.604	5.598	8.610
5	2.757	3.003	3.365	3.634	4.032	4.773	6.869
6	2.612	2.829	3.143	3.372	3.707	4.317	5.959
7	2.517	2.715	2.998	3.203	3.499	4.029	5.408
8	2.449	2.634	2.896	3.085	3.355	3.833	5.041
9	2.398	2.574	2.821	2.998	3.250	3.690	4.781
10	2.359	2.527	2.764	2.932	3.169	3.581	4.587
11	2.328	2.491	2.718	2.879	3.106	3.497	4.437
12	2.303	2.461	2.681	2.836	3.055	3.428	4.318
13	2.282	2.436	2.650	2.801	3.012	3.372	4.221
14	2.264	2.415	2.624	2.771	2.977	3.326	4.140
15	2.249	2.397	2.602	2.746	2.947	3.286	4.073
16	2.235	2.382	2.583	2.724	2.921	3.252	4.015
17	2.224	2.368	2.567	2.706	2.898	3.222	3.965
18	2.214	2.356	2.552	2.689	2.878	3.197	3.922
19	2.205	2.346	2.539	2.674	2.861	3.174	3.883
20	2.197	2.336	2.528	2.661	2.845	3.153	3.849
21	2.189	2.328	2.518	2.649	2.831	3.135	3.819
22	2.183	2.320	2.508	2.639	2.819	3.119	3.792
23	2.177	2.313	2.500	2.629	2.807	3.104	3.768
24	2.172	2.307	2.492	2.620	2.797	3.091	3.745
25	2.167	2.301	2.485	2.612	2.787	3.078	3.725
26	2.162	2.296	2.479	2.605	2.779	3.067	3.707
27	2.158	2.291	2.473	2.598	2.771	3.057	3.690
28	2.154	2.286	2.467	2.592	2.763	3.047	3.674
29	2.150	2.282	2.462	2.586	2.756	3.038	3.659
30	2.147	2.278	2.457	2.581	2.750	3.030	3.646
40	2.123	2.250	2.423	2.542	2.704	2.971	3.551
60	2.099	2.223	2.390	2.504	2.660	2.915	3.460
120	2.076	2.196	2.358	2.468	2.617	2.860	3.373
∞	2.054	2.170	2.326	2.432	2.576	2.807	3.291

TABLE B.3 **Percentiles of the χ^2 Distribution.**

Entry is $\chi^2(A;\nu)$ where $P\{\chi^2(\nu) \le \chi^2(A;\nu)\} = A$

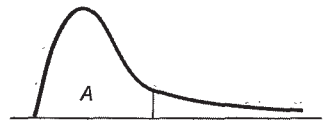


	A									
ν	.005	.010	.025	.050	.100	.900	.950	.975	.990	.995
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

Source: Reprinted, with permission, from C. M. Thompson, "Table of Percentage Points of the Chi-Square Distribution," *Biometrika* 32 (1941), pp. 188–89.

TABLE B.4 Percentiles of the F Distribution.

Entry is $F(A; \nu_1, \nu_2)$ where $P\{F(\nu_1, \nu_2) \leq F(A; \nu_1, \nu_2)\} = A$



$F(A; \nu_1, \nu_2)$

$$F(A; \nu_1, \nu_2) = \frac{1}{F(1 - A; \nu_2, \nu_1)}$$

TABLE B.4 (*continued*) **Percentiles of the *F* Distribution.**

Den. df	*A*	Numerator df 1	2	3	4	5	6	7	8	9
1	.50	1.00	1.50	1.71	1.82	1.89	1.94	1.98	2.00	2.03
	.90	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9
	.95	161	200	216	225	230	234	237	239	241
	.975	648	800	864	900	922	937	948	957	963
	.99	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,981	6,022
	.995	16,211	20,000	21,615	22,500	23,056	23,437	23,715	23,925	24,091
	.999	405,280	500,000	540,380	562,500	576,400	585,940	592,870	598,140	602,280
2	.50	0.667	1.00	1.13	1.21	1.25	1.28	1.30	1.32	1.33
	.90	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.95	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4
	.975	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4
	.99	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4
	.995	199	199	199	199	199	199	199	199	199
	.999	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4
3	.50	0.585	0.881	1.00	1.06	1.10	1.13	1.15	1.16	1.17
	.90	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.95	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	.975	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5
	.99	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3
	.995	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9
	.999	167.0	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.9
4	.50	0.549	0.828	0.941	1.00	1.04	1.06	1.08	1.09	1.10
	.90	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.95	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	.975	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.99	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7
	.995	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1
	.999	74.1	61.2	56.2	53.4	51.7	50.5	49.7	49.0	48.5
5	.50	0.528	0.799	0.907	0.965	1.00	1.02	1.04	1.05	1.06
	.90	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.95	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	.975	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.99	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2
	.995	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8
	.999	47.2	37.1	33.2	31.1	29.8	28.8	28.2	27.6	27.2
6	.50	0.515	0.780	0.886	0.942	0.977	1.00	1.02	1.03	1.04
	.90	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.95	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	.975	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	.99	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.995	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4
	.999	35.5	27.0	23.7	21.9	20.8	20.0	19.5	19.0	18.7
7	.50	0.506	0.767	0.871	0.926	0.960	0.983	1.00	1.01	1.02
	.90	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.95	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	.975	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	.99	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	.995	16.2	12.4	10.9	10.1	9.52	9.16	8.89	8.68	8.51
	.999	29.2	21.7	18.8	17.2	16.2	15.5	15.0	14.6	14.3

TABLE B.4 (*continued*) **Percentiles of the *F* Distribution.**

Den. df	A	Numerator df 10	12	15	20	24	30	60	120	∞
1	.50	2.04	2.07	2.09	2.12	2.13	2.15	2.17	2.18	2.20
	.90	60.2	60.7	61.2	61.7	62.0	62.3	62.8	63.1	63.3
	.95	242	244	246	248	249	250	252	253	254
	.975	969	977	985	993	997	1,001	1,010	1,014	1,018
	.99	6,056	6,106	6,157	6,209	6,235	6,261	6,313	6,339	6,366
	.995	24,224	24,426	24,630	24,836	24,940	25,044	25,253	25,359	25,464
	.999	605,620	610,670	615,760	620,910	623,500	626,100	631,340	633,970	636,620
2	.50	1.34	1.36	1.38	1.39	1.40	1.41	1.43	1.43	1.44
	.90	9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.48	9.49
	.95	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
	.975	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5	39.5
	.99	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5
	.995	199	199	199	199	199	199	199	199	200
	.999	999.4	999.4	999.4	999.4	999.5	999.5	999.5	999.5	999.5
3	.50	1.18	1.20	1.21	1.23	1.23	1.24	1.25	1.26	1.27
	.90	5.23	5.22	5.20	5.18	5.18	5.17	5.15	5.14	5.13
	.95	8.79	8.74	8.70	8.66	8.64	8.62	8.57	8.55	8.53
	.975	14.4	14.3	14.3	14.2	14.1	14.1	14.0	13.9	13.9
	.99	27.2	27.1	26.9	26.7	26.6	26.5	26.3	26.2	26.1
	.995	43.7	43.4	43.1	42.8	42.6	42.5	42.1	42.0	41.8
	.999	129.2	128.3	127.4	126.4	125.9	125.4	124.5	124.0	123.5
4	.50	1.11	1.13	1.14	1.15	1.16	1.16	1.18	1.18	1.19
	.90	3.92	3.90	3.87	3.84	3.83	3.82	3.79	3.78	3.76
	.95	5.96	5.91	5.86	5.80	5.77	5.75	5.69	5.66	5.63
	.975	8.84	8.75	8.66	8.56	8.51	8.46	8.36	8.31	8.26
	.99	14.5	14.4	14.2	14.0	13.9	13.8	13.7	13.6	13.5
	.995	21.0	20.7	20.4	20.2	20.0	19.9	19.6	19.5	19.3
	.999	48.1	47.4	46.8	46.1	45.8	45.4	44.7	44.4	44.1
5	.50	1.07	1.09	1.10	1.11	1.12	1.12	1.14	1.14	1.15
	.90	3.30	3.27	3.24	3.21	3.19	3.17	3.14	3.12	3.11
	.95	4.74	4.68	4.62	4.56	4.53	4.50	4.43	4.40	4.37
	.975	6.62	6.52	6.43	6.33	6.28	6.23	6.12	6.07	6.02
	.99	10.1	9.89	9.72	9.55	9.47	9.38	9.20	9.11	9.02
	.995	13.6	13.4	13.1	12.9	12.8	12.7	12.4	12.3	12.1
	.999	26.9	26.4	25.9	25.4	25.1	24.9	24.3	24.1	23.8
6	.50	1.05	1.06	1.07	1.08	1.09	1.10	1.11	1.12	1.12
	.90	2.94	2.90	2.87	2.84	2.82	2.80	2.76	2.74	2.72
	.95	4.06	4.00	3.94	3.87	3.84	3.81	3.74	3.70	3.67
	.975	5.46	5.37	5.27	5.17	5.12	5.07	4.96	4.90	4.85
	.99	7.87	7.72	7.56	7.40	7.31	7.23	7.06	6.97	6.88
	.995	10.2	10.0	9.81	9.59	9.47	9.36	9.12	9.00	8.88
	.999	18.4	18.0	17.6	17.1	16.9	16.7	16.2	16.0	15.7
7	.50	1.03	1.04	1.05	1.07	1.07	1.08	1.09	1.10	1.10
	.90	2.70	2.67	2.63	2.59	2.58	2.56	2.51	2.49	2.47
	.95	3.64	3.57	3.51	3.44	3.41	3.38	3.30	3.27	3.23
	.975	4.76	4.67	4.57	4.47	4.42	4.36	4.25	4.20	4.14
	.99	6.62	6.47	6.31	6.16	6.07	5.99	5.82	5.74	5.65
	.995	8.38	8.18	7.97	7.75	7.65	7.53	7.31	7.19	7.08
	.999	14.1	13.7	13.3	12.9	12.7	12.5	12.1	11.9	11.7

TABLE B.4 (*continued*) **Percentiles of the *F* Distribution.**

Den. df	A	Numerator df 1	2	3	4	5	6	7	8	9
8	.50	0.499	0.757	0.860	0.915	0.948	0.971	0.988	1.00	1.01
	.90	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	.95	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	.975	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	.99	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	.995	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34
	.999	25.4	18.5	15.8	14.4	13.5	12.9	12.4	12.0	11.8
9	.50	0.494	0.749	0.852	0.906	0.939	0.962	0.978	0.990	1.00
	.90	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	.95	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	.975	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	.99	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	.995	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54
	.999	22.9	16.4	13.9	12.6	11.7	11.1	10.7	10.4	10.1
10	.50	0.490	0.743	0.845	0.899	0.932	0.954	0.971	0.983	0.992
	.90	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	.95	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	.975	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	.99	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	.995	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97
	.999	21.0	14.9	12.6	11.3	10.5	9.93	9.52	9.20	8.96
12	.50	0.484	0.735	0.835	0.888	0.921	0.943	0.959	0.972	0.981
	.90	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	.95	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	.975	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	.99	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	.995	11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20
	.999	18.6	13.0	10.8	9.63	8.89	8.38	8.00	7.71	7.48
15	.50	0.478	0.726	0.826	0.878	0.911	0.933	0.949	0.960	0.970
	.90	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	.95	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	.975	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
	.99	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	.995	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54
	.999	16.6	11.3	9.34	8.25	7.57	7.09	6.74	6.47	6.26
20	.50	0.472	0.718	0.816	0.868	0.900	0.922	0.938	0.950	0.959
	.90	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96
	.95	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
	.975	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84
	.99	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46
	.995	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96
	.999	14.8	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24
24	.50	0.469	0.714	0.812	0.863	0.895	0.917	0.932	0.944	0.953
	.90	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91
	.95	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
	.975	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70
	.99	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26
	.995	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69
	.999	14.0	9.34	7.55	6.59	5.98	5.55	5.23	4.99	4.80

TABLE B.4 (*continued*) **Percentiles of the *F* Distribution.**

Den. df	*A*	Numerator df 10	12	15	20	24	30	60	120	∞
8	.50	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.08	1.09
	.90	2.54	2.50	2.46	2.42	2.40	2.38	2.34	2.32	2.29
	.95	3.35	3.28	3.22	3.15	3.12	3.08	3.01	2.97	2.93
	.975	4.30	4.20	4.10	4.00	3.95	3.89	3.78	3.73	3.67
	.99	5.81	5.67	5.52	5.36	5.28	5.20	5.03	4.95	4.86
	.995	7.21	7.01	6.81	6.61	6.50	6.40	6.18	6.06	5.95
	.999	11.5	11.2	10.8	10.5	10.3	10.1	9.73	9.53	9.33
9	.50	1.01	1.02	1.03	1.04	1.05	1.05	1.07	1.07	1.08
	.90	2.42	2.38	2.34	2.30	2.28	2.25	2.21	2.18	2.16
	.95	3.14	3.07	3.01	2.94	2.90	2.86	2.79	2.75	2.71
	.975	3.96	3.87	3.77	3.67	3.61	3.56	3.45	3.39	3.33
	.99	5.26	5.11	4.96	4.81	4.73	4.65	4.48	4.40	4.31
	.995	6.42	6.23	6.03	5.83	5.73	5.62	5.41	5.30	5.19
	.999	9.89	9.57	9.24	8.90	8.72	8.55	8.19	8.00	7.81
10	.50	1.00	1.01	1.02	1.03	1.04	1.05	1.06	1.06	1.07
	.90	2.32	2.28	2.24	2.20	2.18	2.16	2.11	2.08	2.06
	.95	2.98	2.91	2.84	2.77	2.74	2.70	2.62	2.58	2.54
	.975	3.72	3.62	3.52	3.42	3.37	3.31	3.20	3.14	3.08
	.99	4.85	4.71	4.56	4.41	4.33	4.25	4.08	4.00	3.91
	.995	5.85	5.66	5.47	5.27	5.17	5.07	4.86	4.75	4.64
	.999	8.75	8.45	8.13	7.80	7.64	7.47	7.12	6.94	6.76
12	.50	0.989	1.00	1.01	1.02	1.03	1.03	1.05	1.05	1.06
	.90	2.19	2.15	2.10	2.06	2.04	2.01	1.96	1.93	1.90
	.95	2.75	2.69	2.62	2.54	2.51	2.47	2.38	2.34	2.30
	.975	3.37	3.28	3.18	3.07	3.02	2.96	2.85	2.79	2.72
	.99	4.30	4.16	4.01	3.86	3.78	3.70	3.54	3.45	3.36
	.995	5.09	4.91	4.72	4.53	4.43	4.33	4.12	4.01	3.90
	.999	7.29	7.00	6.71	6.40	6.25	6.09	5.76	5.59	5.42
15	.50	0.977	0.989	1.00	1.01	1.02	1.02	1.03	1.04	1.05
	.90	2.06	2.02	1.97	1.92	1.90	1.87	1.82	1.79	1.76
	.95	2.54	2.48	2.40	2.33	2.29	2.25	2.16	2.11	2.07
	.975	3.06	2.96	2.86	2.76	2.70	2.64	2.52	2.46	2.40
	.99	3.80	3.67	3.52	3.37	3.29	3.21	3.05	2.96	2.87
	.995	4.42	4.25	4.07	3.88	3.79	3.69	3.48	3.37	3.26
	.999	6.08	5.81	5.54	5.25	5.10	4.95	4.64	4.48	4.31
20	.50	0.966	0.977	0.989	1.00	1.01	1.01	1.02	1.03	1.03
	.90	1.94	1.89	1.84	1.79	1.77	1.74	1.68	1.64	1.61
	.95	2.35	2.28	2.20	2.12	2.08	2.04	1.95	1.90	1.84
	.975	2.77	2.68	2.57	2.46	2.41	2.35	2.22	2.16	2.09
	.99	3.37	3.23	3.09	2.94	2.86	2.78	2.61	2.52	2.42
	.995	3.85	3.68	3.50	3.32	3.22	3.12	2.92	2.81	2.69
	.999	5.08	4.82	4.56	4.29	4.15	4.00	3.70	3.54	3.38
24	.50	0.961	0.972	0.983	0.994	1.00	1.01	1.02	1.02	1.03
	.90	1.88	1.83	1.78	1.73	1.70	1.67	1.61	1.57	1.53
	.95	2.25	2.18	2.11	2.03	1.98	1.94	1.84	1.79	1.73
	.975	2.64	2.54	2.44	2.33	2.27	2.21	2.08	2.01	1.94
	.99	3.17	3.03	2.89	2.74	2.66	2.58	2.40	2.31	2.21
	.995	3.59	3.42	3.25	3.06	2.97	2.87	2.66	2.55	2.43
	.999	4.64	4.39	4.14	3.87	3.74	3.59	3.29	3.14	2.97

TABLE B.4 (*continued*) **Percentiles of the *F* Distribution.**

Den. df	*A*	Numerator df 1	2	3	4	5	6	7	8	9
30	.50	0.466	0.709	0.807	0.858	0.890	0.912	0.927	0.939	0.948
	.90	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85
	.95	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
	.975	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57
	.99	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07
	.995	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45
	.999	13.3	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.39
60	.50	0.461	0.701	0.798	0.849	0.880	0.901	0.917	0.928	0.937
	.90	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74
	.95	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
	.975	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33
	.99	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72
	.995	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01
	.999	12.0	7.77	6.17	5.31	4.76	4.37	4.09	3.86	3.69
120	.50	0.458	0.697	0.793	0.844	0.875	0.896	0.912	0.923	0.932
	.90	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68
	.95	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96
	.975	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22
	.99	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56
	.995	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81
	.999	11.4	7.32	5.78	4.95	4.42	4.04	3.77	3.55	3.38
∞	.50	0.455	0.693	0.789	0.839	0.870	0.891	0.907	0.918	0.927
	.90	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63
	.95	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88
	.975	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11
	.99	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41
	.995	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62
	.999	10.8	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10

TABLE B.4 (*concluded*) **Percentiles of the *F* Distribution.**

Den. df	*A*	Numerator df 10	12	15	20	24	30	60	120	∞
30	.50	0.955	0.966	0.978	0.989	0.994	1.00	1.01	1.02	1.02
	.90	1.82	1.77	1.72	1.67	1.64	1.61	1.54	1.50	1.46
	.95	2.16	2.09	2.01	1.93	1.89	1.84	1.74	1.68	1.62
	.975	2.51	2.41	2.31	2.20	2.14	2.07	1.94	1.87	1.79
	.99	2.98	2.84	2.70	2.55	2.47	2.39	2.21	2.11	2.01
	.995	3.34	3.18	3.01	2.82	2.73	2.63	2.42	2.30	2.18
	.999	4.24	4.00	3.75	3.49	3.36	3.22	2.92	2.76	2.59
60	.50	0.945	0.956	0.967	0.978	0.983	0.989	1.00	1.01	1.01
	.90	1.71	1.66	1.60	1.54	1.51	1.48	1.40	1.35	1.29
	.95	1.99	1.92	1.84	1.75	1.70	1.65	1.53	1.47	1.39
	.975	2.27	2.17	2.06	1.94	1.88	1.82	1.67	1.58	1.48
	.99	2.63	2.50	2.35	2.20	2.12	2.03	1.84	1.73	1.60
	.995	2.90	2.74	2.57	2.39	2.29	2.19	1.96	1.83	1.69
	.999	3.54	3.32	3.08	2.83	2.69	2.55	2.25	2.08	1.89
120	.50	0.939	0.950	0.961	0.972	0.978	0.983	0.994	1.00	1.01
	.90	1.65	1.60	1.55	1.48	1.45	1.41	1.32	1.26	1.19
	.95	1.91	1.83	1.75	1.66	1.61	1.55	1.43	1.35	1.25
	.975	2.16	2.05	1.95	1.82	1.76	1.69	1.53	1.43	1.31
	.99	2.47	2.34	2.19	2.03	1.95	1.86	1.66	1.53	1.38
	.995	2.71	2.54	2.37	2.19	2.09	1.98	1.75	1.61	1.43
	.999	3.24	3.02	2.78	2.53	2.40	2.26	1.95	1.77	1.54
∞	.50	0.934	0.945	0.956	0.967	0.972	0.978	0.989	0.994	1.00
	.90	1.60	1.55	1.49	1.42	1.38	1.34	1.24	1.17	1.00
	.95	1.83	1.75	1.67	1.57	1.52	1.46	1.32	1.22	1.00
	.975	2.05	1.94	1.83	1.71	1.64	1.57	1.39	1.27	1.00
	.99	2.32	2.18	2.04	1.88	1.79	1.70	1.47	1.32	1.00
	.995	2.52	2.36	2.19	2.00	1.90	1.79	1.53	1.36	1.00
	.999	2.96	2.74	2.51	2.27	2.13	1.99	1.66	1.45	1.00

Source: Reprinted from Table 5 of Pearson and Hartley, *Biometrika Tables for Statisticians,* Volume 2, 1972, published by the Cambridge University Press, on behalf of The Biometrika Society, by permission of the authors and publishers.

TABLE B.5
Power Values for Two-Sided *t* Test.

	$\alpha = .05$								
	δ								
df	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0
1	.07	.13	.19	.25	.31	.36	.42	.47	.52
2	.10	.22	.39	.56	.72	.84	.91	.96	.98
3	.11	.29	.53	.75	.90	.97	.99	1.00	1.00
4	.12	.34	.62	.84	.95	.99	1.00	1.00	1.00
5	.13	.37	.67	.89	.98	1.00	1.00	1.00	1.00
6	.14	.39	.71	.91	.98	1.00	1.00	1.00	1.00
7	.14	.41	.73	.93	.99	1.00	1.00	1.00	1.00
8	.14	.42	.75	.94	.99	1.00	1.00	1.00	1.00
9	.15	.43	.76	.94	.99	1.00	1.00	1.00	1.00
10	.15	.44	.77	.95	.99	1.00	1.00	1.00	1.00
11	.15	.45	.78	.95	.99	1.00	1.00	1.00	1.00
12	.15	.45	.79	.96	1.00	1.00	1.00	1.00	1.00
13	.15	.46	.79	.96	1.00	1.00	1.00	1.00	1.00
14	.15	.46	.80	.96	1.00	1.00	1.00	1.00	1.00
15	.16	.46	.80	.96	1.00	1.00	1.00	1.00	1.00
16	.16	.47	.80	.96	1.00	1.00	1.00	1.00	1.00
17	.16	.47	.81	.96	1.00	1.00	1.00	1.00	1.00
18	.16	.47	.81	.97	1.00	1.00	1.00	1.00	1.00
19	.16	.48	.81	.97	1.00	1.00	1.00	1.00	1.00
20	.16	.48	.81	.97	1.00	1.00	1.00	1.00	1.00
21	.16	.48	.82	.97	1.00	1.00	1.00	1.00	1.00
22	.16	.48	.82	.97	1.00	1.00	1.00	1.00	1.00
23	.16	.48	.82	.97	1.00	1.00	1.00	1.00	1.00
24	.16	.48	.82	.97	1.00	1.00	1.00	1.00	1.00
25	.16	.49	.82	.97	1.00	1.00	1.00	1.00	1.00
26	.16	.49	.82	.97	1.00	1.00	1.00	1.00	1.00
27	.16	.49	.82	.97	1.00	1.00	1.00	1.00	1.00
28	.16	.49	.83	.97	1.00	1.00	1.00	1.00	1.00
29	.16	.49	.83	.97	1.00	1.00	1.00	1.00	1.00
30	.16	.49	.83	.97	1.00	1.00	1.00	1.00	1.00
40	.16	.50	.83	.97	1.00	1.00	1.00	1.00	1.00
50	.17	.50	.84	.98	1.00	1.00	1.00	1.00	1.00
60	.17	.50	.84	.98	1.00	1.00	1.00	1.00	1.00
100	.17	.51	.84	.98	1.00	1.00	1.00	1.00	1.00
120	.17	.51	.85	.98	1.00	1.00	1.00	1.00	1.00
∞	.17	.52	.85	.98	1.00	1.00	1.00	1.00	1.00

TABLE B.5 (*concluded*) **Power Values for Two-Sided t Test.**

	$\alpha = .01$								
	δ								
df	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0
1	.01	.03	.04	.05	.06	.08	.09	.10	.11
2	.02	.05	.09	.16	.23	.31	.39	.48	.56
3	.02	.08	.17	.31	.47	.62	.75	.85	.92
4	.03	.10	.25	.45	.65	.82	.92	.97	.99
5	.03	.12	.31	.55	.77	.91	.97	.99	1.00
6	.04	.14	.36	.63	.84	.95	.99	1.00	1.00
7	.04	.16	.40	.68	.88	.97	1.00	1.00	1.00
8	.04	.17	.43	.72	.91	.98	1.00	1.00	1.00
9	.04	.18	.45	.75	.93	.99	1.00	1.00	1.00
10	.04	.19	.47	.77	.94	.99	1.00	1.00	1.00
11	.04	.19	.49	.79	.95	.99	1.00	1.00	1.00
12	.04	.20	.50	.80	.96	.99	1.00	1.00	1.00
13	.05	.21	.52	.82	.96	1.00	1.00	1.00	1.00
14	.05	.21	.53	.83	.96	1.00	1.00	1.00	1.00
15	.05	.21	.54	.83	.97	1.00	1.00	1.00	1.00
16	.05	.22	.55	.84	.97	1.00	1.00	1.00	1.00
17	.05	.22	.55	.85	.97	1.00	1.00	1.00	1.00
18	.05	.22	.56	.85	.97	1.00	1.00	1.00	1.00
19	.05	.23	.56	.86	.98	1.00	1.00	1.00	1.00
20	.05	.23	.57	.86	.98	1.00	1.00	1.00	1.00
21	.05	.23	.57	.86	.98	1.00	1.00	1.00	1.00
22	.05	.23	.58	.87	.98	1.00	1.00	1.00	1.00
23	.05	.24	.58	.87	.98	1.00	1.00	1.00	1.00
24	.05	.24	.59	.87	.98	1.00	1.00	1.00	1.00
25	.05	.24	.59	.88	.98	1.00	1.00	1.00	1.00
26	.05	.24	.59	.88	.98	1.00	1.00	1.00	1.00
27	.05	.24	.59	.88	.98	1.00	1.00	1.00	1.00
28	.05	.24	.60	.88	.98	1.00	1.00	1.00	1.00
29	.05	.25	.60	.88	.98	1.00	1.00	1.00	1.00
30	.05	.25	.60	.88	.98	1.00	1.00	1.00	1.00
40	.05	.26	.62	.90	.99	1.00	1.00	1.00	1.00
50	.05	.26	.63	.90	.99	1.00	1.00	1.00	1.00
60	.05	.26	.63	.91	.99	1.00	1.00	1.00	1.00
100	.06	.27	.65	.91	.99	1.00	1.00	1.00	1.00
120	.06	.27	.65	.91	.99	1.00	1.00	1.00	1.00
∞	.06	.28	.66	.92	.99	1.00	1.00	1.00	1.00

TABLE B.6
Critical Values for Coefficient of Correlation between Ordered Residuals and Expected Values under Normality when Distribution of Error Terms Is Normal.

	Level of Significance α				
n	.10	.05	.025	.01	.005
5	.903	.880	.865	.826	.807
6	.910	.888	.866	.838	.820
7	.918	.898	.877	.850	.828
8	.924	.906	.887	.861	.840
9	.930	.912	.894	.871	.854
10	.934	.918	.901	.879	.862
12	.942	.928	.912	.892	.876
14	.948	.935	.923	.905	.890
16	.953	.941	.929	.913	.899
18	.957	.946	.935	.920	.908
20	.960	.951	.940	.926	.916
22	.963	.954	.945	.933	.923
24	.965	.957	.949	.937	.927
26	.967	.960	.952	.941	.932
28	.969	.962	.955	.944	.936
30	.971	.964	.957	.947	.939
40	.977	.972	.966	.959	.953
50	.981	.977	.972	.966	.961
60	.984	.980	.976	.971	.967
70	.986	.983	.979	.975	.971
80	.987	.985	.982	.978	.975
90	.988	.986	.984	.980	.977
100	.989	.987	.985	.982	.979

Source: Reprinted, with permission, from S. W. Looney and T. R. Gulledge, Jr., "Use of the Correlation Coefficient with Normal Probability Plots," *The American Statistician* 39 (1985), pp. 75–79.

TABLE B.7
Durbin-Watson Test Bounds.

	Level of Significance $\alpha = .05$									
	$p-1=1$		$p-1=2$		$p-1=3$		$p-1=4$		$p-1=5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

TABLE B.7 ***(concluded)*** **Durbin-Watson Test Bounds.**

	Level of Significance $\alpha = .01$									
	$p-1=1$		$p-1=2$		$p-1=3$		$p-1=4$		$p-1=5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

Source: Reprinted, with permission, from J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression. II," *Biometrika* 38 (1951), pp. 159–78.

TABLE B.8
Table of z' Transformation of Correlation Coefficient.

r ρ	z' ζ	r ρ	z' ζ	r ρ	z' ζ	r ρ	z' ζ
.00	.0000	.25	.2554	.50	.5493	.75	.973
.01	.0100	.26	.2661	.51	.5627	.76	.996
.02	.0200	.27	.2769	.52	.5763	.77	1.020
.03	.0300	.28	.2877	.53	.5901	.78	1.045
.04	.0400	.29	.2986	.54	.6042	.79	1.071
.05	.0500	.30	.3095	.55	.6184	.80	1.099
.06	.0601	.31	.3205	.56	.6328	.81	1.127
.07	.0701	.32	.3316	.57	.6475	.82	1.157
.08	.0802	.33	.3428	.58	.6625	.83	1.188
.09	.0902	.34	.3541	.59	.6777	.84	1.221
.10	.1003	.35	.3654	.60	.6931	.85	1.256
.11	.1104	.36	.3769	.61	.7089	.86	1.293
.12	.1206	.37	.3884	.62	.7250	.87	1.333
.13	.1307	.38	.4001	.63	.7414	.88	1.376
.14	.1409	.39	.4118	.64	.7582	.89	1.422
.15	.1511	.40	.4236	.65	.7753	.90	1.472
.16	.1614	.41	.4356	.66	.7928	.91	1.528
.17	.1717	.42	.4477	.67	.8107	.92	1.589
.18	.1820	.43	.4599	.68	.8291	.93	1.658
.19	.1923	.44	.4722	.69	.8480	.94	1.738
.20	.2027	.45	.4847	.70	.8673	.95	1.832
.21	.2132	.46	.4973	.71	.8872	.96	1.946
.22	.2237	.47	.5101	.72	.9076	.97	2.092
.23	.2342	.48	.5230	.73	.9287	.98	2.298
.24	.2448	.49	.5361	.74	.9505	.99	2.647

Source: Abridged from Table 14 of Pearson and Hartley, *Biometrika Tables for Statisticians,* Volume 1, 1966, published by the Cambridge University Press, on behalf of The Biometrika Society, by permission of the authors and publishers.

TABLE B.9 **Percentiles of the Studentized Range Distribution.**

Entry is $q(1-\alpha; r, \nu)$ where $P\{q(r, \nu) \leq q(1-\alpha; r, \nu)\} = 1-\alpha$

$1-\alpha = .90$

	r																		
ν	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	8.93	13.4	16.4	18.5	20.2	21.5	22.6	23.6	24.5	25.2	25.9	26.5	27.1	27.6	28.1	28.5	29.0	29.3	29.7
2	4.13	5.73	6.77	7.54	8.14	8.63	9.05	9.41	9.72	10.0	10.3	10.5	10.7	10.9	11.1	11.2	11.4	11.5	11.7
3	3.33	4.47	5.20	5.74	6.16	6.51	6.81	7.06	7.29	7.49	7.67	7.83	7.98	8.12	8.25	8.37	8.48	8.58	8.68
4	3.01	3.98	4.59	5.03	5.39	5.68	5.93	6.14	6.33	6.49	6.65	6.78	6.91	7.02	7.13	7.23	7.33	7.41	7.50
5	2.85	3.72	4.26	4.66	4.98	5.24	5.46	5.65	5.82	5.97	6.10	6.22	6.34	6.44	6.54	6.63	6.71	6.79	6.86
6	2.75	3.56	4.07	4.44	4.73	4.97	5.17	5.34	5.50	5.64	5.76	5.87	5.98	6.07	6.16	6.25	6.32	6.40	6.47
7	2.68	3.45	3.93	4.28	4.55	4.78	4.97	5.14	5.28	5.41	5.53	5.64	5.74	5.83	5.91	5.99	6.06	6.13	6.19
8	2.63	3.37	3.83	4.17	4.43	4.65	4.83	4.99	5.13	5.25	5.36	5.46	5.56	5.64	5.72	5.80	5.87	5.93	6.00
9	2.59	3.32	3.76	4.08	4.34	4.54	4.72	4.87	5.01	5.13	5.23	5.33	5.42	5.51	5.58	5.66	5.72	5.79	5.85
10	2.56	3.27	3.70	4.02	4.26	4.47	4.64	4.78	4.91	5.03	5.13	5.23	5.32	5.40	5.47	5.54	5.61	5.67	5.73
11	2.54	3.23	3.66	3.96	4.20	4.40	4.57	4.71	4.84	4.95	5.05	5.15	5.23	5.31	5.38	5.45	5.51	5.57	5.63
12	2.52	3.20	3.62	3.92	4.16	4.35	4.51	4.65	4.78	4.89	4.99	5.08	5.16	5.24	5.31	5.37	5.44	5.49	5.55
13	2.50	3.18	3.59	3.88	4.12	4.30	4.46	4.60	4.72	4.83	4.93	5.02	5.10	5.18	5.25	5.31	5.37	5.43	5.48
14	2.49	3.16	3.56	3.85	4.08	4.27	4.42	4.56	4.68	4.79	4.88	4.97	5.05	5.12	5.19	5.26	5.32	5.37	5.43
15	2.48	3.14	3.54	3.83	4.05	4.23	4.39	4.52	4.64	4.75	4.84	4.93	5.01	5.08	5.15	5.21	5.27	5.32	5.38
16	2.47	3.12	3.52	3.80	4.03	4.21	4.36	4.49	4.61	4.71	4.81	4.89	4.97	5.04	5.11	5.17	5.23	5.28	5.33
17	2.46	3.11	3.50	3.78	4.00	4.18	4.33	4.46	4.58	4.68	4.77	4.86	4.93	5.01	5.07	5.13	5.19	5.24	5.30
18	2.45	3.10	3.49	3.77	3.93	4.16	4.31	4.44	4.55	4.65	4.75	4.83	4.90	4.98	5.04	5.10	5.16	5.21	5.26
19	2.45	3.09	3.47	3.75	3.97	4.14	4.29	4.42	4.53	4.63	4.72	4.80	4.88	4.95	5.01	5.07	5.13	5.18	5.23
20	2.44	3.08	3.46	3.74	3.95	4.12	4.27	4.40	4.51	4.61	4.70	4.78	4.85	4.92	4.99	5.05	5.10	5.16	5.20
24	2.42	3.05	3.42	3.69	3.90	4.07	4.21	4.34	4.44	4.54	4.63	4.71	4.78	4.85	4.91	4.97	5.02	5.07	5.12
30	2.40	3.02	3.39	3.65	3.85	4.02	4.16	4.28	4.38	4.47	4.56	4.64	4.71	4.77	4.83	4.89	4.94	4.99	5.03
40	2.38	2.99	3.35	3.60	3.80	3.96	4.10	4.21	4.32	4.41	4.49	4.56	4.63	4.69	4.75	4.81	4.86	4.90	4.95
60	2.36	2.96	3.31	3.56	3.75	3.91	4.04	4.16	4.25	4.34	4.42	4.49	4.56	4.62	4.67	4.73	4.78	4.82	4.86
120	2.34	2.93	3.28	3.52	3.71	3.86	3.99	4.10	4.19	4.28	4.35	4.42	4.48	4.54	4.60	4.65	4.69	4.74	4.78
∞	2.33	2.90	3.24	3.48	3.66	3.81	3.93	4.04	4.13	4.21	4.28	4.35	4.41	4.47	4.52	4.57	4.61	4.65	4.69

TABLE B.9 (*continued*) **Percentiles of the Studentized Range Distribution.**

$1 - \alpha = .95$

	r																		
ν	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18.0	27.0	32.8	37.1	40.4	43.1	45.4	47.4	49.1	50.6	52.0	53.2	54.3	55.4	56.3	57.2	58.0	58.8	59.6
2	6.08	8.33	9.80	10.9	11.7	12.4	13.0	13.5	14.0	14.4	14.7	15.1	15.4	15.7	15.9	16.1	16.4	16.6	16.8
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72	9.95	10.2	10.3	10.5	10.7	10.8	11.0	11.1	11.2
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

TABLE B.9 *(concluded)* **Percentiles of the Studentized Range Distribution.**

	$1-\alpha = .99$																		
	r																		
ν	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	90.0	135	164	186	202	216	227	237	246	253	260	266	272	277	282	286	290	294	298
2	14.0	19.0	22.3	24.7	26.6	28.2	29.5	30.7	31.7	32.6	33.4	34.1	34.8	35.4	36.0	36.5	37.0	37.5	37.9
3	8.26	10.6	12.2	13.3	14.2	15.0	15.6	16.2	16.7	17.1	17.5	17.9	18.2	18.5	18.8	19.1	19.3	19.5	19.8
4	6.51	8.12	9.17	9.96	10.6	11.1	11.5	11.9	12.3	12.6	12.8	13.1	13.3	13.5	13.7	13.9	14.1	14.2	14.4
5	5.70	6.97	7.80	8.42	8.91	9.32	9.67	9.97	10.2	10.5	10.7	10.9	11.1	11.2	11.4	11.6	11.7	11.8	11.9
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.49	9.65	9.81	9.95	10.1	10.2	10.3	10.4	10.5
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	4.74	5.63	6.20	6.63	6.96	7.24	7.47	7.68	7.87	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.32	7.49	7.65	7.78	7.91	8.03	8.13	8.23	8.32	8.41	8.49	8.57
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.48	7.60	7.71	7.81	7.91	7.99	8.07	8.15	8.22
11	4.39	5.14	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	4.32	5.04	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	6.90	7.01	7.10	7.19	7.27	7.34	7.42	7.48	7.55
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77	6.87	6.96	7.05	7.12	7.20	7.27	7.33	7.39
15	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
16	4.13	4.78	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48	6.57	6.66	6.73	6.80	6.87	6.94	7.00	7.05
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41	6.50	6.58	6.65	6.72	6.79	6.85	6.91	6.96
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.29	6.37	6.45	6.52	6.59	6.65	6.71	6.76	6.82
24	3.96	4.54	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	3.82	4.37	4.70	4.93	5.11	5.27	5.39	5.50	5.60	5.69	5.77	5.84	5.90	5.96	6.02	6.07	6.12	6.17	6.21
60	3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.79	5.84	5.89	5.93	5.98	6.02
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.38	5.44	5.51	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

Source: Reprinted, with permission, from Henry Scheffé, *The Analysis of Variance* (New York: John Wiley & Sons, 1959), pp. 434–36.

TABLE B.10 Percentiles of H Statistic Distribution.

Entry is $H(1-\alpha; r, df)$ where $P\{H \leq H(1-\alpha; r, df)\} = 1-\alpha$

$1-\alpha = .95$

	r										
df	2	3	4	5	6	7	8	9	10	11	12
2	39.0	87.5	142	202	266	333	403	475	550	626	704
3	15.4	27.8	39.2	50.7	62.0	72.9	83.5	93.9	104	114	124
4	9.60	15.5	20.6	25.2	29.5	33.6	37.5	41.1	44.6	48.0	51.4
5	7.15	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5	28.2	29.9
6	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6	19.7	20.7
7	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3	15.1	15.8
8	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7	12.2	12.7
9	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91	10.3	10.7
10	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66	9.01	9.34
12	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00	7.25	7.48
15	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59	5.77	5.93
20	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37	4.49	4.59
30	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29	3.36	3.39
60	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30	2.33	2.36
∞	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

$1-\alpha = .99$

	r										
df	2	3	4	5	6	7	8	9	10	11	12
2	199	448	729	1,036	1,362	1,705	2,063	2,432	2,813	3,204	3,605
3	47.5	85	120	151	184	216	249	281	310	337	361
4	23.2	37	49	59	69	79	89	97	106	113	120
5	14.9	22	28	33	38	42	46	50	54	57	60
6	11.1	15.5	19.1	22	25	27	30	32	34	36	37
7	8.89	12.1	14.5	16.5	18.4	20	22	23	24	26	27
8	7.50	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9	19.8	21
9	6.54	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3	16.0	16.6
10	5.85	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9	13.4	13.9
12	4.91	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9	10.2	10.6
15	4.07	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7.5	7.8	8.0
20	3.32	3.8	4.3	4.6	4.9	5.1	5.3	5.5	5.6	5.8	5.9
30	2.63	3.0	3.3	3.4	3.6	3.7	3.8	3.9	4.0	4.1	4.2
60	1.96	2.2	2.3	2.4	2.4	2.5	2.5	2.6	2.6	2.7	2.7
∞	1.00	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

Source: Reprinted, with permission, from H. A. David, "Upper 5 and 1% Points of the Maximum F-Ratio," *Biometrika* 39 (1952), pp. 422–24.

TABLE B.11
Power Values for Analysis of Variance (fixed effects).

$v_1 = 2$ and $\alpha = .05$

v_2	ϕ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.08	.11	.14	.18	.21	.24	.27	.31	.34
2	.12	.20	.30	.41	.52	.62	.71	.79	.85
3	.15	.28	.44	.60	.75	.86	.93	.97	.99
4	.18	.34	.54	.73	.86	.94	.98	.99	1.00
5	.20	.39	.61	.80	.92	.97	.99	1.00	1.00
6	.21	.42	.66	.85	.95	.99	1.00	1.00	1.00
7	.22	.45	.70	.88	.96	.99	1.00	1.00	1.00
8	.23	.47	.72	.90	.97	1.00	1.00	1.00	1.00
9	.24	.49	.74	.91	.98	1.00	1.00	1.00	1.00
10	.25	.50	.76	.92	.98	1.00	1.00	1.00	1.00
12	.26	.53	.78	.93	.99	1.00	1.00	1.00	1.00
14	.27	.54	.80	.94	.99	1.00	1.00	1.00	1.00
16	.27	.55	.81	.95	.99	1.00	1.00	1.00	1.00
18	.28	.56	.82	.95	.99	1.00	1.00	1.00	1.00
20	.28	.57	.83	.96	.99	1.00	1.00	1.00	1.00
22	.29	.58	.83	.96	.99	1.00	1.00	1.00	1.00
24	.29	.58	.84	.96	.99	1.00	1.00	1.00	1.00
26	.29	.59	.84	.96	.99	1.00	1.00	1.00	1.00
28	.29	.59	.85	.96	1.00	1.00	1.00	1.00	1.00
30	.29	.60	.85	.97	1.00	1.00	1.00	1.00	1.00
60	.31	.62	.87	.97	1.00	1.00	1.00	1.00	1.00
120	.31	.63	.88	.98	1.00	1.00	1.00	1.00	1.00
∞	.32	.64	.88	.98	1.00	1.00	1.00	1.00	1.00

$v_1 = 2$ and $\alpha = .01$

v_2	ϕ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.02	.02	.03	.04	.04	.05	.06	.06	.07
2	.02	.04	.07	.10	.14	.18	.22	.27	.32
3	.03	.07	.13	.20	.29	.40	.50	.60	.69
4	.04	.10	.19	.31	.46	.61	.73	.83	.91
5	.05	.13	.25	.42	.60	.75	.87	.94	.97
6	.06	.15	.30	.50	.69	.84	.93	.98	.99
7	.07	.17	.35	.57	.76	.90	.96	.99	1.00
8	.07	.19	.39	.62	.81	.93	.98	1.00	1.00
9	.08	.21	.42	.66	.85	.95	.99	1.00	1.00
10	.08	.22	.45	.69	.87	.96	.99	1.00	1.00
12	.09	.24	.49	.74	.91	.98	1.00	1.00	1.00
14	.09	.26	.52	.78	.93	.98	1.00	1.00	1.00
16	.10	.28	.55	.80	.94	.99	1.00	1.00	1.00
18	.10	.29	.57	.82	.95	.99	1.00	1.00	1.00
20	.10	.30	.58	.83	.95	.99	1.00	1.00	1.00
22	.11	.31	.60	.84	.96	.99	1.00	1.00	1.00
24	.11	.31	.61	.85	.96	.99	1.00	1.00	1.00
26	.11	.32	.62	.86	.97	1.00	1.00	1.00	1.00
28	.11	.32	.63	.86	.97	1.00	1.00	1.00	1.00
30	.11	.33	.63	.87	.97	1.00	1.00	1.00	1.00
60	.13	.36	.68	.90	.98	1.00	1.00	1.00	1.00
120	.13	.38	.70	.91	.98	1.00	1.00	1.00	1.00
∞	.14	.40	.72	.92	.99	1.00	1.00	1.00	1.00

TABLE B.11 *(continued)* **Power Values for Analysis of Variance (fixed effects).**

$v_1 = 3$ and $\alpha = .05$

v_2	ϕ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.08	.10	.13	.16	.19	.22	.25	.28	.31
2	.11	.18	.27	.38	.48	.58	.68	.76	.82
3	.14	.26	.42	.58	.73	.84	.92	.96	.98
4	.17	.33	.53	.72	.86	.94	.98	.99	1.00
5	.19	.38	.61	.81	.93	.98	.99	1.00	1.00
6	.21	.43	.67	.86	.96	.99	1.00	1.00	1.00
7	.22	.46	.72	.89	.97	1.00	1.00	1.00	1.00
8	.24	.49	.75	.92	.98	1.00	1.00	1.00	1.00
9	.25	.51	.77	.93	.99	1.00	1.00	1.00	1.00
10	.25	.53	.79	.94	.99	1.00	1.00	1.00	1.00
12	.27	.56	.82	.96	.99	1.00	1.00	1.00	1.00
14	.28	.58	.84	.97	1.00	1.00	1.00	1.00	1.00
16	.29	.60	.86	.97	1.00	1.00	1.00	1.00	1.00
18	.29	.61	.87	.97	1.00	1.00	1.00	1.00	1.00
20	.30	.62	.87	.98	1.00	1.00	1.00	1.00	1.00
22	.31	.63	.88	.98	1.00	1.00	1.00	1.00	1.00
24	.31	.63	.89	.98	1.00	1.00	1.00	1.00	1.00
26	.31	.64	.89	.98	1.00	1.00	1.00	1.00	1.00
28	.32	.65	.89	.98	1.00	1.00	1.00	1.00	1.00
30	.32	.65	.90	.98	1.00	1.00	1.00	1.00	1.00
60	.34	.68	.92	.99	1.00	1.00	1.00	1.00	1.00
120	.35	.70	.93	.99	1.00	1.00	1.00	1.00	1.00
∞	.36	.71	.93	.99	1.00	1.00	1.00	1.00	1.00

$v_1 = 3$ and $\alpha = .01$

v_2	ϕ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.02	.02	.03	.03	.04	.04	.05	.06	.06
2	.02	.04	.06	.09	.12	.16	.20	.24	.29
3	.03	.06	.12	.19	.27	.37	.47	.58	.67
4	.04	.09	.18	.30	.45	.59	.72	.83	.90
5	.05	.12	.25	.42	.60	.76	.87	.94	.98
6	.06	.15	.31	.51	.71	.86	.94	.98	.99
7	.06	.17	.36	.59	.79	.91	.97	.99	1.00
8	.07	.20	.41	.65	.84	.95	.99	1.00	1.00
9	.08	.22	.45	.70	.88	.97	.99	1.00	1.00
10	.08	.23	.48	.74	.91	.98	1.00	1.00	1.00
12	.09	.26	.54	.79	.94	.99	1.00	1.00	1.00
14	.10	.29	.58	.83	.96	.99	1.00	1.00	1.00
16	.10	.31	.61	.86	.97	1.00	1.00	1.00	1.00
18	.11	.32	.63	.87	.97	1.00	1.00	1.00	1.00
20	.11	.34	.65	.89	.98	1.00	1.00	1.00	1.00
22	.12	.35	.67	.90	.98	1.00	1.00	1.00	1.00
24	.12	.36	.68	.91	.98	1.00	1.00	1.00	1.00
26	.12	.37	.69	.91	.99	1.00	1.00	1.00	1.00
28	.12	.37	.70	.92	.99	1.00	1.00	1.00	1.00
30	.13	.38	.71	.92	.99	1.00	1.00	1.00	1.00
60	.14	.43	.77	.95	.99	1.00	1.00	1.00	1.00
120	.15	.46	.80	.96	1.00	1.00	1.00	1.00	1.00
∞	.16	.48	.82	.97	1.00	1.00	1.00	1.00	1.00

TABLE B.11 (*continued*) **Power Values for Analysis of Variance (fixed effects).**

$v_1 = 4$ and $\alpha = .05$

v_2	ϕ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.08	.10	.13	.15	.18	.21	.24	.27	.29
2	.11	.18	.26	.36	.46	.56	.66	.74	.80
3	.14	.26	.41	.57	.72	.83	.91	.96	.98
4	.17	.33	.53	.72	.86	.94	.98	.99	1.00
5	.19	.39	.62	.81	.93	.98	1.00	1.00	1.00
6	.21	.43	.69	.87	.96	.99	1.00	1.00	1.00
7	.23	.47	.73	.91	.98	1.00	1.00	1.00	1.00
8	.24	.50	.77	.93	.99	1.00	1.00	1.00	1.00
9	.25	.53	.80	.95	.99	1.00	1.00	1.00	1.00
10	.26	.55	.82	.96	.99	1.00	1.00	1.00	1.00
12	.28	.59	.85	.97	1.00	1.00	1.00	1.00	1.00
14	.29	.61	.87	.98	1.00	1.00	1.00	1.00	1.00
16	.30	.63	.89	.98	1.00	1.00	1.00	1.00	1.00
18	.31	.65	.90	.99	1.00	1.00	1.00	1.00	1.00
20	.32	.66	.91	.99	1.00	1.00	1.00	1.00	1.00
22	.33	.67	.91	.99	1.00	1.00	1.00	1.00	1.00
24	.33	.68	.92	.99	1.00	1.00	1.00	1.00	1.00
26	.33	.69	.92	.99	1.00	1.00	1.00	1.00	1.00
28	.34	.69	.93	.99	1.00	1.00	1.00	1.00	1.00
30	.34	.70	.93	.99	1.00	1.00	1.00	1.00	1.00
60	.37	.74	.95	1.00	1.00	1.00	1.00	1.00	1.00
120	.38	.76	.96	1.00	1.00	1.00	1.00	1.00	1.00
∞	.39	.77	.96	1.00	1.00	1.00	1.00	1.00	1.00

$v_1 = 4$ and $\alpha = .01$

v_2	ϕ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.02	.02	.03	.03	.04	.04	.05	.05	.06
2	.02	.04	.06	.08	.12	.15	.19	.23	.28
3	.03	.06	.11	.18	.26	.36	.46	.56	.65
4	.04	.09	.18	.30	.44	.59	.72	.82	.90
5	.05	.12	.25	.42	.60	.76	.88	.94	.98
6	.06	.15	.32	.52	.72	.87	.95	.98	1.00
7	.06	.18	.38	.61	.81	.93	.98	.99	1.00
8	.07	.20	.43	.68	.86	.96	.99	1.00	1.00
9	.08	.23	.47	.73	.90	.97	1.00	1.00	1.00
10	.08	.25	.51	.77	.93	.98	1.00	1.00	1.00
12	.09	.28	.58	.83	.96	.99	1.00	1.00	1.00
14	.10	.31	.62	.87	.97	1.00	1.00	1.00	1.00
16	.11	.34	.66	.89	.98	1.00	1.00	1.00	1.00
18	.12	.36	.69	.91	.99	1.00	1.00	1.00	1.00
20	.12	.37	.71	.92	.99	1.00	1.00	1.00	1.00
22	.13	.39	.73	.93	.99	1.00	1.00	1.00	1.00
24	.13	.40	.74	.94	.99	1.00	1.00	1.00	1.00
26	.13	.41	.76	.95	.99	1.00	1.00	1.00	1.00
28	.14	.42	.77	.95	1.00	1.00	1.00	1.00	1.00
30	.14	.43	.78	.96	1.00	1.00	1.00	1.00	1.00
60	.16	.49	.84	.98	1.00	1.00	1.00	1.00	1.00
120	.17	.53	.86	.98	1.00	1.00	1.00	1.00	1.00
∞	.19	.56	.88	.99	1.00	1.00	1.00	1.00	1.00

TABLE B.11 *(continued)* **Power Values for Analysis of Variance (fixed effects).**

$v_1 = 5$ and $\alpha = .05$

v_2	ϕ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.08	.10	.12	.15	.18	.20	.23	.26	.29
2	.11	.17	.26	.35	.45	.55	.64	.72	.79
3	.14	.25	.40	.56	.71	.83	.91	.95	.98
4	.17	.32	.53	.72	.86	.94	.98	.99	1.00
5	.19	.39	.62	.82	.93	.98	1.00	1.00	1.00
6	.21	.44	.70	.88	.97	.99	1.00	1.00	1.00
7	.23	.48	.75	.92	.98	1.00	1.00	1.00	1.00
8	.24	.52	.79	.94	.99	1.00	1.00	1.00	1.00
9	.26	.55	.82	.96	.99	1.00	1.00	1.00	1.00
10	.27	.57	.84	.97	1.00	1.00	1.00	1.00	1.00
12	.29	.61	.88	.98	1.00	1.00	1.00	1.00	1.00
14	.30	.64	.90	.99	1.00	1.00	1.00	1.00	1.00
16	.32	.66	.91	.99	1.00	1.00	1.00	1.00	1.00
18	.33	.68	.92	.99	1.00	1.00	1.00	1.00	1.00
20	.34	.70	.93	.99	1.00	1.00	1.00	1.00	1.00
22	.34	.71	.94	.99	1.00	1.00	1.00	1.00	1.00
24	.35	.72	.94	1.00	1.00	1.00	1.00	1.00	1.00
26	.36	.73	.95	1.00	1.00	1.00	1.00	1.00	1.00
28	.36	.73	.95	1.00	1.00	1.00	1.00	1.00	1.00
30	.36	.74	.95	1.00	1.00	1.00	1.00	1.00	1.00
60	.40	.78	.97	1.00	1.00	1.00	1.00	1.00	1.00
120	.41	.80	.97	1.00	1.00	1.00	1.00	1.00	1.00
∞	.43	.82	.98	1.00	1.00	1.00	1.00	1.00	1.00

$v_1 = 5$ and $\alpha = .01$

v_2	ϕ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.02	.02	.02	.03	.04	.04	.05	.05	.06
2	.02	.04	.06	.08	.11	.15	.18	.22	.27
3	.03	.06	.11	.18	.26	.35	.45	.55	.64
4	.04	.09	.18	.30	.44	.59	.72	.82	.90
5	.05	.12	.25	.42	.61	.77	.88	.95	.98
6	.06	.15	.32	.53	.73	.88	.95	.99	1.00
7	.06	.18	.39	.63	.82	.93	.98	1.00	1.00
8	.07	.21	.44	.70	.88	.97	.99	1.00	1.00
9	.08	.24	.49	.75	.92	.98	1.00	1.00	1.00
10	.09	.26	.54	.80	.94	.99	1.00	1.00	1.00
12	.10	.30	.61	.86	.97	1.00	1.00	1.00	1.00
14	.11	.34	.66	.90	.98	1.00	1.00	1.00	1.00
16	.12	.36	.70	.92	.99	1.00	1.00	1.00	1.00
18	.12	.39	.73	.94	.99	1.00	1.00	1.00	1.00
20	.13	.41	.76	.95	.99	1.00	1.00	1.00	1.00
22	.14	.43	.78	.96	1.00	1.00	1.00	1.00	1.00
24	.14	.44	.79	.96	1.00	1.00	1.00	1.00	1.00
26	.14	.45	.80	.97	1.00	1.00	1.00	1.00	1.00
28	.15	.46	.82	.97	1.00	1.00	1.00	1.00	1.00
30	.15	.47	.82	.97	1.00	1.00	1.00	1.00	1.00
60	.18	.55	.88	.99	1.00	1.00	1.00	1.00	1.00
120	.20	.59	.91	.99	1.00	1.00	1.00	1.00	1.00
∞	.21	.62	.93	1.00	1.00	1.00	1.00	1.00	1.00

TABLE B.11 (*concluded*) **Power Values for Analysis of Variance (fixed effects).**

$v_1 = 6$ and $\alpha = .05$

v_2	ϕ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.07	.10	.12	.15	.17	.20	.23	.25	.28
2	.10	.17	.25	.34	.44	.54	.63	.71	.78
3	.14	.25	.40	.56	.71	.82	.90	.95	.98
4	.16	.32	.53	.72	.86	.94	.98	.99	1.00
5	.19	.39	.63	.82	.94	.98	1.00	1.00	1.00
6	.21	.44	.70	.89	.97	.99	1.00	1.00	1.00
7	.23	.49	.76	.93	.98	1.00	1.00	1.00	1.00
8	.25	.53	.80	.95	.99	1.00	1.00	1.00	1.00
9	.26	.56	.83	.96	1.00	1.00	1.00	1.00	1.00
10	.28	.59	.86	.97	1.00	1.00	1.00	1.00	1.00
12	.30	.63	.89	.98	1.00	1.00	1.00	1.00	1.00
14	.32	.66	.91	.99	1.00	1.00	1.00	1.00	1.00
16	.33	.69	.93	.99	1.00	1.00	1.00	1.00	1.00
18	.34	.71	.94	.99	1.00	1.00	1.00	1.00	1.00
20	.35	.73	.95	1.00	1.00	1.00	1.00	1.00	1.00
22	.36	.74	.95	1.00	1.00	1.00	1.00	1.00	1.00
24	.37	.75	.96	1.00	1.00	1.00	1.00	1.00	1.00
26	.37	.76	.96	1.00	1.00	1.00	1.00	1.00	1.00
28	.38	.77	.96	1.00	1.00	1.00	1.00	1.00	1.00
30	.39	.77	.97	1.00	1.00	1.00	1.00	1.00	1.00
60	.42	.82	.98	1.00	1.00	1.00	1.00	1.00	1.00
120	.45	.84	.99	1.00	1.00	1.00	1.00	1.00	1.00
∞	.47	.86	.99	1.00	1.00	1.00	1.00	1.00	1.00

$v_1 = 6$ and $\alpha = .01$

v_2	ϕ 1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
1	.01	.02	.02	.03	.04	.04	.05	.05	.06
2	.02	.04	.06	.08	.11	.14	.18	.22	.26
3	.03	.06	.11	.17	.25	.35	.45	.55	.64
4	.04	.09	.18	.30	.44	.59	.72	.82	.90
5	.05	.12	.25	.43	.61	.77	.88	.95	.98
6	.06	.15	.33	.54	.74	.88	.96	.99	1.00
7	.07	.19	.39	.64	.83	.94	.98	1.00	1.00
8	.07	.22	.46	.71	.89	.97	.99	1.00	1.00
9	.08	.24	.51	.77	.93	.98	1.00	1.00	1.00
10	.09	.27	.56	.82	.95	.99	1.00	1.00	1.00
12	.10	.32	.64	.88	.98	1.00	1.00	1.00	1.00
14	.11	.36	.69	.92	.99	1.00	1.00	1.00	1.00
16	.12	.39	.73	.94	.99	1.00	1.00	1.00	1.00
18	.13	.42	.77	.95	1.00	1.00	1.00	1.00	1.00
20	.14	.44	.79	.96	1.00	1.00	1.00	1.00	1.00
22	.14	.46	.81	.97	1.00	1.00	1.00	1.00	1.00
24	.15	.48	.83	.98	1.00	1.00	1.00	1.00	1.00
26	.16	.49	.84	.98	1.00	1.00	1.00	1.00	1.00
28	.16	.50	.85	.98	1.00	1.00	1.00	1.00	1.00
30	.16	.51	.86	.98	1.00	1.00	1.00	1.00	1.00
60	.20	.60	.92	.99	1.00	1.00	1.00	1.00	1.00
120	.22	.65	.94	1.00	1.00	1.00	1.00	1.00	1.00
∞	.24	.69	.96	1.00	1.00	1.00	1.00	1.00	1.00

TABLE B.12 Table for Determining Sample Size for Analysis of Variance (fixed factor levels model).

Power $1 - \beta = .70$

	$\Delta/\sigma = 1.0$				$\Delta/\sigma = 1.25$				$\Delta/\sigma = 1.50$				$\Delta/\sigma = 1.75$				$\Delta/\sigma = 2.0$				$\Delta/\sigma = 2.5$				$\Delta/\sigma = 3.0$			
	α				α				α				α				α				α				α			
r	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	7	11	14	21	5	7	9	15	4	6	7	11	3	4	6	9	3	4	5	7	2	3	4	5	2	3	3	5
3	9	13	17	25	6	9	11	17	5	7	8	12	4	5	7	10	3	4	5	8	3	3	4	6	2	3	3	5
4	11	15	19	28	7	10	13	19	5	7	9	13	4	6	7	10	4	5	6	8	3	4	4	6	2	3	4	5
5	12	17	21	30	8	11	14	20	6	8	10	14	5	6	8	11	4	5	6	9	3	4	5	6	3	3	4	5
6	13	18	22	32	9	12	15	21	6	9	11	15	5	7	8	12	4	5	7	9	3	4	5	7	3	3	4	5
7	14	19	24	34	9	13	16	22	7	9	11	16	5	7	9	12	4	6	7	10	3	4	5	7	3	3	4	5
8	15	20	25	35	10	13	16	23	7	10	12	17	6	7	9	13	5	6	7	10	3	4	5	7	3	3	4	5
9	15	21	26	37	10	14	17	24	7	10	12	17	6	8	9	13	5	6	8	10	3	4	5	7	3	4	4	6
10	16	22	27	38	11	14	18	25	8	10	13	18	6	8	10	14	5	6	8	11	4	5	6	7	3	4	4	6

Power $1 - \beta = .80$

	$\Delta/\sigma = 1.0$				$\Delta/\sigma = 1.25$				$\Delta/\sigma = 1.50$				$\Delta/\sigma = 1.75$				$\Delta/\sigma = 2.0$				$\Delta/\sigma = 2.5$				$\Delta/\sigma = 3.0$			
	α				α				α				α				α				α				α			
r	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	10	14	17	26	7	9	12	17	5	7	9	13	4	5	7	10	3	4	6	8	3	3	4	6	2	3	4	5
3	12	17	21	30	8	11	14	20	6	8	10	14	5	6	8	11	4	5	6	9	3	4	5	7	3	3	4	5
4	14	19	23	33	9	13	15	22	7	9	11	16	5	7	9	12	4	6	7	10	3	4	5	7	3	3	4	5
5	16	21	25	35	10	14	17	23	8	10	12	17	6	8	9	13	5	6	7	10	4	4	5	7	3	4	4	6
6	17	22	27	38	11	15	18	25	8	11	13	18	6	8	10	13	5	7	8	11	4	5	6	8	3	4	4	6
7	18	24	29	39	12	16	19	26	9	11	14	18	7	9	10	14	5	7	8	11	4	5	6	8	3	4	5	6
8	19	25	30	41	12	16	20	27	9	12	14	19	7	9	11	15	6	7	9	12	4	5	6	8	3	4	5	6
9	20	26	31	43	13	17	21	28	9	12	15	20	7	9	11	15	6	7	9	12	4	5	6	8	3	4	5	6
10	21	27	33	44	14	18	21	29	10	13	15	21	8	10	12	16	6	8	9	12	4	5	6	8	3	4	5	6

TABLE B.12 (*concluded*) **Table for Determining Sample Size for Analysis of Variance (fixed factor levels model).**

Power $1-\beta = .90$

	$\Delta/\sigma = 1.0$				$\Delta/\sigma = 1.25$				$\Delta/\sigma = 1.50$				$\Delta/\sigma = 1.75$				$\Delta/\sigma = 2.0$				$\Delta/\sigma = 2.5$				$\Delta/\sigma = 3.0$			
	α				α				α				α				α				α				α			
r	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	14	18	23	32	9	12	15	21	7	9	11	15	5	7	8	12	4	6	7	10	3	4	5	7	3	3	4	6
3	17	22	27	37	11	15	18	24	8	11	13	18	6	8	10	13	5	7	8	11	4	5	6	8	3	4	5	6
4	20	25	30	40	13	16	20	27	9	12	14	19	7	9	11	15	6	7	9	12	4	5	6	8	3	4	5	6
5	21	27	32	43	14	18	21	28	10	13	15	20	8	10	12	15	6	8	9	12	4	5	6	9	4	4	5	7
6	22	29	34	46	15	19	23	30	11	14	16	21	8	10	12	16	7	8	10	13	5	6	7	9	4	4	5	7
7	24	31	36	48	16	20	24	31	11	14	17	22	9	11	13	17	7	9	10	13	5	6	7	9	4	5	5	7
8	26	32	38	50	17	21	25	33	12	15	18	23	9	11	13	17	7	9	11	14	5	6	7	9	4	5	6	7
9	27	33	40	52	17	22	26	34	13	16	18	24	9	12	14	18	8	9	11	14	5	6	8	10	4	5	6	7
10	28	35	41	54	18	23	27	35	13	16	19	25	10	12	14	19	8	10	11	15	5	7	8	10	4	5	6	7

Power $1-\beta = .95$

	$\Delta/\sigma = 1.0$				$\Delta/\sigma = 1.25$				$\Delta/\sigma = 1.50$				$\Delta/\sigma = 1.75$				$\Delta/\sigma = 2.0$				$\Delta/\sigma = 2.5$				$\Delta/\sigma = 3.0$			
	α				α				α				α				α				α				α			
r	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01	.2	.1	.05	.01
2	18	23	27	38	12	15	18	25	9	11	13	18	7	8	10	14	5	7	8	11	4	5	6	8	3	4	5	6
3	22	27	32	43	14	18	21	29	10	13	15	20	8	10	12	16	6	8	9	12	5	6	7	9	4	4	5	7
4	25	30	36	47	16	20	23	31	12	14	17	22	9	11	13	17	7	9	10	13	5	6	7	9	4	5	5	7
5	27	33	39	51	18	22	25	33	13	15	18	23	10	12	14	18	8	9	11	14	5	6	7	10	4	5	6	7
6	29	35	41	53	19	23	27	35	13	16	19	25	10	12	14	19	8	10	11	15	6	7	8	10	4	5	6	8
7	30	37	43	56	20	24	28	36	14	17	20	26	11	13	15	19	8	10	12	15	6	7	8	10	4	5	6	8
8	32	39	45	58	21	25	29	38	15	18	21	27	11	14	16	20	9	11	12	16	6	7	8	11	5	5	6	8
9	33	40	47	60	22	26	30	39	15	19	22	28	12	14	16	21	9	11	13	16	6	8	9	11	5	6	6	8
10	34	42	48	62	22	27	31	40	16	19	22	29	12	15	17	21	9	11	13	17	6	8	9	11	5	6	7	8

Source: Reprinted, with permission, from T. L. Bratcher, M. A. Moran, and W. J. Zimmer, "Tables of Sample Sizes in the Analysis of Variance," *Journal of Quality Technology* 2 (1970), pp. 156–64. Copyright American Society for Quality Control, Inc.

TABLE B.13 **Table of $\lambda\sqrt{n}/\sigma$ for Determining Sample Size to Find "Best" of r Population Means.**

Number of Populations (r)	Probability of Correct Identification ($1-\alpha$)		
	.90	.95	.99
2	1.8124	2.3262	3.2900
3	2.2302	2.7101	3.6173
4	2.4516	2.9162	3.7970
5	2.5997	3.0552	3.9196
6	2.7100	3.1591	4.0121
7	2.7972	3.2417	4.0861
8	2.8691	3.3099	4.1475
9	2.9301	3.3679	4.1999
10	2.9829	3.4182	4.2456

Source: Reprinted, with permission, from R. E. Bechhofer, "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances," *The Annals of Mathematical Statistics* 25 (1954), pp. 16–39.

TABLE B.14 **Selected Standard Latin Squares.**

3 × 3

A	B	C
B	C	A
C	A	B

4 × 4

1

A	B	C	D
B	A	D	C
C	D	B	A
D	C	A	B

2

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

3

A	B	C	D
B	D	A	C
C	A	D	B
D	C	B	A

4

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

5 × 5

A	B	C	D	E
B	A	E	C	D
C	D	A	E	B
D	E	B	A	C
E	C	D	B	A

6 × 6

A	B	C	D	E	F
B	F	D	C	A	E
C	D	E	F	B	A
D	A	F	E	C	B
E	C	A	B	F	D
F	E	B	A	D	C

7 × 7

A	B	C	D	E	F	G
B	C	D	E	F	G	A
C	D	E	F	G	A	B
D	E	F	G	A	B	C
E	F	G	A	B	C	D
F	G	A	B	C	D	E
G	A	B	C	D	E	F

8 × 8

A	B	C	D	E	F	G	H
B	C	D	E	F	G	H	A
C	D	E	F	G	H	A	B
D	E	F	G	H	A	B	C
E	F	G	H	A	B	C	D
F	G	H	A	B	C	D	E
G	H	A	B	C	D	E	F
H	A	B	C	D	E	F	G

9 × 9

A	B	C	D	E	F	G	H	I
B	C	D	E	F	G	H	I	A
C	D	E	F	G	H	I	A	B
D	E	F	G	H	I	A	B	C
E	F	G	H	I	A	B	C	D
F	G	H	I	A	B	C	D	E
G	H	I	A	B	C	D	E	F
H	I	A	B	C	D	E	F	G
I	A	B	C	D	E	F	G	H

TABLE B.15 Selected Balanced Incomplete Block Designs.

Design 1: $r=4$, $r_b=2$, $n_b=6$, $n=3$, $n_p=1$

Block	Treatments
1	1 2
2	3 4
3	1 3
4	2 4
5	1 4
6	2 3

Design 2: $r=4$, $r_b=3$, $n_b=4$, $n=3$, $n_p=2$

Block	Treatments
1	1 2 3
2	1 2 4
3	1 3 4
4	2 3 4

Design 3: $r=5$, $r_b=2$, $n_b=10$, $n=4$, $n_p=1$

Block	Treatments
1	1 2
2	3 4
3	2 5
4	1 3
5	4 5
6	1 4
7	2 3
8	3 5
9	1 5
10	2 4

Design 4: $r=5$, $r_b=3$, $n_b=10$, $n=6$, $n_p=3$

Block	Treatments
1	1 2 3
2	1 2 5
3	1 4 5
4	2 3 4
5	3 4 5
6	1 2 4
7	1 3 4
8	1 3 5
9	2 3 5
10	2 4 5

Design 5: $r=5$, $r_b=4$, $n_b=5$, $n=4$, $n_p=3$

Block	Treatments
1	1 2 3 4
2	1 2 3 5
3	1 2 4 5
4	1 3 4 5
5	2 3 4 5

Design 6: $r=6$, $r_b=2$, $n_b=15$, $n=5$, $n_p=1$

Block	Treatments
1	1 2
2	3 4
3	5 6
4	1 3
5	2 5
6	4 6
7	1 4
8	2 6
9	3 5
10	1 5
11	2 4
12	3 6
13	1 6
14	2 3
15	4 5

Design 7: $r=6$, $r_b=3$, $n_b=10$, $n=5$, $n_p=2$

Block	Treatments
1	1 2 5
2	1 2 6
3	1 3 4
4	1 3 6
5	1 4 5
6	2 3 4
7	2 3 5
8	2 4 6
9	3 5 6
10	4 5 6

Design 8: $r=6$, $r_b=3$, $n_b=20$, $n=10$, $n_p=4$

Block	Treatments
1	1 2 3
2	4 5 6
3	1 2 4
4	3 5 6
5	1 2 5
6	3 4 6
7	1 2 6
8	3 4 5
9	1 3 4
10	2 5 6
11	1 3 5
12	2 4 6
13	1 3 6
14	2 4 5
15	1 4 5
16	2 3 6
17	1 4 6
18	2 3 5
19	1 5 6
20	2 3 4

TABLE B.15 (*continued*) **Selected Balanced Incomplete Block Designs.**

Design 9: $r = 6$, $r_b = 4$, $n_b = 15$, $n = 10$, $n_p = 6$

Block	Treatments			
1	1	2	3	4
2	1	4	5	6
3	2	3	5	6
4	1	2	3	5
5	1	2	4	6
6	3	4	5	6
7	1	2	3	6
8	1	3	4	5
9	2	4	5	6
10	1	2	4	5
11	1	3	5	6
12	2	3	4	6
13	1	2	5	6
14	1	3	4	6
15	2	3	4	5

Design 10: $r = 6$, $r_b = 5$, $n_b = 6$, $n = 5$, $n_p = 4$

Block	Treatments				
1	1	2	3	4	5
2	1	2	3	4	6
3	1	2	3	5	6
4	1	2	4	5	6
5	1	3	4	5	6
6	2	3	4	5	6

Design 11: $r = 7$, $r_b = 2$, $n_b = 21$, $n = 6$, $n_p = 1$

Block	Treatments	
1	1	2
2	2	6
3	3	4
4	4	7
5	1	5
6	5	6
7	3	7
8	1	3
9	2	4
10	3	5
11	4	6
12	5	7
13	1	6
14	2	7
15	1	4
16	2	3
17	3	6
18	4	5
19	2	5
20	6	7
21	1	7

Design 12: $r = 7$, $r_b = 3$, $n_b = 7$, $n = 3$, $n_p = 1$

Block	Treatments		
1	1	2	4
2	2	3	5
3	3	4	6
4	4	5	7
5	5	6	1
6	6	7	2
7	7	1	3

Design 13: $r = 7$, $r_b = 4$, $n_b = 7$, $n = 4$, $n_p = 2$

Block	Treatments			
1	3	5	6	7
2	1	4	6	7
3	1	2	5	7
4	1	2	3	6
5	2	3	4	7
6	1	3	4	5
7	2	4	5	6

Design 14: $r = 7$, $r_b = 6$, $n_b = 7$, $n = 6$, $n_p = 5$

Block	Treatments					
1	1	2	3	4	5	6
2	1	2	3	4	5	7
3	1	2	3	4	6	7
4	1	2	3	5	6	7
5	1	2	4	5	6	7
6	1	3	4	5	6	7
7	2	3	4	5	6	7

TABLE B.15 (*continued*) **Selected Balanced Incomplete Block Designs.**

Design 15: $r = 8$, $r_b = 2$, $n_b = 28$, $n = 7$, $n_p = 1$

Block	Treatments	
1	1	2
2	3	4
3	5	6
4	7	8
5	1	3
6	2	8
7	4	5
8	6	7
9	1	4
10	2	7
11	3	6
12	5	8
13	1	5
14	2	3
15	4	7
16	6	8
17	1	6
18	2	4
19	3	8
20	5	7
21	1	7
22	2	6
23	3	5
24	4	8
25	1	8
26	2	5
27	3	7
28	4	6

Design 16: $r = 8$, $r_b = 4$, $n_b = 14$, $n = 7$, $n_p = 3$

Block	Treatments			
1	1	2	3	4
2	5	6	7	8
3	1	2	7	8
4	3	4	5	6
5	1	3	6	8
6	2	4	5	7
7	1	4	6	7
8	2	3	5	8
9	1	2	5	6
10	3	4	7	8
11	1	3	5	7
12	2	4	6	8
13	1	4	5	8
14	2	3	6	7

Design 17: $r = 8$, $r_b = 7$, $n_b = 8$, $n = 7$, $n_p = 6$

Block	Treatments						
1	1	2	3	4	5	6	7
2	1	2	3	4	5	6	8
3	1	2	3	4	5	7	8
4	1	2	3	4	6	7	8
5	1	2	3	5	6	7	8
6	1	2	4	5	6	7	8
7	1	3	4	5	6	7	8
8	2	3	4	5	6	7	8

Design 18: $r = 9$, $r_b = 3$, $n_b = 12$, $n = 4$, $n_p = 1$

Block	Treatments		
1	1	2	3
2	1	4	7
3	1	5	9
4	1	6	8
5	2	4	9
6	2	5	8
7	2	6	7
8	3	4	8
9	3	5	7
10	3	6	9
11	4	5	6
12	7	8	9

Appendix C

Data Sets

Data Set C.1 SENIC

The primary objective of the Study on the Efficacy of Nosocomial Infection Control (**SENIC** Project) was to determine whether infection surveillance and control programs have reduced the rates of nosocomial (hospital-acquired) infection in United States hospitals. This data set consists of a random sample of 113 hospitals selected from the original 338 hospitals surveyed.

Each line of the data set has an identification number and provides information on 11 other variables for a single hospital. The data presented here are for the 1975–76 study period. The 12 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–113
2	Length of stay	Average length of stay of all patients in hospital (in days)
3	Age	Average age of patients (in years)
4	Infection risk	Average estimated probability of acquiring infection in hospital (in percent)
5	Routine culturing ratio	Ratio of number of cultures performed to number of patients without signs or symptoms of hospital-acquired infection, times 100
6	Routine chest X-ray ratio	Ratio of number of X-rays performed to number of patients without signs or symptoms of pneumonia, times 100
7	Number of beds	Average number of beds in hospital during study period
8	Medical school affiliation	1 = Yes, 2 = No
9	Region	Geographic region, where: 1 = NE, 2 = NC, 3 = S, 4 = W
10	Average daily census	Average number of patients in hospital per day during study period
11	Number of nurses	Average number of full-time equivalent registered and licensed practical nurses during study period (number full time plus one half the number part time)
12	Available facilities and services	Percent of 35 potential facilities and services that are provided by the hospital

Reference: Special Issue, "The SENIC Project," *American Journal of Epidemiology* **111** (1980), pp. 465–653. Data obtained from Robert W. Haley, M.D., Hospital Infections Program, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia 30333.

1	2	3	4	5	6	7	8	9	10	11	12
1	7.13	55.7	4.1	9.0	39.6	279	2	4	207	241	60.0
2	8.82	58.2	1.6	3.8	51.7	80	2	2	51	52	40.0
3	8.34	56.9	2.7	8.1	74.0	107	2	3	82	54	20.0
...	...	...	...	...	...	...	...	...	...	...	...
111	7.70	56.9	4.4	12.2	67.9	129	2	4	85	136	62.9
112	17.94	56.2	5.9	26.4	91.8	835	1	1	791	407	62.9
113	9.41	59.5	3.1	20.6	91.7	29	2	3	20	22	22.9

Data Set C.2 CDI

This data set provides selected county demographic information (CDI) for 440 of the most populous counties in the United States. Each line of the data set has an identification number with a county name and state abbreviation and provides information on 14 variables for a single county. Counties with missing data were deleted from the data set. The information generally pertains to the years 1990 and 1992. The 17 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–440
2	County	County name
3	State	Two-letter state abbreviation
4	Land area	Land area (square miles)
5	Total population	Estimated 1990 population
6	Percent of population aged 18–34	Percent of 1990 CDI population aged 18–34
7	Percent of population 65 or older	Percent of 1990 CDI population aged 65 years old or older
8	Number of active physicians	Number of professionally active nonfederal physicians during 1990
9	Number of hospital beds	Total number of beds, cribs, and bassinets during 1990
10	Total serious crimes	Total number of serious crimes in 1990, including murder, rape, robbery, aggravated assault, burglary, larceny-theft, and motor vehicle theft, as reported by law enforcement agencies
11	Percent high school graduates	Percent of adult population (persons 25 years old or older) who completed 12 or more years of school
12	Percent bachelor's degrees	Percent of adult population (persons 25 years old or older) with bachelor's degree
13	Percent below poverty level	Percent of 1990 CDI population with income below poverty level
14	Percent unemployment	Percent of 1990 CDI labor force that is unemployed
15	Per capita income	Per capita income of 1990 CDI population (dollars)
16	Total personal income	Total personal income of 1990 CDI population (in millions of dollars)
17	Geographic region	Geographic region classification is that used by the U.S. Bureau of the Census, where: 1 = NE, 2 = NC, 3 = S, 4 = W

Source: Geospatial and Statistical Data Center, University of Virginia.

1	2	3	4	5	6	7	8	9	10
1	Los_Angeles	CA	4060	8863164	32.1	9.7	23677	27700	688936
2	Cook	IL	946	5105067	29.2	12.4	15153	21550	436936
3	Harris	TX	1729	2818199	31.3	7.1	7553	12449	253526
...	...	...	...	...	...	...	...	...	...
438	Montgomery	TN	539	100498	35.7	7.9	87	188	6537
439	Maui	HI	1159	100374	26.2	11.3	192	182	7130
440	Morgan	AL	582	100043	26.3	11.7	122	464	4693

11	12	13	14	15	16	17
70.0	22.3	11.6	8.0	20786	184230	4
73.4	22.8	11.1	7.2	21729	110928	2
74.9	25.4	12.5	5.7	19517	55003	3
...	...	...	...	...	...	...
77.9	16.5	10.8	8.0	13169	1323	3
77.0	17.8	5.7	3.2	18504	1857	4
69.4	15.5	9.4	7.1	16458	1647	3

Data Set C.3 Market Share

Company executives from a large packaged foods manufacturer wished to determine which factors influence the market share of one of its products. Data were collected from a national database (Nielsen) for 36 consecutive months. Each line of the data set has an identification number and provides information on 6 other variables for each month. The data presented here are for September, 1999, through August, 2002. The variables are:

Variable Number	Variable Name	Description
1	Identification number	1–36
2	Market share	Average monthly market share for product (percent)
3	Price	Average monthly price of product (dollars)
4	Gross Nielsen rating points	An index of the amount of advertising exposure that the product received
5	Discount price	Presence or absence of discount price during period: 1 if discount, 0 otherwise
6	Package promotion	Presence or absence of package promotion during period: 1 if promotion present, 0 otherwise
7	Month	Month (Jan–Dec)
8	Year	Year (1999–2002)

1	2	3	4	5	6	7	8
1	3.15	2.198	498	1	1	Sep	1999
2	2.52	2.186	510	0	0	Oct	1999
3	2.64	2.293	422	1	1	Nov	1999
...	...	...	...	...	...	...	...
34	2.80	2.518	270	1	0	Jun	2002
35	2.48	2.497	322	0	1	Jul	2002
36	2.85	2.781	317	1	1	Aug	2002

Data Set C.4 University Admissions

The director of admissions at a state university wanted to determine how accurately students' grade-point averages at the end of their freshman year could be predicted by entrance test scores and high school class rank. The academic years cover 1996 through 2000. Each line of the data set has an identification number and information on 4 other variables for each student. The 5 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–705
2	GPA	Grade-point average following freshman year
3	High school class rank	High school class rank as percentile: lower percentiles imply higher class ranks
4	ACT score	ACT entrance examination score
5	Academic year	Calendar year that freshman entered university

1	2	3	4	5
1	0.980	61	20	1996
2	1.130	84	20	1996
3	1.250	74	19	1996
...	...	...	...	...
703	4.000	97	29	2000
704	4.000	97	29	2000
705	4.000	99	32	2000

Data Set C.5 Prostate Cancer

A university medical center urology group was interested in the association between prostate-specific antigen (PSA) and a number of prognostic clinical measurements in men with advanced prostate cancer. Data were collected on 97 men who were about to undergo radical prostectomies. Each line of the data set has an identification number and provides information on 8 other variables for each person. The 9 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–97
2	PSA level	Serum prostate-specific antigen level (mg/ml)
3	Cancer volume	Estimate of prostate cancer volume (cc)
4	Weight	Prostate weight (gm)
5	Age	Age of patient (years)
6	Benign prostatic hyperplasia	Amount of benign prostatic hyperplasia (cm^2)
7	Seminal vesicle invasion	Presence or absence of seminal vesicle invasion: 1 if yes; 0 otherwise
8	Capsular penetration	Degree of capsular penetration (cm)
9	Gleason score	Pathologically determined grade of disease using total score of two patterns (summed scores were either 6, 7, or 8 with higher scores indicating worse prognosis)

1	2	3	4	5	6	7	8	9
1	0.651	0.5599	15.959	50	0	0	0	6
2	0.852	0.3716	27.660	58	0	0	0	7
3	0.852	0.6005	14.732	74	0	0	0	7
...	...	...	...	...	...	...	...	...
95	170.716	18.3568	29.964	52	0	1	11.7048	8
96	239.847	17.8143	43.380	68	4.7588	1	4.7588	8
97	265.072	32.1367	52.985	68	1.5527	1	18.1741	8

Adapted in part from: Hastie, T. J.; R. J. Tibshirani; and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001.

Data Set C.6 Website Developer

Management of a company that develops websites was interested in determining which variables have the greatest impact on the number of websites developed and delivered to customers per quarter. Data were collected on website production output for 13 three-person website development teams, from January 2001 through August 2002. Each line of the data set has an identification number and provides information on 6 other variables for thirteen teams over time. The 8 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–73
2	Websites delivered	Number of websites completed and delivered to customers during the quarter
3	Backlog of orders	Number of website orders in backlog at the close of the quarter
4	Team number	1–13
5	Team experience	Number of months team has been together
6	Process change	A change in the website development process occurred during the second quarter of 2002: 1 if quarter 2 or 3, 2002; 0 otherwise
7	Year	2001 or 2002
8	Quarter	1, 2, 3, or 4

1	2	3	4	5	6	7	8
1	1	12	1	3	0	2001	1
2	2	18	1	6	0	2001	2
3	7	26	1	9	0	2001	3
...	...	...	...	...	...	...	...
71	7	36	13	14	0	2002	1
72	19	37	13	17	1	2002	2
73	12	26	13	20	1	2002	3

Data Set C.7 Real Estate Sales

The city tax assessor was interested in predicting residential home sales prices in a midwestern city as a function of various characteristics of the home and surrounding property. Data on 522 arms-length transactions were obtained for home sales during the year 2002. Each line of the data set has an identification number and provides information on 12 other variables. The 13 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–522
2	Sales price	Sales price of residence (dollars)
3	Finished square feet	Finished area of residence (square feet)
4	Number of bedrooms	Total number of bedrooms in residence
5	Number of bathrooms	Total number of bathrooms in residence
6	Air conditioning	Presence or absence of air conditioning: 1 if yes; 0 otherwise
7	Garage size	Number of cars that garage will hold
8	Pool	Presence or absence of swimming pool: 1 if yes; 0 otherwise
9	Year built	Year property was originally constructed
10	Quality	Index for quality of construction: 1 indicates high quality; 2 indicates medium quality; 3 indicates low quality
11	Style	Qualitative indicator of architectural style
12	Lot size	Lot size (square feet)
13	Adjacent to highway	Presence or absence of adjacency to highway: 1 if yes; 0 otherwise

1	2	3	4	5	6	7	8	9	10	11	12	13
1	360000	3032	4	4	1	2	0	1972	2	1	22221	0
2	340000	2058	4	2	1	2	0	1976	2	1	22912	0
3	250000	1780	4	3	1	2	0	1980	2	1	21345	0
...	...	...	...	...	...	...	...	...	...	...	...	...
520	133500	1922	3	1	0	2	0	1950	3	1	14805	0
521	124000	1480	3	2	1	2	0	1953	3	1	28351	0
522	95500	1184	2	1	0	1	0	1951	3	1	14786	0

Data Set C.8 Heating Equipment

A manufacturer of heating equipment was interested in forecasting the volume of monthly orders as a function of various economic indicators, supply-chain factors, and weather in a particular sales region. Data by month over a four-year period (1999–2002) for this region were available for analysis. Each line of the data set has an identification number and provides information on 9 other variables. The 10 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–43
2	Number of orders	Number of heating equipment orders during month
3	Interest rate	Prime rate in effect during month
4	New homes	Number of new homes completed and for sale in sales region during month
5	Discount	Percent discount (0–5) offered to distributors during month; value is usually 0, indicating no discount
6	Inventories	Distributor inventories in warehouses during month
7	Sell through	Number of units sold by distributor to contractors in previous month
8	Temperature deviation	Difference between average temperature for month and 30-year average for that month
9	Year	1999, 2000, 2001, or 2002
10	Month	Coded 1–12

1	2	3	4	5	6	7	8	9	10
1	121	0.0750	64	0	3536	615	2.22	1999	1
2	227	0.0750	64	0	3042	813	0.28	1999	2
3	446	0.0750	65	0	2456	704	0.79	1999	3
...	...	...	...	...	...	...	...	...	...
41	754	0.0475	64	0	1417	927	0.81	2002	6
42	1098	0.0475	65	0	1244	877	0.28	2002	7
43	1158	0.0475	65	0	1465	809	0.50	2002	8

Data Set C.9 Ischemic Heart Disease

A health insurance company collected information on 788 of its subscribers who had made claims resulting from ischemic (coronary) heart disease. Data were obtained on total costs of services provided for these 788 subscribers and the nature of the various services for the period of January 1, 1998 through December 31, 1999. Each line in the data set has an identification number and provides information on 9 other variables for each subscriber. The 10 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–788
2	Total cost	Total cost of claims by subscriber (dollars)
3	Age	Age of subscriber (years)
4	Gender	Gender of subscriber: 1 if male; 0 otherwise
5	Interventions	Total number of interventions or procedures carried out
6	Drugs	Number of tracked drugs prescribed
7	Emergency room visits	Number of emergency room visits
8	Complications	Number of other complications that arose during heart disease treatment
9	Comorbidities	Number of other diseases that the subscriber had during period
10	Duration	Number of days of duration of treatment condition

1	2	3	4	5	6	7	8	9	10
1	179.1	63	0	2	1	4	0	3	300
2	319.0	59	0	2	0	6	0	0	120
3	9310.7	62	0	17	0	2	0	5	353
...	...	...	...	...	...	...	...	...	...
786	2677.7	68	0	3	2	6	0	10	303
787	1282.2	58	0	7	2	2	0	7	244
788	586.0	56	0	4	4	6	0	3	336

Data Set C.10 Disease Outbreak

This data set provides information from a study based on 196 persons selected in a probability sample within two sectors in a city. Each line of the data set has an identification number and provides information on 5 other variables for a single person. The 6 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–196
2	Age	Age of person (in years)
3	Socioeconomic status	1 = upper, 2 = middle, 3 = lower
4	Sector	Sector within city, where: 1 = sector 1, 2 = sector 2
5	Disease status	1 = with disease, 0 = without disease
6	Savings account status	1 = has savings account, 0 = does not have savings account

Adapted in part from H. G. Dantes, J. S. Koopman, C. L. Addy, et al., "Dengue Epidemics on the Pacific Coast of Mexico," *International Journal of Epidemiology* 17 (1988), pp. 178–86.

1	2	3	4	5	6
1	33	1	1	0	1
2	35	1	1	0	1
3	6	1	1	0	0
...	...	...	...	...	...
194	31	3	1	0	0
195	85	3	1	0	1
196	24	2	1	0	0

Data Set C.11 IPO

Private companies often go public by issuing shares of stock referred to as initial public offerings (IPOs). A study of 482 IPOs was conducted to determine what are the characteristics of companies that attract venture capital funding. The response of interest is whether or not a company was financed with venture capital funds. Potential predictors include the face value of the company, the number of shares offered, and whether or not the company

underwent a leveraged buyout. Each line of the data set has an identification number and provides information on 4 other variables for a single person. The 5 variables are:

Variable Number	Variable Name	Description
1	Identification number	1–482
2	Venture capital funding	Presence or absence of venture capital funding: 1 if yes; 0 otherwise
3	Face value of company	Estimated face value of company from prospectus (in dollars)
4	Number of shares offered	Total number of shares offered
5	Leveraged buyout	Presence or absence of leveraged buyout: 1 if yes; 0 otherwise

1	2	3	4	5
1	0	1,200,000	3,000,000	0
2	0	1,454,000	1,454,000	1
3	0	1,500,000	300,000	0
...	...	...	...	...
480	0	159,500,000	7,250,000	0
481	0	165,000,000	11,000,000	0
482	0	234,600,000	9,200,000	0

Data Set C.12 Drug Effect Experiment

This data set provides results adapted from an experiment in which the effects of a drug on the behavior of rats were studied. The behavior under consideration was the rate at which a rat deprived of water presses a lever to obtain water. The experiment was carried out in two parts. Variable 2 identifies the two parts of the study (1, 2).

In Part I of the study, 12 male albino rats of the same strain and approximately the same weight were utilized. Variable 3 identifies each rat (1, . . . , 12). Prior to the experiment, each rat was trained to press a lever for water until a stable rate of pressing was reached. Two factors were studied in this experiment—initial lever press rate (factor A) and dosage of the drug (factor B). The 12 rats were classified into one of three groups according to their initial lever press rate. Variable 4 identifies the level of the initial lever press rate (1, 2, 3). Level 1 is a slow rate, level 2 a moderate rate, and level 3 a fast rate. The levels were defined such that one third of the rats were classified into each of the three levels.

Four dosage levels of the drug were studied, including a zero level consisting of a saline solution. Variable 5 identifies the drug dosage (1, . . . , 4). All dosage levels were specified in terms of milligrams of drug per kilogram of weight of the rat.

One hour after a drug dosage injection was administered, an experimental session began during which the rat received water each time after the second lever press. This reinforcement schedule will be denoted by FR-2. Each rat received all four drug dosage levels in a random order. Each of the four drug dosages was administered twice, thus providing two observation units for each treatment. Variable 6 identifies the observation unit (1, 2).

The response variable was defined as the total number of lever presses divided by the elapsed time (in seconds) during a session for the given treatment. Variable 7 is the response variable.

In Part II of the study, another 12 albino male rats of the same strain and approximately the same weight as the rats used in Part I were used. Variable 2 identifies this part of the study, and variable 3 identifies the 12 additional rats (13, . . . , 24). The experimental design for Part II of the study was exactly the same as for Part I, except that each rat received water each time after the fifth lever press. This reinforcement schedule will be denoted by FR-5. Variable 2 identifies the reinforcement schedule since Part I of the study used schedule FR-2 while Part II of the study used schedule FR-5. The reinforcement schedule thus is another factor (factor *C*) that was studied in the combined experiment.

To summarize, the variables for this experimental design are:

Variable Number	Variable Name	Description
1	Identification number	1–192
2	Part of study (factor *C*: reinforcement schedule)	1:Part I (FR-2) 2:Part II (FR-5)
3	Rat identification	1–24
4	Initial lever press rate (factor *A*)	1:Slow 2:Moderate 3:Fast
5	Dosage level (mg/kg) (factor *B*)	1:0 (saline solution) 2:.5 3:1.0 4:1.8
6	Observation unit	1, 2
7	Response variable—lever press rate	Total number of lever presses divided by elapsed time in seconds

Reference: T. G. Heffner; R. B. Drawbaugh; and M. J. Zigmond. "Amphetamine and Operant Behavior in Rats: Relationship between Drug Effect and Control Response Rate," *Journal of Comparative and Physiological Psychology* 86 (1974), pp. 1031–43.

1	2	3	4	5	6	7
1	1	1	1	1	1	.81
2	1	1	1	2	1	.80
3	1	1	1	3	1	.82
. . .	. . .	. . .	. . .	. . .	. . .	. . .
190	2	24	3	2	2	2.98
191	2	24	3	3	2	2.47
192	2	24	3	4	2	1.51

Appendix D

Rules for Developing ANOVA Models and Tables for Balanced Designs

In this appendix, we present and illustrate rules for developing models for nested and/or crossed factor designs, for finding the appropriate sums of squares and degrees of freedom for the needed mean squares, and for finding the expected values of the mean squares. The rules in Sections D.1–D.3 apply to all balanced designs with two or more replications and with no interactions assumed to equal zero. The rule modifications in Section D.4 show how these rules need to be modified to make them applicable to balanced designs with no replications and/or with some interaction terms assumed to equal zero.

As noted earlier, a design is *balanced* in the nested case when (1) the number of factor levels of a nested factor is the same for each level of the factor in which the nesting takes place, and (2) the number of replications is constant for the different factor level combinations. In the crossed case, a design is balanced whenever the number of replications is constant for all factor level combinations. In a subsampling design, balance requires that the subsample sizes at each stage of sampling be constant.

D.1 Rule for Model Development

We begin by presenting a rule for the development of a nested and/or crossed factor design model. *This rule is applicable when no interactions are assumed to equal zero.* We shall utilize as an illustration the training school example of Table 26.1, where the effects of three schools (factor A) and two instructors within each school (factor B) were studied and two replications were made in each instance.

Rule (D.1)

Step 1. *Include an overall constant and a main effect term for each factor, taking into account when one factor is nested within another.*

Example For the training school example, we include:

$$\mu_{..} \qquad \alpha_i \qquad \beta_{j(i)}$$

Note that factor B is nested within factor A.

Step 2. *Include all interaction terms except those containing both a nested factor and the factor within which it is nested.*

Example Since factor B is nested within factor A, the AB interaction term (the only possible interaction term here) is not included.

Step 3. *Interactions between a nested factor and another factor with which the nested factor is crossed are always themselves nested.*

Example For the training school example, this situation does not arise.

Step 4. *Include the error term, which is nested within all factors.*

Since the model formulation will be used for developing the needed ANOVA sums of squares, degrees of freedom, and expected mean squares, we now need to recognize that the error term ε is nested within a factor level combination. That is, the kth experimental unit when factor A is at level 1 and factor B is at level 1 is not the same unit as the kth experimental unit for another factor level combination.

Example For the training school example the error term is $\varepsilon_{k(ij)}$, and the appropriate model therefore is:

$$\begin{gathered} Y_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \varepsilon_{k(ij)} \\ i = 1, 2, 3; \quad j = 1, 2; \quad k = 1, 2 \end{gathered} \tag{D.2}$$

D.2 Rule for Finding Sums of Squares and Degrees of Freedom

This rule is applicable to all balanced designs with two or more replications and with no interaction terms assumed to equal zero. We shall continue to consider the training school example where factor B is nested within factor A. It does not matter for this rule whether the factor effects are fixed or random.

Rule (D.3) for Definitional Forms of Sums of Squares and Degrees of Freedom

Step 1. *Write the model equation.*

Example The model equation for the training school example was given earlier. We show this model now in its general form, where factor A has a levels, factor B has b levels, and there are n replications:

$$\begin{gathered} Y_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \varepsilon_{k(ij)} \\ i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, n \end{gathered} \tag{D.2a}$$

Step 2. *For each model term other than the overall constant, write the associated SS notation.*

Example We do this for the training school example in columns 1 and 2 of Table D.1 for α_i, $\beta_{j(i)}$, and $\varepsilon_{k(ij)}$. The line for Total will not be completed until step 9.

Step 3. *Each sum of squares will have as coefficient the product of the limits of the subscripts not appearing in the model term. The coefficient is taken to be* 1 *if all subscripts appear in the model term.*

TABLE D.1 Derivation of Sums of Squares Formulas for Nested Two-Factor Experiment (B nested within A).

(1) Model Term	(2) SS	(3) Coefficient	(4) $\sum$	(5) Symbolic Product	(6) Term to Be Squared	(7) Sum of Squares	(8) Degrees of Freedom
α_i	SSA	bn	$\sum_i$	$i-1$	$\bar{Y}_{i..}-\bar{Y}_{...}$	$bn\sum_i(\bar{Y}_{i..}-\bar{Y}_{...})^2$	$a-1$
$\beta_{j(i)}$	$SSB(A)$	n	$\sum_i\sum_j$	$i(j-1)=ij-i$	$\bar{Y}_{ij.}-\bar{Y}_{i..}$	$n\sum_i\sum_j(\bar{Y}_{ij.}-\bar{Y}_{i..})^2$	$a(b-1)$
$\varepsilon_{k(ij)}$	SSE	1	$\sum_i\sum_j\sum_k$	$(k-1)ij=ijk-ij$	$Y_{ijk}-\bar{Y}_{ij.}$	$\sum_i\sum_j\sum_k(Y_{ijk}-\bar{Y}_{ij.})^2$	$ab(n-1)$
Total	$SSTO$				$Y_{ijk}-\bar{Y}_{...}$	$\sum_i\sum_j\sum_k(Y_{ijk}-\bar{Y}_{...})^2$	$abn-1$

Example The coefficients for our example are shown in column 3 of Table D.1. For instance, α_i does not contain j and k. These subscripts have limits of b and n, respectively. The coefficient for the SSA term is therefore bn. Since the model term $\varepsilon_{k(ij)}$ contains all subscripts, the coefficient is taken to be 1 here.

Step 4. *Each sum of squares is summed over all of the subscripts of the model term, whether in parentheses or not.*

Example The summations for our example are shown in column 4. For instance, the sum of squares term corresponding to α_i is summed over i, the only subscript in that model term. Similarly, the sum of squares term corresponding to $\varepsilon_{k(ij)}$ is summed over i, j, and k since all of these appear in the model term.

Step 5. *Form a symbolic product from the subscripts of the model term, using the subscript if it is in parentheses, and the subscript minus* 1 *if it is not in parentheses. Expand the product.*

Example The symbolic products for our example are shown in column 5. For instance, for α_i the symbolic product is $i-1$. For $\beta_{j(i)}$, the symbolic product is $i(j-1)=ij-i$. For $\varepsilon_{k(ij)}$, the symbolic product is $(k-1)ij=ijk-ij$.

Step 6. *The typical term to be squared consists of means of the observations with the subscripts consisting of the symbolic product term and dots elsewhere. The sign of each mean is that of the symbolic product. A* 1 *refers to the overall mean.*

Example The terms to be squared for our example are shown in column 6. Note that for α_i, the symbolic product is $i-1$, and the typical term to be squared therefore is:

$$\bar{Y}_{i..}-\bar{Y}_{...}$$

For $\beta_{j(i)}$ the symbolic product is $ij-i$, and hence the typical term to be squared is:

$$\bar{Y}_{ij.}-\bar{Y}_{i..}$$

Similarly, for $\varepsilon_{k(ij)}$, the symbolic product is $ijk - ij$. Hence the typical term to be squared is:

$$Y_{ijk} - \bar{Y}_{ij}.$$

Note that we write the first term as Y_{ijk} since it is not averaged over any subscript.

Step 7. *Combining the steps of squaring, summing, and multiplying by the coefficient yields the appropriate sums of squares.*

Example The sums of squares for our example are shown in column 7.

Step 8. *The degrees of freedom are obtained by replacing in each symbolic product the subscript variable by its limit.*

Example For our example, the degrees of freedom are shown in column 8. For instance, for α_i the symbolic product is $i - 1$; hence $df = a - 1$. Similarly for $\varepsilon_{k(ij)}$, the symbolic product is $ijk - ij$; hence $df = abn - ab = ab(n - 1)$.

Step 9. *The total sum of squares is always defined as the sum, over all observations, of the squared deviations of the observations from the overall mean. The total degrees of freedom are always defined as one less than the total number of observations.*

D.3 Rule for Finding Expected Mean Squares

The rule for finding expected mean squares that we shall now present enables us to avoid tedious derivations. The rule applies to both nested factors and crossed factors. *The rule is applicable to all balanced designs with two or more replications and with no interaction terms assumed to equal zero.* We continue to use the training school example of Table 26.1 as our illustration. Here factor A (school) and factor B (instructor) are both fixed factors, factor B is nested within factor A, factor B has b levels within each level of factor A, factor A has a levels, and there are n replications.

Rule (D.4)

The rule for finding expected mean squares to be presented may appear to be a bit complex on first reading. However, with a little practice the desired expected mean squares can be obtained very quickly and easily.

Step 1. *List the model equation.*

Example The model equation is that of (D.2a):

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \varepsilon_{k(ij)}$$

Step 2. *For each term other than the overall constant, write the associated random effects variance term.*

Example

α_i	$\beta_{j(i)}$	$\varepsilon_{k(ij)}$
σ_α^2	σ_β^2	σ^2

If factors have fixed effects, as in this example, we shall at the end replace these variance terms by sums of squared effects divided by degrees of freedom. For instance, in the training school example the term σ_α^2 later will be replaced by $\sum \alpha_i^2/(a-1)$, and likewise σ_β^2 will be replaced by $\sum\sum \beta_{j(i)}^2/a(b-1)$. In the meantime, however, it is easier to write the variance term rather than a sum of squared effects divided by degrees of freedom.

Step 3. *Set up a table, with the rows consisting of the model elements other than the overall constant.*

Example

α_i	
$\beta_{j(i)}$	
$\varepsilon_{k(ij)}$	

Step 4. *The column headings for the table are the subscripts in the model. Under each heading, write F if the factor indexed by the subscript is fixed, and write R if it is random. Also write the number of levels for that factor.*

Example

	i	*j*	*k*
	F	*F*	*R*
	a	*b*	*n*
α_i			
$\beta_{j(i)}$			
$\varepsilon_{k(ij)}$			

For instance, i refers to school, a fixed factor that occurs at a levels. Note that the subscript k refers to replication, which is a random "factor" and occurs at n levels.

Step 5. *In each row where one or more subscripts are in parentheses, enter a* 1 *in the column(s) corresponding to the subscript(s) in parentheses.*

Example

	i	*j*	*k*
	F	*F*	*R*
	a	*b*	*n*
α_i			
$\beta_{j(i)}$	1		
$\varepsilon_{k(ij)}$	1	1	

Thus, in the $\beta_{j(i)}$ row, we enter a 1 in the i column, and so on.

Step 6. *In each row where one or more subscripts are not in parentheses, enter in the column(s) corresponding to the subscript(s) not in parentheses a* 1 *if the subscript refers to a random factor, and a* 0 *if the factor is fixed.*

Example

	i	j	k
	F	F	R
	a	b	n
α_i	0		
$\beta_{j(i)}$	1	0	
$\varepsilon_{k(ij)}$	1	1	1

Thus, for the $\beta_{j(i)}$ row, the subscript not in parentheses is j, which refers to factor B, a fixed factor. Hence, a 0 is entered in the j column.

Step 7. *Fill in all remaining empty cells with the number of levels appearing in the column heading.*

Example

	i	j	k
	F	F	R
	a	b	n
α_i	0	b	n
$\beta_{j(i)}$	1	0	n
$\varepsilon_{k(ij)}$	1	1	1

Each $E\{MS\}$ will consist of a linear combination of the variance terms enumerated in step 2, with the coefficients obtained by taking additional steps in the table just completed. Some of the coefficients may be zero, which means that the corresponding variance term is not present in the $E\{MS\}$.

Step 8. *Adjoin on the right of the table just completed the variance term associated with the effect in that row. In addition, adjoin a column for each expected mean square to be found. Under each expected mean square, indicate all of the subscripts (including any parentheses) associated with the corresponding model term.*

Example

	i	j	k				
	F	F	R		$E\{MSA\}$	$E\{MSB(A)\}$	$E\{MSE\}$
	a	b	n	**Variance**	i	$(i)j$	$(ij)k$
α_i	0	b	n	σ_α^2			
$\beta_{j(i)}$	1	0	n	σ_β^2			
$\varepsilon_{k(ij)}$	1	1	1	σ^2			

Note that all of the subscripts of the associated model term, whether in parentheses or not, are shown under the expected mean square. For example, $E\{MSB(A)\}$ has associated with it the model term $\beta_{j(i)}$, so that the subscripts shown are (i) and j. Similarly, $E\{MSE\}$ has associated with it the model term $\varepsilon_{k(ij)}$, so that (ij) and k are shown.

Step 9. *For each expected mean square column, the coefficient of any variance term is zero if the subscript(s) of the model term in that row (whether in parentheses or not) do not include all of the subscript(s) in the heading of that E{MS} column (whether in parentheses or not).*

Example

	i	*j*	*k*				
	F	*F*	*R*		*E{MSA}*	*E{MSB(A)}*	*E{MSE}*
	a	*b*	*n*	**Variance**	*i*	*(i)j*	*(ij)k*
α_i	0	b	n	σ_α^2		0	0
$\beta_{j(i)}$	1	0	n	σ_β^2			0
$\varepsilon_{k(ij)}$	1	1	1	σ^2			

For the $E\{MSA\}$ column, it will be noted that the model terms in all rows contain the subscript i. Hence, none of the variances receives a zero coefficient as a result of this step.

For the $E\{MSB(A)\}$ column, note that the first row has a model term not containing both i and j. Hence, σ_α^2 receives a zero coefficient in the $E\{MSB(A)\}$ column.

Finally, for the $E\{MSE\}$ column, the first and second rows have model terms that do not contain the three subscripts i, j, and k. Hence, both σ_α^2 and σ_β^2 receive zero coefficients in the $E\{MSE\}$ column.

Step 10. *The coefficients of the variance terms that have not been assigned a zero coefficient as a result of step 9 are found as follows:*

a. *For each expected mean square column, delete (e.g., mask or cover) the column(s) on the left corresponding to the subscripts not in parentheses in the heading of the E{MS} column.*
b. *Multiply the entries in the remaining columns for each row being considered.*

Step 11. *The expected mean square equals the sum of the products of each coefficient times the associated variance term, with the variance terms for fixed effects replaced by sums of squared effects divided by degrees of freedom.*

Example

	i	*j*	*k*				
	F	*F*	*R*		*E{MSA}*	*E{MSB(A)}*	*E{MSE}*
	a	*b*	*n*	**Variance**	*i*	*(i)j*	*(ij)k*
α_i	0	b	n	σ_α^2	bn	0 (step 9)	0 (step 9)
$\beta_{j(i)}$	1	0	n	σ_β^2	0	n	0 (step 9)
$\varepsilon_{k(ij)}$	1	1	1	σ^2	1	1	1

To find the coefficients for the $E\{MSA\}$ column, for example, we noted earlier that no zero coefficient is assigned as a result of step 9. Step 10a calls for column i on the left to

be deleted. Hence, we obtain by multiplying the terms in the j and k columns:

	j	k		
	F	R		$E\{MSA\}$
	b	n	**Variance**	i
α_i	b	n	σ_α^2	bn
$\beta_{j(i)}$	0	n	σ_β^2	0
$\varepsilon_{k(ij)}$	1	1	σ^2	1

Thus:

$$E\{MSA\} = bn\sigma_\alpha^2 + (0)\sigma_\beta^2 + (1)\sigma^2 = bn\sigma_\alpha^2 + \sigma^2$$

Since factor A has fixed effects, we finally obtain:

$$E\{MSA\} = bn\frac{\sum \alpha_i^2}{a-1} + \sigma^2$$

We find the remaining coefficients for $E\{MSB(A)\}$ in similar fashion. We delete column j on the left, the subscript not in parentheses, and obtain:

	i	k		
	F	R		$E\{MSB(A)\}$
	a	n	**Variance**	$(i)j$
α_i	0	n	σ_α^2	0 (step 9)
$\beta_{j(i)}$	1	n	σ_β^2	n
$\varepsilon_{k(ij)}$	1	1	σ^2	1

Thus:

$$E\{MSB(A)\} = (0)\sigma_\alpha^2 + n\sigma_\beta^2 + (1)\sigma^2 = n\sigma_\beta^2 + \sigma^2$$

Since factor B has fixed effects, we finally obtain:

$$E\{MSB(A)\} = n\frac{\sum\sum \beta_{j(i)}^2}{a(b-1)} + \sigma^2$$

To find the remaining coefficient in the $E\{MSE\}$ column, we delete column k, and the product on the σ^2 line is $1 \cdot 1 = 1$. Thus:

$$E\{MSE\} = (0)\sigma_\alpha^2 + (0)\sigma_\beta^2 + (1)\sigma^2 = \sigma^2$$

Assembling our results, we have:

$$E\{MSA\} = bn\frac{\sum \alpha_i^2}{a-1} + \sigma^2 \qquad \textbf{(D.5a)}$$

$$E\{MSB(A)\} = n\frac{\sum\sum \beta_{j(i)}^2}{a(b-1)} + \sigma^2 \qquad \textbf{(D.5b)}$$

$$E\{MSE\} = \sigma^2 \qquad \textbf{(D.5c)}$$

Comment

Some computer packages provide the expected mean squares for any balanced ANOVA study. An example is shown in Figure 26.7. ■

D.4 No Replications and/or Some Interactions Equal Zero

Modification of Rules

When a balanced design includes no replications and/or some interactions are assumed to equal zero—as, for instance, in a randomized complete block design with fixed block effects—rules (D.1) and (D.3) need to be modified slightly. Rule (D.4) requires no modification.

The modification of rule (D.1) is very slight. Step 2 now becomes:

Rule (D.1) modification: Step 2. *Include all interaction terms except those assumed to equal zero and those containing both a nested factor and the factor within which it is nested.* **(D.6)**

The modification of rule (D.3) is also a simple one:

Rule (D.3) modification: Steps 2 through 8 do not apply to the model error term ε. Instead, the sum of squares associated with the model error term ε is obtained as a remainder from the total sum of squares. Likewise, the degrees of freedom associated with this remainder sum of squares are obtained as a remainder from the total degrees of freedom. **(D.7)**

The sum of squares associated with the model error term ε in balanced designs where there are no replications and/or where some interaction terms are assumed to equal zero will be denoted by *SSRem*, which stands for the *remainder sum of squares*. Frequently, the remainder sum of squares will turn out to be an interaction sum of squares for the interaction terms in the model that are assumed to equal zero. The *remainder mean square* will be denoted by *MSRem*.

Additional Modification for Latin Square Designs

For latin square design model (28.12), one of the subscripts in Y_{ijk} is redundant since the row and column indices define the treatment for a given latin square design. Hence, when using the rules presented in the case of a latin square design, one of the subscripts must be treated as redundant, i.e., it needs to be ignored.

D.5 Additional Examples of Use of Rules

Crossed Two-Factor Study—Mixed Factor Effects

Consider a two-factor experiment in a completely randomized design, where factors A and B are crossed, factor A has fixed effects and factor B has random effects, and n replications are obtained for each factor combination. The model equation is that of (25.42):

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{k(ij)}$$

where we now recognize the nesting of the error term ε.

Table D.2 contains the derivation of the sums of squares. Table D.3a contains the preliminary tabulations for finding the expected mean squares, while Table D.3b presents the results of steps 9 and 10 of rule (D.4). The random effects variance terms corresponding to the model terms are:

$$\begin{array}{cccc} \alpha_i & \beta_j & (\alpha\beta)_{ij} & \varepsilon_{k(ij)} \\ \sigma_\alpha^2 & \sigma_\beta^2 & \sigma_{\alpha\beta}^2 & \sigma^2 \end{array}$$

Here, only the α_i are fixed effects, so at the end σ_α^2 will need to be replaced by a sum of squared effects divided by degrees of freedom. Note in Table D.3b that for finding $E\{MSA\}$, σ_β^2 receives a zero coefficient as a result of step 9 since the subscript in the β_j model term does not contain the subscript i in the $E\{MSA\}$ column. Column i is deleted for step 10 for finding the coefficients in the $E\{MSA\}$ column since it is the only subscript in the column heading and is not in parentheses. The other expected mean squares coefficients are found in similar fashion. Table D.3b indicates for each expected mean square whether the zero coefficients are obtained from step 9, and also which columns are deleted. The final expected mean squares, presented in Table D.3c, are identical to those shown in Table 25.5.

Subsampling in Randomized Block Design

The model usually employed for a randomized block design when only a single observation is made on an experimental unit is ANOVA model (21.1) in the case of fixed treatment and block effects:

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \varepsilon_{ij} \tag{D.8}$$

We shall now consider a slightly more complex case, namely when subsampling is used in a randomized block design—that is, when more than one observation is made on each experimental unit. Consider, for instance, an experiment to study how three different motivational stimuli affect the length of time a person requires to perform a task. The persons in the experiment are blocked into groups of three, according to age, and each person is assigned at random one of the three motivational stimuli. Three observations are then made on the time required to complete the task; that is, the subject is asked to perform the same task three times.

In this type of situation, we simply add a random observation error component to ANOVA model (D.8). Assuming that the treatment and block effects (motivational stimuli and age groups in our example) are fixed, an appropriate model is:

$$Y_{ijk} = \mu_{..} + \rho_i + \tau_j + \varepsilon_{(ij)} + \eta_{k(ij)} \tag{D.9}$$

where:

$$\sum \rho_i = 0$$
$$\sum \tau_j = 0$$

$\varepsilon_{(ij)}$ and $\eta_{k(ij)}$ are independent normal random variables with expectations 0 and variances σ^2 and σ_η^2, respectively

$$i = 1, \ldots, n_b; \qquad j = 1, \ldots, r; \qquad k = 1, \ldots, m$$

TABLE D.2 Derivation of Sums of Squares Formulas for Crossed Two-Factor Experiment in Completely Randomized Design.

(1) Model Term	(2) SS	(3) Coefficient	(4) $\sum$	(5) Symbolic Product	(6) Term to Be Squared	(7) Sum of Squares	(8) Degrees of Freedom
α_i	SSA	bn	$\sum_i$	$i-1$	$\bar{Y}_{i..}-\bar{Y}_{...}$	$bn\sum_i(\bar{Y}_{i..}-\bar{Y}_{...})^2$	$a-1$
β_j	SSB	an	$\sum_j$	$j-1$	$\bar{Y}_{.j.}-\bar{Y}_{...}$	$an\sum_j(\bar{Y}_{.j.}-\bar{Y}_{...})^2$	$b-1$
$(\alpha\beta)_{ij}$	$SSAB$	n	$\sum_i\sum_j$	$(i-1)(j-1)=ij-i-j+1$	$\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...}$	$n\sum_i\sum_j(\bar{Y}_{ij.}-\bar{Y}_{i..}-\bar{Y}_{.j.}+\bar{Y}_{...})^2$	$(a-1)(b-1)$
$\varepsilon_{k(ij)}$	SSE	1	$\sum_i\sum_j\sum_k$	$(k-1)ij=ijk-ij$	$Y_{ijk}-\bar{Y}_{ij.}$	$\sum_i\sum_j\sum_k(Y_{ijk}-\bar{Y}_{ij.})^2$	$ab(n-1)$
Total	$SSTO$				$Y_{ijk}-\bar{Y}_{...}$	$\sum_i\sum_j\sum_k(Y_{ijk}-\bar{Y}_{...})^2$	$abn-1$

TABLE D.3 ***E{MS}* Derivations for Crossed Two-Factor Experiment (*A* fixed, *B* random).**

(a) Table

	i	j	k
	F	R	R
	a	b	n
α_i	0	b	n
β_j	a	1	n
$(\alpha\beta)_{ij}$	0	1	n
$\varepsilon_{k(ij)}$	1	1	1

(b) Coefficients

Variance	$E\{MSA\}$ i	$E\{MSB\}$ j	$E\{MSAB\}$ ij	$E\{MSE\}$ $(ij)k$
σ_α^2	$b \cdot n$	0 (step 9)	0 (step 9)	0 (step 9)
σ_β^2	0 (step 9)	$a \cdot n$	0 (step 9)	0 (step 9)
$\sigma_{\alpha\beta}^2$	$1 \cdot n$	$0 \cdot n$	n	0 (step 9)
σ^2	$1 \cdot 1$	$1 \cdot 1$	1	$1 \cdot 1$
	(i col. deleted)	(j col. deleted)	(i, j cols. deleted)	(k col. deleted)

(c) $E\{MS\}$

$$E\{MSA\} = bn\frac{\sum \alpha_i^2}{a-1} + n\sigma_{\alpha\beta}^2 + \sigma^2$$

$$E\{MSB\} = an\sigma_\beta^2 + \sigma^2$$

$$E\{MSAB\} = n\sigma_{\alpha\beta}^2 + \sigma^2$$

$$E\{MSE\} = \sigma^2$$

Here ρ_i is the block effect, τ_j the treatment effect, $\varepsilon_{(ij)}$ the random effect associated with the experimental unit, and $\eta_{k(ij)}$ the random effect associated with the kth observation on the experimental unit. Note that the experimental error ε is nested within the (ij) block-treatment combination; there is no additional subscript since only one experimental unit is assigned to a treatment within a block. Thus, there are no replications for experimental units. Also note that the observation error η is nested within the (ij) block-treatment combination.

Since there are no replications present and the block-treatment interactions are assumed to equal zero, we need to use the modified rules, as explained in Section D.4. Table D.4 contains the derivation of the sums of squares for ANOVA model (D.9), and Table D.5 contains the derivation of the expected mean squares. Note that the sum of squares for experimental units is obtained as a remainder in Table D.4 because there is only one experimental unit assigned to a treatment within a block. As expected, *SSRem* turns out to be the block-treatment interaction sum of squares, as for a randomized block design without subsampling.

TABLE D.4 Derivation of Sums of Squares Formulas for Randomized Block Design with Subsampling—ANOVA Model (D.9).

Model Term	Symbolic Product	Sum of Squares	Degrees of Freedom
ρ_i	$i-1$	$SSBL = rm\sum(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$n_b - 1$
τ_j	$j-1$	$SSTR = n_b m\sum(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$r-1$
$\varepsilon_{(ij)}$		$SSRem = SSBL.TR$ $= m\sum\sum(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$	Remainder $=(n_b-1)(r-1)$
$\eta_{k(ij)}$	$(k-1)ij = ijk - ij$	$SSOE = \sum\sum\sum(Y_{ijk} - \bar{Y}_{ij\cdot})^2$	$n_b r(m-1)$
Total		$SSTO = \sum\sum\sum(Y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$	$n_b rm - 1$

TABLE D.5 Derivation of Expected Mean Squares for Randomized Block Design with Subsampling—ANOVA Model (D.9).

(a) Table

	i	j	k	Variance	Expected Mean Square of BL	TR	Rem	OE
	F n_b	F r	R m		i	j	(ij)	$(ij)k$
ρ_i	0	r	m	σ_ρ^2	rm	0	0	0
τ_j	n_b	0	m	σ_τ^2	0	$n_b m$	0	0
$\varepsilon_{(ij)}$	1	1	m	σ^2	m	m	m	0
$\eta_{k(ij)}$	1	1	1	σ_η^2	1	1	1	1

(b) Expected Mean Squares

$$E\{MSBL\} = rm\frac{\sum \rho_i^2}{n_b - 1} + m\sigma^2 + \sigma_\eta^2$$

$$E\{MSTR\} = n_b m\frac{\sum \tau_j^2}{r-1} + m\sigma^2 + \sigma_\eta^2$$

$$E\{MSRem\} = m\sigma^2 + \sigma_\eta^2$$

$$E\{MSOE\} = \sigma_\eta^2$$

Table D.5b indicates that for ANOVA model (D.9) with fixed treatment and block effects, the test statistic for examining the presence of treatment effects is $F^* = MSTR/MSRem$, as is also the case when no subsampling occurs in a randomized complete block design—see (21.7b). Remember that $MSRem$ denotes simply the interaction mean square $MSBL.TR$ here.

Problems

D.1. Refer to ANOVA model (25.39). Use rule (D.4) to obtain the expected mean squares in Table 25.5 for this model.

D.2. Refer to ANOVA model (25.77).

a. Use rule (D.3) to obtain the sums of squares formulas in (24.22) and the associated degrees of freedom.

b. Use rule (D.4) to obtain the expected mean squares in Table 25.9.

D.3. Refer to ANOVA model (25.79).

a. Use rule (D.3) to obtain the sums of squares formulas in (24.22) and the associated degrees of freedom.

b. Use rule (D.4) to obtain the expected mean squares in Table 25.10.

D.4. Refer to nested design model (26.7), but assume that factor A is nested within factor B, factor A effects are random, and factor B effects are fixed. (See also "Random Factor Effects" on page 1093.)

a. Use rule (D.3) to obtain the sums of squares formulas and the associated degrees of freedom.

b. Use rule (D.4) to obtain the expected mean squares.

c. What is the appropriate mean square to be used in constructing a confidence interval for $\mu_{\cdot j}$?

D.5. Refer to randomized complete block model (21.1).

a. Use rule (D.3) and modification (D.7) to obtain the sums of squares formulas in (21.6) and the associated degrees of freedom.

b. Use rule (D.4) to obtain the expected mean squares in Table 21.2 for this model.

D.6. Refer to randomized complete block model (21.1), but assume that treatment effects are random. (See also Comment 2 on page 897.)

a. Use rule (D.3) and modification (D.7) to obtain the sums of squares formulas in (21.6) and the associated degrees of freedom.

b. Use rule (D.4) to obtain the expected mean squares in Table 21.2 for this model.

D.7. Refer to randomized complete block model (25.67).

a. Use rule (D.3) and modification (D.7) to obtain the sums of squares formulas in (21.6) and the associated degrees of freedom.

b. Use rule (D.4) to obtain the expected mean squares in Table 25.8 for this model.

D.8. Refer to randomized complete block model (D.9), but assume that block effects are random.

a. Use rule (D.3) and modification (D.7) to obtain the sums of squares formulas and the associated degrees of freedom.

b. Use rule (D.4) to obtain the expected mean squares.

D.9. In a balanced three-factor study, factors A and C are crossed and factor B is nested within factor C. Factor A has fixed effects, and factors B and C have random effects. There are n replications for each treatment.

a. Use rule (D.3) to obtain the sums of squares formulas and the associated degrees of freedom.

b. Use rule (D.4) to obtain the expected mean squares.

c. What is the appropriate denominator mean square for testing for factor A main effects?

D.10. **Swimmer motivation.** A large metropolitan swim club for youths studied the effects of three motivational stimuli on performance. The three motivational stimuli were: (1) presentation of merit award, (2) granting of team leadership privileges, and (3) publicity in the club newsletter.

Since age is known to be related to performance, the nine female swimmers included in the study were grouped according to age into three blocks of three each. Within each age block, the three swimmers were randomly assigned to one of the motivation treatments. After a suitable amount of training, each swimmer was timed on three separate occasions while swimming a fixed distance. The coded data on the time for each of the three trials follow.

		Motivation Treatment		
Block	Observation	$j = 1$ Merit Award	$j = 2$ Leadership	$j = 3$ Publicity
$i = 1$	$k = 1$:	28	26	27
(7–8 years)	$k = 2$:	32	24	29
	$k = 3$:	31	27	30
$i = 2$	$k = 1$:	24	22	20
(9–10 years)	$k = 2$:	26	19	21
	$k = 3$:	23	18	22
$i = 3$	$k = 1$:	18	13	17
(11–12 years)	$k = 2$:	21	16	19
	$k = 3$:	20	15	19

Obtain the residuals for randomized block model (D.9) and plot them against the fitted values. Also prepare a normal probability plot of the residuals. What are your findings about the appropriateness of model (D.9)?

D.11. Refer to **Swimmer motivation** Problem D.10. Assume that randomized block model (D.9) with fixed block and treatment effects is appropriate.

a. Obtain the analysis of variance table.

b. Test whether or not the mean times are the same for the three motivational stimuli; use $\alpha = .05$. State the alternatives, decision rule, and conclusion. What is the P-value of the test?

c. Make all pairwise comparisons among the three treatment means; use the Tukey procedure with a 90 percent family confidence coefficient. State your findings.

d. Obtain point estimates of σ^2 and σ_η^2. Does one variance appear to be much larger than the other? Discuss.

D.12. Refer to repeated measures model (27.21). Consider a simpler model, in which interactions SA and SB are not present. The parameters $\mu_{...}$, ρ_i, α_j, β_k, $(\alpha\beta)_{jk}$, and ε_{ijk} are defined in the same way as (27.21).

a. Use rule (D.3) and modification (D.7) to obtain the sums of squares formulas similar to those in Table 27.11b and the associated degrees of freedom similar to those in Table 27.11a.

b. Use rule (D.4) to obtain the expected mean squares similar to those in Table 27.11a.

D.13. Refer to repeated measures model (27.11).

a. Use rule (D.3) and modification (D.7) to obtain the sums of squares formulas and the associated degrees of freedom in Table 27.5.

b. Use rule (D.4) to obtain the expected mean squares in Table 27.6.

D.14. Refer to the **Drug effect experiment** data set. Consider the combined study. Assume that subjects (rats) and observation units have random effects, and that factor A (initial lever press rate), factor B (dosage level), and factor C (reinforcement schedule) have fixed effects. Also assume that there are no interactions between subjects and treatments.

a. Use rule (D.1) and modification (D.6) to develop the model for this experiment.

b. Use rule (D.3) and modification (D.7) to obtain the sums of squares formulas and the associated degrees of freedom.

c. Use rule (D.4) to obtain the expected mean squares.

D.15. Derive the expected mean squares in Table 28.5 for latin square model (28.12) by using rule (D.4). (See also "Additional Modification for Latin Square Designs" on page 1366.)

D.16. Derive the expected mean squares for latin square model (28.27) with n replications by using rule (D.4). (See also "Additional Modification for Latin Square Designs" on page 1366.)

D.17. Derive the expected mean squares in Table 28.10 for latin square cross-over model (28.29) with n subjects for each treatment order pattern by using rule (D.4). (See also "Additional Modification for Latin Square Designs" on page 1366.)

Appendix

Selected Bibliography

The selected references are grouped into the following categories:

1. General regression books
2. General linear models books
3. Diagnostics and model building
4. Statistical computing
5. Nonlinear regression
6. Miscellaneous regression topics
7. General experimental design and analysis of variance books
8. Miscellaneous experimental design and analysis of variance topics

1. General Regression Books

Allison, P. D. *Multiple Regression: A Primer*. Thousand Oaks, Calif.: Sage Publications, 1999.

Bowerman, B. L., and R. T. O'Connell. *Linear Statistical Models: An Applied Approach*. 2nd ed. Boston: Duxbury Press, 1990.

Chatterjee, S.; A. S. Hadi; and B. Price. *Regression Analysis by Example*. 3rd ed. New York: John Wiley & Sons, 1999.

Cohen, J.; P. Cohen; S. G. West; and L. S. Aiken. *Applied Multiple Regression/ Correlation Analysis for the Behavioral Sciences*. 3rd ed. Hillsdale, N.J.: Lawrence Erlbaum Associates, 2003.

Cook, R. D., and S. Weisberg. *Applied Regression Including Computing and Graphics*. New York: John Wiley & Sons, 1999.

Daniel, C., and F. S. Wood. *Fitting Equations to Data: Computer Analysis of Multifactor Data*. 2nd ed. New York: John Wiley & Sons, 1999.

Draper, N. R., and H. Smith. *Applied Regression Analysis*. 3rd ed. New York: John Wiley & Sons, 1998.

Freund, R. J., and R. C. Littell. *SAS System for Regression*. 3rd ed. New York: John Wiley & Sons, 2000.

Graybill, F. A., and H. Iyer. *Regression Analysis: Concepts and Applications*. Belmont, Calif.: Duxbury Press, 1994.

Hamilton, L. C. *Regression with Graphics: A Second Course in Applied Statistics*. Pacific Grove, Calif.: Brooks/Cole Publishing, 1992.

Kleinbaum, D. G.; L. L. Kupper; K. E. Muller; and A. Nizam. *Applied Regression Analysis and Other Multivariate Methods*. 3rd ed. Belmont, Calif.: Duxbury Press, 1998.

Mendenhall, W., and T. Sincich. *A Second Course in Business Statistics: Regression Analysis*. 5th ed. Upper Saddle River, N.J.: Prentice Hall, 1996.

Muller, K. E., and B. A. Fetterman. *Regression and ANOVA: An Integrated Approach Using SAS Software*. New York: John Wiley & Sons, 2003.

Myers, R. H. *Classical and Modern Regression with Applications*. 2nd ed. Boston: Duxbury Press, 1990.

Pedhazur, E. J. *Multiple Regression in Behavioral Research*. 3rd ed. Belmont, Calif.: Duxbury Press, 1997.

Rawlings, J. O; S. G. Pantula; and D. A. Dickey. *Applied Regression Analysis: A Research Tool*. New York: Springer-Verlag, 1998.

Ryan, T. P. *Modern Regression Methods*. New York: John Wiley & Sons, 1997.

Seber, G. A. F., and A. S. Lee. *Linear Regression Analysis*. 2nd ed. New York: John Wiley & Sons, 2003.

Sen, A., and M. Srivastava. *Regression Analysis: Theory, Methods, and Applications*. 4th ed. New York: Springer-Verlag, 1997.

2. General Linear Models Books

Graybill, F. A. *Theory and Application of the Linear Model*. Boston: Duxbury Press, 1976.

Hocking, R. R. *Methods and Applications of Linear Models: Regression and the Analysis of Variance*. 2nd ed. New York: John Wiley & Sons, 2003.

Littell, R. C.; W. W. Stroup; and R. J. Freund. *SAS System for Linear Models*. 4th ed. New York: John Wiley & Sons, 2002.

Searle, S. R. *Linear Models*. New York: John Wiley & Sons, 1997.

Searle, S. R. *Linear Models for Unbalanced Data*. New York: John Wiley & Sons, 1987.

3. Diagnostics and Model Building

Allen, D. M. "Mean Square Error of Prediction as a Criterion for Selecting Variables." *Technometrics* 13 (1971), pp. 469–75.

Anscombe, F. J., and J. W. Tukey. "The Examination and Analysis of Residuals." *Technometrics* 5 (1963), pp. 141–60.

Atkinson, A. C. "Two Graphical Displays for Outlying and Influential Observations in Regression." *Biometrika* 68 (1981), pp. 13–20.

Atkinson, A. C. *Plots, Transformations, and Regression*. Oxford: Clarendon Press, 1987.

Barnett, V., and T. Lewis. *Outliers in Statistical Data*. 3rd ed. New York: John Wiley & Sons, 1994.

Belsley, D. A. *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. New York: John Wiley & Sons, 1991.

Belsley, D. A.; E. Kuh; and R. E. Welsch. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: John Wiley & Sons, 1980.

Box, G. E. P., and D. R. Cox. "An Analysis of Transformations." *Journal of the Royal Statistical Society B* 26 (1964), pp. 211–43.

Box, G. E. P., and N. R. Draper. *Empirical Model-Building and Response Surfaces*. New York: John Wiley & Sons, 1987.

Box, G. E. P., and P. W. Tidwell. "Transformations of the Independent Variables." *Technometrics* 4 (1962), pp. 531–50.

Breiman, L., and P. Spector. "Submodel Selection and Evaluation in Regression: The X-Random Case." *International Statistical Review* 60 (1992), pp. 291–319.

Breusch, T. S., and A. R. Pagan. "A Simple Test for Heteroscedasticity and Random Coefficient Variation." *Econometrica* 47 (1979), pp. 1287–94.

Brown, M. B., and A. B. Forsythe. "Robust Tests for Equality of Variances," *Journal of the American Statistical Association* 69 (1974), pp. 364–67.

Carroll, R. J., and D. Ruppert. *Transformation and Weighting in Regression*. New York: Chapman & Hall, 1988.

Chatterjee, S., and A. S. Hadi. *Sensitivity Analysis in Linear Regression*. New York: John Wiley & Sons, 1988.

Conover, W. J.; M. E. Johnson; and M. M. Johnson. "A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data." *Technometrics* 23 (1981), pp. 351–61.

Cook, R. D. "Exploring Partial Residual Plots." *Technometrics* 35 (1993), pp. 351–62.

Cook, R. D., and S. Weisberg. "Diagnostics for Heteroscedasticity in Regression." *Biometrika* 70 (1983), pp. 1–10.

Cox, D. R. *Planning of Experiments*. New York: John Wiley & Sons, 1958.

Davidian, M., and R. J. Carroll. "Variance Function Estimation." *Journal of the American Statistical Association* 82 (1987), pp. 1079–91.

Durbin, J., and G. S. Watson. "Testing for Serial Correlation in Least Squares Regression. II." *Biometrika* 38 (1951), pp. 159–78.

Faraway, J. J. "On the Cost of Data Analysis." *Journal of Computational and Graphical Statistics* 1 (1992), pp. 213–29.

Flack, V. F., and P. C. Chang. "Frequency of Selecting Noise Variables in Subset-Regression Analysis: A Simulation Study." *The American Statistician* 41 (1987), pp. 84–86.

Freedman, D. A. "A Note on Screening Regression Equations." *The American Statistician* 37 (1983), pp. 152–55.

Hoaglin, D. C.; F. Mosteller; and J. W. Tukey. *Exploring Data Tables, Trends, and Shapes*. New York: John Wiley & Sons, 1985.

Hoaglin, D. C., and R. Welsch. "The Hat Matrix in Regression and ANOVA." *The American Statistician* 32 (1978), pp. 17–22.

Hocking, R. R. "The Analysis and Selection of Variables in Linear Regression." *Biometrics* 32 (1976), pp. 1–49.

Hoerl, A. E., and R. W. Kennard. "Ridge Regression: Applications to Nonorthogonal Problems." *Technometrics* 12 (1970), pp. 69–82.

Joglekar, G.; J. H. Schuenemeyer; and V. LaRiccia. "Lack-of-Fit Testing When Replicates Are Not Available." *The American Statistician* 43 (1989), pp. 135–43.

Levene, H. "Robust Tests for Equality of Variances," in *Contributions to Probability and Statistics*, ed. I. Olkin. Palo Alto, Calif.: Stanford University Press, 1960, pp. 278–92.

Lindsay, R. M., and A. S. C. Ehrenberg. "The Design of Replicated Studies." *The American Statistician* 47 (1993), pp. 217–28.

Looney, S. W., and T. R. Gulledge, Jr. "Use of the Correlation Coefficient with Normal Probability Plots." *The American Statistician* 39 (1985), pp. 75–79.

Mallows, C. L. "Some Comments on C_p." *Technometrics* 15 (1973), pp. 661–75.

Mansfield, E. R., and M. D. Conerly. "Diagnostic Value of Residual and Partial Residual Plots." *The American Statistician* 41 (1987), pp. 107–16.

Mantel, N. "Why Stepdown Procedures in Variable Selection." *Technometrics* 12 (1970), pp. 621–25.

Miller, A. J. *Subset Selection in Regression*. 2nd ed. London: Chapman & Hall, 2002.

Pope, P. T., and J. T. Webster. "The Use of an F-Statistic in Stepwise Regression Procedures." *Technometrics* 14 (1972), pp. 327–40.

Rousseeuw, P. J., and A. M. Leroy. *Robust Regression and Outlier Detection*. New York: John Wiley & Sons, 1987.

Shapiro, S. S., and M. B. Wilk. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika* 52 (1965), pp. 591–611.

Snee, R. D. "Validation of Regression Models: Methods and Examples." *Technometrics* 19 (1977), pp. 415–28.

Stone, M. "Cross-Validatory Choice and Assessment of Statistical Prediction." *Journal of the Royal Statistical Society B* 36 (1974), pp. 111–47.

Velleman, P. F., and D. C. Hoaglin. *Applications, Basics, and Computing of Exploratory Data Analysis*. Boston: Duxbury Press, 1981.

4. Statistical Computing

BMDP New System 2.0. Statistical Solutions, Inc.

JMP Version 5. SAS Institute Inc.

Kennedy, W. J., and J. E. Gentle. *Statistical Computing*. New York: Marcel Dekker, 1980.

LogXact 5. Cytel Software Corporation. Cambridge, Mass., 2003.

MATLAB 6.5. The MathWorks, Inc.

MINITAB Release 13. Minitab Inc.

S-Plus 6 for Windows. Insightful Corporation

SAS/STAT Release 8.2. SAS Institute, Inc.

SPSS 11.5 for Windows. SPSS Inc.

SYSTAT 10.2. SYSTAT Software Inc.

Tierney, L. *LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*. New York: John Wiley & Sons, 1990.

5. Nonlinear Regression

Allison, P. D. *Logistic Regression Using the SAS System: Theory and Applications*. New York: John Wiley & Sons, 1999.

Bates, D. M., and D. G. Watts. *Nonlinear Regression Analysis and Its Applications*. New York: John Wiley & Sons, 1988.

Begg, C. B., and R. Gray. "Calculation of Polytomous Logistic Regression Parameters Using Individualized Regressions." *Biometrika* 71 (1984), pp. 11–18.

Box, M. J. "Bias in Nonlinear Estimation." *Journal of the Royal Statistical Society B* 33 (1971), pp. 171–201.

DeVeaux, R. D.; J. Schumi; J. Schweinsberg; and L. H. Ungar. "Prediction Intervals for Neural Networks via Nonlinear Regression," *Technometrics* 40 (1998), pp. 273–282.

DeVeaux, R. D., and L. H. Ungar. "A Brief Introduction to Neural Networks," http://www.williams.edu/Mathematics/rdeveaux/pubs.html (1996).

Gallant, A. R. "Nonlinear Regression." *The American Statistician* 29 (1975), pp. 73–81.

Gallant, A. R. *Nonlinear Statistical Models*. New York: John Wiley & Sons, 1987.

Halperin, M.; W. C. Blackwelder; and J. I. Verter. "Estimation of the Multivariate Logistic Risk Function: A Comparison of Discriminant Function and Maximum Likelihood Approaches." *Journal of Chronic Diseases* 24 (1971), pp. 125–58.

Hartley, H. O. "The Modified Gauss-Newton Method for the Fitting of Non-linear Regression Functions by Least Squares." *Technometrics* 3 (1961), pp. 269–80.

Hosmer, D. W., and S. Lemeshow. "Goodness of Fit Tests for the Multiple Logistic Regression Model." *Communications in Statistics A* 9 (1980), pp. 1043–69.

Hosmer, D. W., and S. Lemeshow. *Applied Logistic Regression*. 2nd ed. New York: John Wiley & Sons, 2000.

Hougaard, P. "The Appropriateness of the Asymptotic Distribution in a Nonlinear Regression Model in Relation to Curvature." *Journal of the Royal Statistical Society B* 47 (1985), pp. 103–14.

Kleinbaum, D. G.; L. L. Kupper; and L. E. Chambless. "Logistic Regression Analysis of Epidemiologic Data: Theory and Practice." *Communications in Statistics A* 11 (1982), pp. 485–547.

Landwehr, J. M.; D. Pregibon; and A. C. Shoemaker. "Graphical Methods for Assessing Logistic Regression Models (with discussion)." *Journal of the American Statistical Association* 79 (1984), pp. 61–83.

Marquardt, D. W. "An Algorithm for Least Squares Estimation of Non-linear Parameters." *Journal of the Society of Industrial and Applied Mathematics* 11 (1963), pp. 431–41.

Menard, S. *Applied Logistic Regression Analysis*. Thousand Oaks, Calif.: Sage Publications, 1995.

Pregibon, D. "Logistic Regression Diagnostics." *Annals of Statistics* 9 (1981), pp. 705–24.

Prentice, R. L. "Use of the Logistic Model in Retrospective Studies." *Biometrics* 32 (1976), pp. 599–606.

Ratkowsky, D. A. *Nonlinear Regression Modeling*. New York: Marcel Dekker, 1983.

Truett, J.; J. Cornfield; and W. Kannel. "A Multivariate Analysis of the Risk of Coronary Heart Disease in Framingham." *Journal of Chronic Diseases* 20 (1967), pp. 511–24.

6. Miscellaneous Regression Topics

Agresti, A. *Categorical Data Analysis*. 2nd ed. New York: John Wiley & Sons, 2002.

Altman, N. S. "An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression." *The American Statistician* 46 (1992), pp. 175–85.

Berkson, J. "Are There Two Regressions?" *Journal of the American Statistical Association* 45 (1950), pp. 164–80.

Bishop, Y. M. M.; S. E. Fienberg; and P. W. Holland. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Mass.: MIT Press, 1975.

Box, G. E. P. "Use and Abuse of Regression." *Technometrics* 8 (1966), pp. 625–29.

Box, G. E. P., and G. M. Jenkins. *Time Series Analysis, Forecasting and Control*. 3rd ed. San Francisco: Holden-Day, 1994.

Breiman, L.; J. H. Friedman; R. A. Olshen; and C. J. Stone. *Classification and Regression Trees*. New York: Chapman & Hall, 1993.

Christensen, R. *Log-Linear Models and Logistic Regression*. 2nd ed. New York: Springer-Verlag, 1997.

Cleveland, W. S. "Robust Locally Weighted Regression and Smoothing Scatterplots." *Journal of the American Statistical Association* 74 (1979), pp. 829–36.

Cleveland, W. S., and S. J. Devlin. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting." *Journal of the American Statistical Association* 83 (1988), pp. 596–610.

Collett, D. *Modelling Binary Data*. 2nd ed. London: Chapman & Hall, 2002.

Cox, D. R. "Notes on Some Aspects of Regression Analysis." *Journal of the Royal Statistical Society A* 131 (1968), pp. 265–79.

Cox, D. R. *The Analysis of Binary Data*. 2nd. ed. London: Chapman & Hall, 1989.

Efron, B. *The Jackknife, The Bootstrap, and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics, 1982.

Efron, B. "Better Bootstrap Confidence Intervals" (with discussion). *Journal of the American Statistical Association* 82 (1987), pp. 171–200.

Efron, B., and R. Tibshirani. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy." *Statistical Science* 1 (1986), pp. 54–77.

Efron, B., and R. J. Tibshirani. *An Introduction to the Bootstrap*. New York: Chapman & Hall, 1993.

Eubank, R. L. *Nonparametric Regression and Spline Smoothing*. 2nd ed. New York: Marcel Dekker, 1999.

Finney, D. J. *Probit Analysis*. 3rd ed. Cambridge, England: Cambridge University Press, 1971.

Frank, I. E., and J. H. Friedman. "A Statistical View of Some Chemometrics Regression Tools." *Technometrics* 35 (1993), pp. 109–35.

Friedman, J. H., and W. Stuetzle. "Projection Pursuit Regression." *Journal of the American Statistical Association* 76 (1981), pp. 817–23.

Fuller, W. A. *Measurement Error Models*. New York: John Wiley & Sons, 1987.

Gibbons, J. D. *Nonparametric Methods for Quantitative Analysis*. 2nd ed. Columbus, Ohio: American Sciences Press, 1985.

Graybill, F. A. *Matrices with Applications in Statistics*. 2nd ed. Belmont, Calif.: Duxbury Press, 2001.

Greene, W. H. *Econometric Analysis*. 5th ed. Upper Saddle River, N.J.: Prentice Hall, 2003.

Haerdle, W. *Applied Nonparametric Regression*. Cambridge, England: Cambridge University Press, 1990.

Harrell, F. E. *Regression Modeling Strategies: With Application to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer-Verlag, 2001.

Hastie, T., and C. Loader. "Local Regression: Automatic Kernel Carpentry" (with discussion). *Statistical Science* 8 (1993), pp. 120–43.

Hastie, T. J., and R. J. Tibshirani. *Generalized Additive Models*. New York: Chapman & Hall, 1990.

Hastie, T. J.; R. J. Tibshirani; and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer-Verlag, 2001.

Hochberg, Y., and A. C. Tamhane. *Multiple Comparison Procedures*. New York: John Wiley and Sons, 1987.

Hogg, R. V. "Statistical Robustness: One View of Its Use in Applications Today." *The American Statistician* 33 (1979), pp. 108–15.

Johnson, R. A., and D. W. Wichern. *Applied Multivariate Statistical Analysis*. 5th ed. Englewood Cliffs, N.J.: Prentice Hall, 2002.

Kendall, M. G., and J. D. Gibbons. *Rank Correlation Methods*. 5th ed. London: Charles Griffin, 1990.

Lachenbruch, P. A. *Discriminant Analysis*. New York: Hafner Press, 1975.

McCulloch, P., and J. A. Nelder. *Generalized Linear Models*. 2nd ed. New York: Chapman & Hall, 1989.

Miller, R. G., Jr. *Simultaneous Statistical Inference*. 2nd ed. New York: Springer-Verlag, 1991.

Nelder, J. A., and R. W. M. Wedderburn. "Generalized Linear Models." *Journal of the Royal Statistical Society A* 135 (1972), pp. 370–84.

Pindyck, R. S., and D. L. Rubinfeld. *Econometric Models and Economic Forecasts*. 4th ed. New York: McGraw-Hill, 1997.

Satterthwaite, F. E. "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2 (1946), pp. 110–14.

Searle, S. R. *Matrix Algebra Useful for Statistics*. New York: John Wiley & Sons, 1982.

Snedecor, G. W., and W. G. Cochran. *Statistical Methods*. 8th ed. Ames, Iowa: Iowa State University Press, 1989.

Theil, H., and A. L. Nagar. "Testing the Independence of Regression Disturbances." *Journal of the American Statistical Association* 56 (1961), pp. 793–806.

7. General Experimental Design and Analysis of Variance Books

Atkinson, A. C., and A. N. Donev. *Optimum Experimental Designs*. Oxford: Clarendon Press, 1992.

Box, G. E. P., and N. R. Draper. *Empirical Model-Building and Response Surfaces*. New York: John Wiley & Sons, 1987.

Box, G. E. P.; W. G. Hunter; and J. S. Hunter. *Statistics for Experimenters*. New York: John Wiley & Sons, 1978.

Cochran, W. G., and G. M. Cox. *Experimental Designs*. 2nd ed. New York: John Wiley & Sons, 1992.

Cook, R. D., and C. J. Nachtsheim. "Computer-Aided Blocking of Factorial and Response Surface Designs." *Technometrics* 31 (1989), pp. 339–346.

Cox, D. R. *Planning of Experiments*. New York: John Wiley & Sons, 1992.

Dean, A., and D. Voss. *Design and Analysis of Experiments*. New York: Springer-Verlag, 1999.

Fisher, R. A. *The Design of Experiments*. 8th ed. New York: Hafner Publishing Co., 1966.

Hicks, C. R., and K. V. Turner. *Fundamental Concepts in the Design of Experiments*. 3rd ed. New York: Holt, Rinehart and Winston, 1999.

Hinkelmann, K., and O. Kempthorne. *Design and Analysis of Experiments: Introduction to Experimental Design*. Vol. 1. New York: John Wiley & Sons, 1994.

Hoaglin, D. C.; F. Mosteller; and J. W. Tukey. *Fundamentals of Exploratory Analysis of Variance*. New York: John Wiley & Sons, 1991.

Hsu, J. C. *Multiple Comparisons: Theory and Methods*. London: Chapman & Hall, 1996.

Kempthorne, O. *The Design and Analysis of Experiments*. New York: John Wiley & Sons, 1952.

Kirk, R. E. *Experimental Design: Procedures for the Behavioral Sciences*. 3rd. ed. Monterey Calif.: Brooks/Cole Publishing Co., 1994.

Lorenzen, T. J., and V. L. Anderson. *Design of Experiments: A No-Name Approach*. New York: Marcel Dekker, 1993.

Mason, R. L.; R. F. Gunst; and J. L. Hess. *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*. 2nd ed. New York: John Wiley & Sons, 2003.

Montgomery, D. C. *Design and Analysis of Experiments*. 5th ed. New York: John Wiley & Sons, 2000.

Oehlert, G. W. *A First Course in Design and Analysis of Experiments*. New York: W. H. Freeman, 2000.

Scheffé, H. *The Analysis of Variance*. New York: John Wiley & Sons, 1959.

Searle, S. R.; G. Casella; and C. E. McCulloch. *Variance Components*. New York: John Wiley & Sons, 1992.

Snedecor, G. W., and W. G. Cochran. *Statistical Methods*. 8th ed. Ames, Iowa: Iowa State University Press, 1989.

Steel, R. G. D.; J. H. Torrie; and D. A. Dickey. *Principles and Procedures of Statistics: A Biometrical Approach*. New York: McGraw-Hill, 1996.

Winer, B. J.; D. R. Brown; and K. M. Michels. *Statistical Principles in Experimental Design*. 3rd ed. New York: McGraw-Hill, 1991.

Wu, C. F. J., and M. Hamada. *Experiments: Planning, Analysis, and Parameter Design Optimization*. New York: John Wiley & Sons, 2000.

8. Miscellaneous Experimental Design and Analysis of Variance Topics

Beckman, R. J.; R. D. Cook; and C. J. Nachtsheim. "Diagnostics for Mixed Model Analysis of Variance." *Technometrics* 29 (1987), pp. 413–26.

Berger, V. W. "Pros and Cons of Permutation Tests in Clinical Trials." *Statistics in Medicine* 19 (2000), pp. 1319–28.

Burdick, R. K. "Using Confidence Intervals to Test Variance Components." *Journal of Quality Technology* 26 (1994), pp. 30–38.

Burdick, R. K., and F. A. Graybill. *Confidence Intervals on Variance Components*. New York: Marcel Dekker, Inc., 1992.

Dunnett, C. W. "A Multiple Comparisons Procedure for Comparing Several Treatments with a Control." *Journal of the American Statistical Association* 50 (1955), p. 1096.

Gaylor, D. W., and F. N. Hopper. "Estimating the Degrees of Freedom for Linear Combinations of Mean Squares by Satterthwaite's Formula." *Technometrics* 11 (1969), pp. 691–706.

Hartley, H. O. "Testing the Homogeneity of a Set of Variances." *Biometrika*, 31 (1940), pp. 249–55.

Greenhouse, S. W., and S. Geisser. "On Methods in the Analysis of Profile Data." *Psychometrika* 24 (1959), pp. 95–112.

Hocking, R. R. "A Discussion of the Two-Way Mixed Model." *The American Statistician* 27 (1973), pp. 148–52.

Holland, B., and Copenhaven, M. D. "An Improved Sequentially Rejective Bonferroni Test Procedure." *Biometrics* 43 (1987), pp. 417–23.

Holland, B., and Copenhaven, M. D. "Improved Bonferroni-Type Multiple Testing Procedures." *Psychological Bulletin* 104 (1988), pp. 145–49.

Hsiao, C., *Analysis of Panel Data.* Cambridge, England: Cambridge University Press, 1986.

Huynh, H., and L. Feldt. "Estimation of the Box Correction for Degrees of Freedom from Sample Data in the Randomized Block and Split-Plot Designs." *Journal of Educational Statistics* 1 (1976), pp. 69–82.

Johnson, D. E., and F. A. Graybill. "Estimation of σ^2 in a Two-Way Classification Model with Interaction." *Journal of the American Statistical Association* 67 (1972), pp. 388–94.

Koch, G. G.; J. D. Elashoff; and I. A. Amara. "Repeated Measurements–Design and Analysis." In *Encyclopedia of Statistical Sciences,* vol. 8, ed. S. Kotz and N. L. Johnson. New York: John Wiley & Sons, 1988, pp. 46–73.

Kruskal, W. H., and W. A. Wallis. "Use of Ranks on One-Criterion Variance Analysis," *Journal of the American Statistical Association,* 47 (1952), pp. 583–621 (corrections appear in Vol. 48, pp. 907–11).

Meyer, R. K., and C. J. Nachtsheim. "The Coordinate Exchange Algorithm for Constructing Exact Optimal Experimental Designs." *Technometrics* 37 (1995), pp. 60–69.

Monlezun, C. J. "Two-Dimensional Plots for Interpreting Interactions in the Three-Factor Analysis of Variance Model," *The American Statistician* 33 (1989), pp. 63–69.

Nelson, L. S. "Exact Critical Values for Use With the Analysis of Means." *Journal of Quality Technology* 15 (1983), pp. 40–44.

Nelson, P. R. "Additional Uses for the Analysis of Means and Extended Tables of Critical Values." *Technometrics* 35 (1993), pp. 61–71.

Ott, E. R. "Analysis of Means—A Graphical Procedure." *Industrial Quality Control* 24 (1967), pp. 101–109.

Plackett, R. L., and J. P. Burman. "The Design of Optimum Multifactorial Experiments." *Biometrika* 33 (1946), pp. 305–25.

Puri, M. L., and P. K. Sen. *Nonparametric Methods in General Linear Models*. New York: John Wiley & Sons, 1985.

Satterthwaite, F. E. "An Approximate Distribution of Estimates of Variance Components." *Biometrics Bulletin* 2 (1946), pp. 110–14.

Schwarz, C. J. "The Mixed-Model ANOVA: The Truth, the Computer Packages, the Books. Part I: Balanced Data." *The American Statistician* 47 (1993), pp. 48–59.

Shaffer, J. P. "Modified Sequentially Rejective Multiple Test Procedures." *Journal of the American Statistical Association* 81 (1986), pp. 826–31.

Shoemaker, A. C.; K. L. Tsui; and C. F. J. Wu. "Economical Experimentation Methods for Robust Design." *Technometrics* 33 (1991), pp. 415–27.

Snee, R. D. "Computer-Aided Design of Experiments—Some Practical Experiences." *Journal of Quality Technology* 17 (1985), pp. 222–36.

Taguchi, G. *Introduction to Quality Engineering*. Tokyo: Asian Productivity Organization, 1986.

Ting, N.; R. K. Burdick; F. A. Graybill; S. Jeyaratnam; and T. F. C. Lu. "Confidence Intervals on Linear Combinations of Variance Components that Are Unrestricted in Sign." *Journal of Statistical Computation and Simulation* 35 (1990), pp. 135–43.

Welch, W. J.; T. K. Yu; S. M. Kang; and J. Sacks. "Computer Experiments for Quality Control by Parameter Designs." *Journal of Quality Technology* 40 (1990), pp. 62–71.

Index

A

ABT Electronics Corporation, 783
Activation functions, 538
Active explanatory factors, 1209
Added-variable plots, 384–390
Addition theorem, 1298
Additive effects, 216
Additive factor effects, 819–822
Additive model for random block effects, 1061–1064
Adjusted coefficient of multiple determination, 226–227
Adjusted estimated treatment mean, 932
Adjusted variable plots (*see* Added-variable plots)
Adjustment factors, 1250
Akaike's information criterion (AIC_p), 359–360
Algorithms, 361–364
Aliased, 1225
Aligned dot plots, 777, 778
Allocated codes, 321–322
America's Smallest School: The Family, 441
Analysis of covariance, 658, 917–920
  alternative to blocking, 939
  correction for bias, 940
  estimation of effects, 930–932, 940
  F tests, 928–929
  models
    appropriateness, 925–926
    multifactor, 934–937
    single-factor, 920–925
    two-factor, 934
  parallel slopes test, 932–933
  randomized block design, 937–938
  regression approach
    multifactor, 934–935
    single-factor, 924–925
  uses of differences, 939–940
Analysis of means, 758–759
Analysis of means plot, 758–759
Analysis of variance (ANOVA), 679–681, 833–842
  coefficient of multiple correlation, 227
  coefficient of multiple determination, 226–227
  concomitant variables, 919–920
  degrees of freedom, 66
  empty cells, 964–967
  estimation of effects
    latin square design, 1190
    nested design, 1100–1104
    quantitative factor, 762–766
    repeated measures design, 1137–1138, 1157–1158
      with interaction, 1148–1152
      without interaction, 1145–1148
    single-factor, 762–766
    three-factor, 1013–1017, 1069–1070
    two-factor, 848–861, 959–964, 970–980, 1055–1060
    two-level factorial design, 1212–1214
  estimation of factor level means, 737–761
  expected mean squares, 68–69
  F tests, 69–71, 226
    latin square design, 1190
    multiple pairwise testing procedure, 1138–1139
    nested design, 1097–1099
    power value charts, 1337–1341
    randomized block design, 898–901, 1138–1139
    repeated measures design, 1130–1134, 1138–1139, 1142–1143, 1155–1157
    single-factor, 698–701, 716–718, 744, 795–798
    single-factor ANOVA model, 704
    three-factor, 1009–1010, 1067–1068
    two-factor, 843–847, 1053–1054
    two-level factorial design, 1214–1215
  mean squares, 66–67, 225
  no replications and/or some interactions equal zero, 1366
  partitioning
    latin square design, 1188–1189
    nested design, 1093–1099
    randomized block design, 898–900, 908–909
    repeated measures design, 1130–1134, 1138–1139, 1142–1143, 1155–1157
    single-factor ANOVA model, 690–693
    three-factor, 1008–1009
    two-factor, 836–840
  partition of sum of squares, 63–66
  planning of sample size
    equal sample sizes, 759–761
    estimation approach, 759–761
    to find "best" treatment, 721–722
    latin square design, 1193–1194
    power approach, 716–723
    randomized block design, 909–912, 939
    single-factor, 716–718
    tables, 1342–1344
    unequal sample sizes, 761
  regression approach
    randomized block design, 967–969
    repeated measures design, 1161
    single-factor ANOVA model, 704–712
    three-factor, 1019–1020
    two-factor, 953–959
  rule for finding expected mean squares, 1361–1365
  rule for finding sums of squares and degrees of freedom, 1359–1361
  statistical computing packages, 980–981
  sums of squares, 204–206, 225
  unequal sample sizes, 761, 1019–1021, 1070–1077
  unequal treatment importance, 970–980
Analysis of variance (ANOVA) models, 329, 642, 659, 679–681
  diagnosis of departures from, 778–781
  effects of departures from model, 793–795
  factor effects model
    with unweighted mean, 705–708
    with weighted mean, 709–710
  fitting of, 685–689, 1003–1011
  meaning of model elements, 817–829
  no-interaction model, 880–886
  randomized block design, 1061–1065
  rule for model development, 1358–1359
  single-factor, 681–685, 692–704
    model II, 1030–1034, 1047
    model I vs. model II, 685
    repeated measures design, 1129–1139
    residual analysis for aptness, 775–781
    subsampling, 1106–1113
    weighted least squares, 786–789
  tests for constancy of error variance, 780–785
  three-factor
    ANOVA model, 992–998
    fitting of model, 1003–1005
    model II, 1066
    model III, 1066–1067
    partially nested design, 1114–1119
    residual analysis for appropriateness, 1006, 1007

Analysis of variance models (*cont.*)
transformations of response variables, 789–793
treatment means comparisons, 856–861
two-factor
crossed, 1366–1369
estimation of effects, 848–861
fitting of model, 834–836
fixed-factor levels, 829–833
model II, 1047–1049
model III, 1049–1052
pooling sums of squares, 861–862
repeated measures design, 1153–1161
residual analysis, 842–843
strategy for analysis, 847–858
Tukey test for additivity, 886–888
two-level factorial design, 1210–1212
two-level fractional factorial design, 1223–1239
unbalanced nested design, 1104–1106
unequal sample sizes, 951–964
Analysis of variance table, 67–68, 120, 124–127, 225, 261, 262, 694, 840–842
ANOVA (*see* Analysis of variance)
ANOVA models (*See* Analysis of variance models)
ANOVA table (*see* Analysis of variance table)
Antagonistic interaction effect, 308
A-optimality, 1282
Apex Enterprises, 1031–1034
Arc sine transformation, 790
Asymptotic normality, 50
Autocorrelation, 481–501
Durbin-Watson test, 487–490
parameters, 484
remedial measures, 490–498
Autocorrelation function, 485
Autocovariance function, 485
Automatic model selection, 582–585
Automatic search procedures, 361–369
Autoregressive error model (*see* First-order autoregressive error model)
Axial points, 1269

B

Back-propagation, 543
Backward elimination selection procedure, 368
Balanced incomplete block designs (BIBDs), 664–665, 1173–1183
advantages and disadvantages of, 1175–1176
analysis of, 1177–1183
table, 1345–1347
Balanced nested design, 1088–1091
Bar graphs, 736, 853
Bar-interval graph, 738–739
Baseline (referent) category, 611
Bates, D. M., 529
Bechhofer procedure, 721
Berkson, J., 167–168, 172
Berkson model, 167–168
Bernoulli random variables, 563
"Best" subsets algorithms, 361–364
"Best" treatment, 721–722, 864
Beta coefficients, 278
Bias correction, 940
Biased estimation, 369, 432–433
BIBDs (*See* Balanced incomplete block designs)
Bilinear interaction term, 306
Binary variables, 315
outcome variables, 556–557
response variables, 555–563 (*See also* Indicator variables)
Binomial distribution, 569
Bisquare weight function, 439–440
Bivariate normal distribution, 78–80
Blind studies, 658
Blocked experiments, 656
Blocking, 656–658
and cause-and-effect inferences, 895
covariance analysis alternative, 939
criteria for, 893–894
Blocks, 656, 892
BMDP, 107, 694, 695
BMDP3V, 1075
Bonferroni inequality, 155–156, 1215
Bonferroni joint estimation procedure, 287, 396–398
analysis of covariance, 930, 931, 934
confidence coefficient, 155–157
latin square design, 1190
logistic regression, 580
mean responses, 159, 230
multiple pairwise testing, 1038–1039
nested design, 1101, 1102
nonlinear regression, 532
prediction of new observations, 160–161, 231
randomized block design, 904
regression coefficients, 228
repeated measures design, 1157
single-factor ANOVA model, 756–758
three-factor analysis of variance, 1015, 1017, 1074
two-factor analysis of variance, 851, 852, 856–857
Bootstrapping, 458–464, 529, 530, 1075–1076
Bounded influence regression methods, 449
Box, G. E. P., 236
Box, M. J., 529
Box-Behnken designs, 1276
Box-Cox transformation, 134–137, 142, 236, 791–793
Box plots, 102, 108, 110, 779, 781
Breiman, L., 369
Breusch-Pagan test, 115, 118–119, 142–144, 234–235
Brown-Forsythe test, 115, 116–118, 234
Brushing, 233
Bubble (proportional influence) plot, 600

C

Calibration problem, 170
Caliper (interval) matching, 669
Candidate list, 1278
Carryover effect, 1128, 1201
Case, 4
Case-control studies (*See* Retrospective studies)
Castle Bakery Company, 833–838
Causality, 8–9
CDI data set, 1349–1350
Cell means model, 681–682, 704, 710–712, 830–831, 834, 996–997
Centering, 272
Center point, 1222, 1269
Center point replications, 1222–1223
Central composite designs, 1268–1276
Central limit theorem, 1302
Centroid, 398
Changeover design, 1198
Cheng, C., 464
Cleveland, W S., 138, 146, 450
Close-to-linear nonlinear regression estimates, 529
Cochrane-Orcutt procedure, 492–495
Cochran's theorem, 69–70, 699, 843
Coefficient(s):
of correlation, 76, 80
of determination, 74–76, 86–87
of multiple correlation, 227
of multiple determination, 226–227
of partial correlation, 270–271
of partial determination, 268–271
of simple correlation, 227 (*See also* Multiple regression coefficients; Simple linear regression coefficients)
Cohort studies, 667
Column sum of squares, 1189
Column vector, 178
Comparative experimental studies, 643, 644
Comparative observational studies, 644–645
Complementary events, 1298
Complementary log-log transformation, 562–563, 568
Complete block design, 1183

Completely randomized design, 13, 644, 659–660
Completely randomized factorial design, 660
Complete replicates, 653
Components of variance model, 1033
Compound symmetry, 1062
Concomitant variables, 919–920
Concordance index, 607
Conditional effects plot, 307, 1285
Conditional probability distributions, 80–83
Confidence band, for regression line, 61–63
Confidence coefficient, 46, 49–50, 54–55, 744–745
  Bonferroni procedure, 155–157
  family, 154–155
  and risk of errors, 50
Confidence intervals, bootstrap, 460
Confirmatory experiments, 1209
Confounding, 1224–1227
  worst-case degree of, 1232
Confounding factors, 656
Confounding scheme, 1226
Conjugate gradient method, 543
Consistent estimator, 1305
Constancy of error variance, tests for, 115–119, 234
Constrained randomization, 655–658
Contour diagram, 1284–1286
Contrast, 741–742
Control factors, 1251
Control group, 643
Controlled experiments, 343–344, 347
Control treatment, 651–652
Control variables, 344, 919
Cook's distance, 402–405, 598–601
Corner points, 1269
Correlation coefficients, 78–89, 1301
  and bivariate normal distribution, 78–80
  and conditional inferences, 80–83
  inferences on, 83–87
  and regression vs. correlation models, 78
  Spearman rank, 87–88
  table of critical values, 1329
  table of z' transformation, 1332
Correlation matrix of transformed variables, 274–275
Correlation models:
  bivariate normal distribution, 78–80
  compared to regression models, 78
  conditional probability distributions, 80–83
  multivariate normal distribution, 196–197
  regression analysis, 82–83 (*See also* Regression models)
Correlation operator, 1301
Correlation test for normality, 115, 234
Correlation transformation, 272–273
Covariance, 1300–1301 (*See also* Variance-covariance matrix)
Covariance analysis (*see* Analysis of covariance)
Covariance models, 329 (*See also* Analysis of covariance)
Covariance operator, 1300
Covariates, 919
Cox, D. R., 171, 172
C_p criterion, Mallows', 357–359
Crossed factor, 648, 1088–1091
Crossed-nested design, 662–663, 1114
Crossed two-factor study, 1366–1369
Crossover designs, 1198–1200
Cross-sectional studies, 666–667
Cross-validation, 372
Cumulative logits, 616
Curvilinear relationship, 4
Cutoff point, 604

D

Data collection, 343–346, 370–371
Data sets, 1348–1357
Data snooping, 745
Data splitting, 372
Decision rule, 70
Defining relation, 1227–1228
Degree of linear association, 74–77
Degrees of freedom, 66, 693, 839, 1009, 1095, 1303–1304
  rule for finding, 1359–1361
Deleted residuals, 395–396
Delta chi-square statistic, 598
Delta deviance statistic, 598
Density function, 1073
  of normal random variable, 1302
Dependent variables, 2–3, 3–4
Derived predictor values, 537
Design criterion, 1278
Design generators, 1229
Design matrix, 1212
Design of experiments, 647
Design resolution, 1231–1232
Determinant criterion, 1279–1280
Determinant of matrix, 190
Deviance, 589
Deviance goodness of fit test, 588–590
DFBETAS, 404–405
DFFITS, 401–402
Diagnostic plots, 901–903
  for predictor variables, 100–102
  for residuals, 103–114
Diagnostics (*see* Logistic regression diagnostics)
Diagonal matrix, 185–186
Dichotomous responses, 556 (*See also* Binary variables)
Differences of treatment means, 1002
Direct numerical search, 518–524
Discriminant analysis, 608
Disease outbreak data set, 1355
Disordinal interaction, 326
Dispersion model, 1246–1247
Disturbances, 482
Dixon, W. J., 33
D-optimal design, 1279
Dorle Exterior Trim, 1288–1289
Dose-response relationships, 510
Dot plots, 100–101, 108, 110, 777, 778, 781
  of residuals, 778
Double-blind study, 658
Double crossover design, 1201
Double cross-validation procedure, 375
Drug effect experiment data set, 1356–1357
Dummy variables, 12, 315 (*See also* Indicator variables)
Durbin-Watson test, 114, 487–490, 492
  table of test bounds, 1330–1331

E

Educational Testing Service, 441
Efron, B., 459
Eigenvalues, 1287
Empty cells, 964–967
Error mean square, 66, 126
Error sum of squares, 25, 64, 72, 126, 691, 836–837
Error terms, 9–12, 778–779, 794–795
  autocorrelated, 481–484
  constancy of variance tests, 116–119
  nonindependence of, 108–110, 128
  nonnormality, 110–112, 128–129
Error term variance, 24–26, 527–528
Error variability reduction, 917–918
Error variance, 107, 128, 421–431, 778, 793–794
Estimates (*see* Tests)
Estimation, 737–738
Estimation approach to sample size planning, 759–761, 863–864, 1182–1183
Estimators, 1305–1306
Exact F test, 1067–1068
Exchange algorithms, 1282
Expectation operator, 1299
Expected mean squares, 68–69, 694–698, 840, 1052–1053
  rule for finding, 1361–1365
Expected value, of random variable, 1299
Experimental data, 13
  factor level means, 684

Experimental designs, 647
blocking, 656–658
completely randomized design, 13, 659–660
crossed-nested, 662–663
crossover, 1198–1200
double crossover, 1201
exploratory, 1209
factorial experiment, 660–661
fractional factorial design, 665–666
latin square, 1183–1186
nested, 662–663, 1088–1091
one-factor-at-a-time approach, 815, 816
randomization tests, 712–715
randomized block, 892–896
randomized complete block, 661–662
repeated measures, 663–664, 1127–1129
response surface design, 666
response surface methodology, 1267–1268
screening designs, 1239–1240
sequential search for optimal conditions, 1290–1292
split-plot design, 664, 1162–1163
two-level factorial design, 665–666, 1210–1212
two-level fractional factorial design, 1223–1239
Experimental error sums of squares, 1108–1109
Experimental factors, 644, 647
Experimental group, 643
Experimental studies, 643–644
mixed observational and, 646–647
observational vs., 677–679
Experimental units, 13, 643, 652, 893–894, 1112
Experiments, 643
Explanatory variables, 3, 347–349
omission of, 780–781
Exponential family of probability distributions, 623
Exponential regression function, 128
Exponential regression model, 511–512
Ex post facto studies (*See* Retrospective studies)
Externally studentized residuals, 396
Extra sums of squares, 256–262, 285–286
Extreme value (Gumbel density function), 562

F

Face-centered design, 1273
Factor *A* main effects, 844
Factor *A* sum of squares, 838
Factor *B* main effects, 845
Factor *B* sum of squares, 838
Factor effects model, 701–704, 831–833, 835–836, 997–998
with unweighted mean, 705–708
with weighted mean, 709–710
Factorial experiments, 660–661
Factor level, 647–648
Factor level means, 684, 698–701, 704, 818
estimation, 848–853
estimation and testing, 737–761
plots of, 735–737
line plot, 735–736
main effects plot, 736–737
weighted least squares estimation, 786–789
Factors, 344, 647
and choice of treatment, 649–652
crossed, 648
nested, 649
nuisance (confounding), 656
(*See also* Control variables)
Family:
of conclusions, 1013
of estimates, 154–155
of tests, 745, 846, 1010
Family confidence coefficient, 154–155
Far-from-linear nonlinear regression estimates, 529
F distribution, 1304
Scheffé procedure, 160–161
table of percentiles, 1320–1326
First differences procedure, 496–498
First-order autoregressive error model, 484–487
Cochrane-Orcutt procedure, 492–495
Durbin-Watson test, 487–490
first differences procedure, 496–498
forecasting with, 499–501
Hildreth-Lu procedure, 495–496
First-order interactions, 995
First-order regression model, 9, 215–217, 318–319 (*See also* Regression models)
Fish-bone diagrams, 649
Fisher Company, 1272–1274
Fisher *z* transformation, 85
Fitted logit response function, 565
Fitted values, 202–203, 688, 835, 1004
influences on, 401–404
and multicollinearity, 286–288
and residuals, 224–225
total mean square error, 357–359
Fitting, 298–299
of ANOVA model, 685–689, 1003–1011
Fixed effects contrasts, 1058
Fixed effects model, 685
Fixed *X* sampling, 459
Folding over, 1240
Foldover design, 1240
Forecasting with autoregressive error model, 499–501
Forward selection procedure, 368
Forward stepwise regression, 364–367
Fraction, 1209
Fractional factorial designs, 665–666, 1209
Frequencies, proportional, 980
Friedman test, 900–901, 1138
F test:
for analysis of variance, 69–71
equivalence of *t* test, 71
for lack of fit, 119–127, 235
nonparametric rank, 795–798
for regression relation, 226
Full model, 72, 121–123, 700, 711–712
Functional relation, 2–3

G

Galton, Francis, 5
Gauss-Markov theorem, 18, 43, 884
Gauss-Newton method, 518–524
Generalized interaction, 1230
Generalized least squares, 430
Generalized randomized block design, 906–908
Generalized randomized block model, 907
General linear regression models, 217–221, 510–511, 623–624
General linear test, 72–73, 121–127
approach, 972–974
Goodness of fit tests, 586–590
deviance, 588–590
Hosmer-Lemeshow, 589–590
Pearson chi-square, 586–588, 590
G-optimality, 1282
Gulledge, T. R., Jr., 115, 146, 1329
Gumbel density function (extreme value), 562

H

Half-fraction design, 1229
Half-normal probability plot, 595–598, 1222
Hartley test, 782–784, 1144
Hat matrix, 202–203, 392–394, 398–400
Heating equipment data set, 1353–1354
Hessian matrix, 578, 1074
Heteroscedasticity, 429
Hidden nodes, 540
Hidden replication, 816
Hierarchical fitting, 298–299
Hildreth-Lu procedure, 495–496
Histograms, 110, 778, 781
Holm simultaneous testing procedure, 850
Homoscedasticity, 429

Honestly significant difference tests, 752
Hosmer-Lemeshow goodness of fit test, 589–590
Hougaard, P., 529
H statistic, 782
table of percentiles, 1336
Huber weight function, 439–440
Hyperplane, 217

I

IC Technologies, 1282–1283
Idempotent matrix, 203
Identity matrix, 186
Important interactions, 824–825, 1016
Incomplete block designs, 664–665, 1183
two-level factorial, 1240–1244
Independent random variables, 1302
Independent samples, 1309–1311
Independent variables, 2–3, 3–4
Index of response, 939
Indicator variables, 314–315
allocated codes vs., 321–322
alternative codings, 323–324
in analysis of variance, 680–681
for comparing regression functions, 329–335
interaction effects, 324–327
quantitative variables vs., 322–323
time series applications, 319–321
Individual outcome, 56
Individual test, 745
Influential cases, 400–406
Influential observations, detection of, 598–601
Instrumental variables, 167
Interaction effect coefficient, 297
Interaction effects, 220
with indicator variables, 324–327
interference/antagonistic, 308
reinforcement type, 308
Interaction model for random block effects, 1064–1065
Interaction regression models, 306–313
Interactions, 823
in analysis of variance, 822–829
generalized, 1230
multiple two-factor, 999–1000
single two-factor, 1000–1002
tests for, 844
three-factor, 996, 998–999, 1016
two-factor, 856–861
two-level factorial design, 1218–1219
Interaction sum of squares, 838
Interaction sum of squares between blocks and treatments, 898
Interaction sum of squares between treatments and subjects, 1130
Intercorrelation (*see* Multicollinearity)
Interference interaction effect, 308
Internally studentized residuals, 394
Interval (caliper) matching, 669
Interval plot, 738
Intraclass correlation coefficient, 1035
Intrinsically linear response functions, 514
Inverse of matrix, 189–193
Inverse predictions, 168–170
Inverse regression, 170
Iowa Aluminum Corporation, 1233–1239
IPO data set, 1355–1356
IRLS robust regression, 439–441
Irregular experimental regions, 1276–1277
Ischemic heart disease data set, 1354–1355
Ishakawa diagrams, 649
Iteratively reweighted least squares, 426

J

JMP, 981
$J-1$ nominal response logits, 610–614
Joint density function, 1073
Joint estimation, 154–157
Joint probability function, 1300

K

Kendall's coefficient of concordance, 1139
Kendall's τ, 89
Kenton Food Company, 694–695, 712
Kimball inequality, 846, 1010, 1011, 1215
Kolmogorov-Smirnov test, 115
Kruskal-Wallis rank test, 796–797
Kurtosis, 793

L

Lack of fit mean square, 124
Lack of fit test, 119–127, 235, 764–766, 1222–1223
LAD (least absolute deviations) regression, 438
Large-sample theory, 528–530
LAR (least absolute residuals) regression, 438
Latent explanatory variables, 348
Latin square changeover design, 1198
Latin square design, 1183–1186
ANOVA partitioning, 1188–1189
crossover design, 1198–1200
double crossover design, 1201
efficiency, 1193–1194
estimation of effects, 1190
factorial treatment, 1192
fitting of model, 1188
F test, 1190
model, 1187
notation, 1188
planning of sample sizes, 1193–1194
random blocking effects, 1193
randomization, 1185–1186
repeated measures, 1198–1201
replications, 1193
replications within cells, 1195–1196
residual analysis, 1191
rule modification, 1366
sums of squares, 1188–1189
table, 1344
Tukey test for additivity, 1191–1192
use of independent squares, 1200–1201
use of several squares, 1196–1198
Learning curve models, 533–537
Least absolute deviations (LAD) regression, 438
Least squares estimation, 161–162, 1305–1306
criterion, 15–19
generalized, 430
and maximum likelihood estimation, 32–33
multiple regression, 223–224
penalized, 436
randomized complete block model, 898
simple linear regression, 199–201
single-factor ANOVA model, 687–689
standardized regression coefficients, 275–278
three-factor analysis of variance, 1003–1005
two-factor analysis of variance, 834–836, 975–976
weighted, 421–431
Levene test (*see* Modified Levene test)
Leverage, 398
Likelihood function, 29–33, 564, 1305
Likelihood ratio test, 580–582
Likelihood value, 28
Lilliefors test, 115
Linear-by-linear interaction term, 306
Linear combination of factor level means, 744
Linear dependence, 188
Linear effect coefficient, 296
Linear function of normal random variable, 1302
Linear independence, 188
Linearity, test for, 119–127
Linearization method (*see* Gauss-Newton method)
Linear model, 221
ANOVA model, 683–684
Linear predictor, 560, 623
Linear regression functions, 7

Line plot of estimated factor level means, 735–736
Link function, 623–624
LMS (least median of squares) regression, 439
Locally weighted regression scatter plot smoothing, 138–139
Location model, 1247, 1250
Logistic mean response function, 560–562
Logistic regression, polytomous, 608–618
Logistic regression diagnostics, 591–601
  influential observations, detection of, 598–601
  plots, residual, 594–598
  residuals, 591–594
Logistic regression models, 512–513 (*See also* Regression models)
Logit response function, 562
Logit transformation, 562
Looney, S. W., 115, 146, 1329
Lowess method, 138–139, 449–450

M

Main effects, 818–819, 1012
Main effects plot of estimated factor level means, 736–737
Mallows' C_p criterion, 357–359
Marginal probability function, 1300
Market share data set, 1350
Marquardt algorithm, 525
Matched-pairs design, 669
Matched studies, 668–669
Matching, 668–669
Mathematics proficiency, 441–448
Matrix(-ces):
  addition, 180–181
  with all elements unity, 187
  basic theorems, 193
  definition, 176–178
  determinant, 190
  diagonal, 185–186
  dimension, 176–177
  elements, 176–177
  equality of two, 179–180
  hat, 202–203, 392–394, 398–400
  Hessian, 578
  idempotent, 203
  identity, 186
  inverse, 189–193
  linear dependence, 188
  multiplication by matrix, 182–185
  multiplication by scalar, 182
  nonsingular, 190
  of quadratic form, 205–206
  random, 193–196
  rank, 188–189
  scalar, 187
  scatter plot, 232–233
  simple linear regression model, 197–199
  singular, 190
  square, 178
  subtraction, 180–181
  symmetric, 185
  transpose, 178–179
  vector, 178
  zero vector, 187
Maximum likelihood estimation, 27–33, 612–614, 617, 1305
  logistic regression, 564–567
  mixed ANOVA models, 1072–1076
  Poisson regression model, 620
  single-factor ANOVA model, 687–689
Mean:
  of the distribution, 56
  prediction of, 60–61
  of residuals, 102
Mean response:
  interval estimation, 52
  logistic regression
    interval estimation, 602–603
    point estimation, 602–604
  multiple regression
    estimation, 229–232
    joint estimation, 230
  simple linear regression
    interval estimation, 52–55, 157–159, 208–209
    joint estimation, 157–159
    point estimation, 21–22
Mean squared error:
  of regression coefficient, 433
  total, of n fitted values, 357–359
Mean square prediction error, 370–371
Mean squares, 25, 66–67, 693–694, 839–840, 1009
  analysis of variance, 225
  expected, 68–69
Measurement errors in observation, 165–168
Median absolute deviation (MAD), 440–441
Method of steepest descent, 525
Minimax, 1285
Minimum L_1-norm regression, 438
Minimum variance estimator, 1305
MINITAB, 20–21, 46, 47, 49–50, 101, 104, 671, 777, 840, 981, 1117, 1249
MINITAB Fractional Factorial procedure, 1235–1238
Minnesota Department of Transportation, 464–471
Mixed experimental and observational studies, 646–647
Mixed factor effects model, 1049–1052
MLS procedure, 1045–1047
Model-building set, 372
Modified large sample procedure (*see* MLS procedure)
Modified Levene test, 115, 116–118, 784–785, 1144
Modified Levene test statistic, 234
Moving average method, 137
MSE_p criterion, 355–356
Multicategory logistic regression models (*see* Polytomous logistic regression)
Multicollinearity, 278–289
  detection of, 406–410
  remedial measures, 431–437
  ridge regression, 431–437
Multifactor covariance analysis, 934–937
Multifactor studies, 648
  sample size planning, 1021–1022
  unequal sample sizes, 1019–1021
Multiple comparison procedures, 746–759, 1059
Multiple logistic regression, 570–577
  geometric interpretation, 572–573
  model, 570–573
  polynomial logistic regression, 575–576
  prediction of new observation, 604–608
Multiple pairwise comparisons, 797–798, 850–851, 856–861
Multiple pairwise testing procedure, 1138–1139
Multiple regression (*see* Mean response; Prediction of new observation; Regression coefficients; Regression function)
Multiple regression coefficients, 216–217
  danger in simultaneous tests, 287–288
  interval estimation, 228, 229
  joint inferences, 228
  least squares estimation, 223–224
  tests concerning, 228, 263–268
  variance-covariance matrix of, 227–228
Multiple regression mean response:
  estimation, 229–232
  joint estimation, 230
Multiple regression models, 214–221
  ANOVA table, 225
  diagnostics, 232–236
  extra sum of squares, 256–262
  general model, 217–221
  interaction effects, 220
  logistic regression, 570–577
  lowess method, 449–450
  in matrix terms, 222–223
  multicollinearity effects, 278–289
  remedial measures, 236
  standardized, 271–278
  two predictor variables, 236–248
Multiplication theorem, 1298
Multivariate normal distribution, 196–197

N

Nested design, 662–663, 1088–1091
balanced, 1088–1091
residual analysis, 1100
rule for model development, 1091–1092
subsampling, 1106–1114
three-factor partially nested, 1114–1119
two-factor
ANOVA partitioning, 1093–1099
estimation of effects, 1100–1104
fitting of model, 1093
F test, 1097–1099
model, 1091–1092
residual analysis, 1100
unbalanced, 1104–1106
Nested factor, 649, 1088–1091
Neural networks, 537–547
conditional effects plots, 546
example illustrating, 543–546
as generalization of linear regression, 541
network representation, 540–541
and penalized least squares, 542–543
single-hidden-layer, feedforward, 537
training the network, 542
No-interaction model, 880–886
Noise factors, 1246, 1250–1252
Noncentral F distribution, 70, 699
Noncentrality measure, 51
Noncentrality parameter, 717
Nonconstant error variance, 557–558, 778
Nonindependence of error terms, 778–779, 794–795
Nonindependent residuals, 102–103
Nonlinear regression models, 511–512
transformations for, 129–132 (*See also* Regression models)
Nonnormal error terms, 557
Nonnormality, 793–794
of error terms, 781
transformations for, 132–134
Nonparametric rank F test, 795–798, 900–901, 1138–1139
Nonparametric regression, 449–458
Nonparametric regression curves, 137
Nonsingular matrix, 190
Nonstandard models, 1277
Nonstandard sample sizes, 1278
Nontransformable interactions, 826–827
Normal equations, 17–18, 271–272, 517–518
Normal error regression model, 26–33, 82
confidence band for regression line, 61–63
inferences concerning β_0, 48–51
sampling distribution of b_1, 41–46
X and Y random, 78–89
Normality:
assessing, 112
correlation test for, 115
tests for, 115
Normal population:
one population mean, 1306–1308
population variance, 1311–1312
two population means, 1309–1311
two population variances, 1312–1314
Normal probability distribution, 1302–1303
Normal probability plot, 110–112, 781, 1221
of residuals, 778
Nuisance factors, 656
Numerator degrees of freedom, 1304

O

Observational data, 12–13
Observational factors, 645, 647
Observational studies, 344–345, 347–349, 368–369, 644–646
cross-sectional studies, 666–667
design of, 666
experimental vs., 677–679
factor level means, 684
mixed experimental and, 646–647
prospective (cohort) studies, 667
retrospective studies, 667–668
Observation units, 1112
Observed value, 21
Odds ratio, 562, 567
One-factor-at-a-time (OFAAT) approach to experimentation, 815, 816
One-sided test, 47–48
Optimal response surface design, 1276–1283
Optimum conditions, 1290–1292
Order effect, 1128
Order position sum of squares, 1199
Ordinal interaction, 326
Orthogonal coding, 1214
Orthogonal decomposition, 838
Orthogonality, 1273, 1275
Orthogonally blocked design, 1276
Orthogonal polynomials, 305
Ott, E. R., 758
Outliers:
detection of, 779–780
observations, outlying, 108–109, 129
tests for, 115, 396–398
Outlying cases, 390–391, 437–438
Overall F test, 264, 266

P

Paired-comparison design, 669
Paired comparison plot, 748
Paired observations, 1311
Pair-wise comparison, 739, 746, 797–798, 850–851, 856–861, 962–964, 1182
Pareto plot, 1219–1220
Partial F test, 264, 267–268
Partially hierarchical nested design, 1114
Partially nested design, 1114
Partial regression coefficients, 216
Partial regression plots (*see* Added-variable plots)
Partitioning:
degrees of freedom, 66
sum of squares total, 63–66
Path of steepest ascent/descent, 1290
Pattern sum of squares, 1199
Pearson chi-square goodness of fit test, 586–588, 590
Pearson product-moment correlation coefficient, 84, 87
Pearson semistudentized residuals, 591
Pearson studentized residuals, 592
Pecos Foods Corporation, 1216–1222
Penalized least squares, 436, 542–543
Penalty weight, 541
Permutation test, 714
Plackett-Burman designs, 1240
Plots of estimated factor level means, 735–737
line plot, 735–736
main effects plot, 736–737
Plots of residuals against fitted values, 778, 779
Plutonium measurement, 141–144
Point clouds, 233
Point estimators, 17, 21–22, 24–26, 602, 1056–1057
Poisson regression model, 618–623
Polynomial logistic regression, 575–576
Polynomial regression model, 219–220, 294–305
Polytomous (multicategory) logistic regression:
for nominal response, 608–614
for ordinal response, 614–618
Pooling sums of squares, 861–862
Population of consumers, 970
Power approach to sample planning, 716–723, 862–863
Power of tests:
latin square design, 1193
planning of sample size, 716–717
randomized block design, 909–910
regression coefficients, 50–51, 228
Power transformations, 134, 135
Prediction error rate, 607–608
Prediction interval, 57–60
Prediction of mean, 60–61
Prediction of new observation:
logistic regression, 604–608
multiple regression, 231
simple linear regression, 209

Prediction of new observation (*cont.*)
simple linear regression model, 55–61
Prediction set, 372
Predictor variables, 3–4, 6–7
added-variable plots, 384–390
in analysis of variance, 679–681
diagnostics, 100–102
measurement errors, 165–168
multicollinearity, 278–289
multiple regression, 214–217, 236–248
omission of important, 129
polynomial regression, 295–298
qualitative, 218–219, 313–318
residuals and omission of, 112–114
$PRESS_p$ criterion, 360–361
Primary variables, 344
Principal components regression, 432
Probability distribution, 50, 52–53
in single-factor ANOVA model, 681
Probability theorems, 1298
Probit mean response function, 559–560
Probit response function, 560, 568
Product operator, 1298
Projection property, 1232
Proportional frequencies, 980
Proportional influence (bubble) plot, 600
Proportionality constant, 424
Proportional odds model, 615, 616
Prospective studies, 667
Prostate cancer data set, 1351–1352
Pseudo F test statistic, 1068–1069
Pure error estimate, 1222
Pure error mean square, 124
Pure error sum of squares, 122
Puri, M. L., 798

Q

Quadratic effect coefficient, 296
Quadratic forms, 205–206
Quadratic linear regression model, 220–221
Quadratic regression function, 7, 128
Quadratic response function, 295, 305, 764–766
Qualitative factor, 647
Qualitative predictor variables, 218–219, 313–318
in analysis of variance, 679–681
more than one variable, 328
with more than two classes, 318–320
with two classes, 314–318
variables only, 329
Quantitative factors, 647, 762–766
Quantitative predictor variables, in analysis of variance, 679, 681
Quantitative variables, 322–323
Quarter-fraction design, 1229–1231
Quasi F test statistic, 1068–1069

R

R^2a criterion, 355–356
R^2p criterion, 354–356
Radial basis function, 541
Random ANOVA model, 1031
Random cell means model, 1031–1034
Random effects model, 685
Random factor effects model, 1047
Randomization, 643, 652–658, 895, 1128–1129
constrained, 655–658
latin square design, 1185–1186
restricted, 656
Randomization distribution, 713–714
Randomization tests, 712–715
Randomized complete block design, 661–662, 892–896
analysis of covariance, 937–938
ANOVA partitioning, 898–900, 908–909
appropriateness of model, 901–903
diagnostic plots, 901–903
estimation of effects, 904–905
factorial treatments, 908–909
fitting of model, 898
F test, 898–900, 900–901, 1138–1139
generalized design, 906–908
missing observations, 967–969
models, 897–898, 1061–1065
multiple pairwise testing procedure, 1138–1139
nonparametric rank F test, 900–901, 1138–1139
planning of sample sizes, 909–912, 939
random block effects, 1060–1065
rank data, 900–901, 1138–1139
regression approach, 938, 967–969
residual analysis, 901–903
subsampling, 1367, 1369–1370
Tukey test for additivity, 903–904
use of more than one blocking variable, 905–906
use of more than one replicate in each block, 906–908
Random matrix, 193–196
Randomness tests, 114
Random variables, 1299–1302
Random vector, 193–196
Random X sampling, 459
Rank correlation procedure, 87
Rank of matrix, 188–189
Real estate sales data set, 1353
Receiver operating characteristic (ROC) curve, 606
Reduced model, 700, 711–712
Reduced model general linear test, 123
Reduced model general test, 72
Referent (baseline) category, 611
Reflection method, 460
Regression:
and causality, 8–9
as term, 5
Regression analysis, 2
analysis of variance approach, 63–71
approach to balanced incomplete block designs, 1177–1179
approach to single-factor ANOVA model, 704–712
compared to analysis of variance, 679–681
completely randomized design, 13
computer calculations, 9
considerations in applying, 77–78
experimental data, 13
inferences concerning β_1, 40–48
observational data, 12–13
overview of steps, 13–15
transformations of variables, 129–137
uses, 8
Regression approach:
to analysis of covariance, 924–925, 934–935
to analysis of variance models
three-factor, 1019–1020
to randomized block design, 938, 967–969
to repeated measures design, 1161
two-factor, 953–959
Regression coefficients, 404–405
bootstrapping, 458–464
effects of multicollinearity, 284–285
interaction regression, 306–309
interpretation for predictor variables, 315–318, 324–327
lack of comparability, 272
multiple regression (*see* Multiple regression coefficients)
partial, 216
simple linear regression (*see* Simple linear regression coefficients)
standardized, 275–278
Regression curve, 6
Regression function, 6
comparison of two or more, 329–330
constraints, 558–559
estimation, 15–24
exploration of shape, 137–144
exponential, 128
hyperplane, 217
interaction models, 309
intrinsically linear, 514
nonlinearity, 104–107, 128
through origin, 161–165
outlying cases, 390–391

polynomial regression, 299–300
quadratic, 128
test for fit, 235
test for lack of fit, 119–127
test for regression relation, 226
Regression line, 61–63
Regression mean square, 66
Regression models:
autocorrelation problems, 481–484
basic concepts, 5–7
with binary response variable, 555–559
bootstrapping, 458–464
building, 343–350, 368–369
building process diagnostics, 384–414
choice of levels, 170–171
coefficient of correlation, 76
coefficient of determination, 74
compared to correlation models, 78
construction of, 7–8
degree of linear association, 74–77
effect of measurement errors, 165–168
estimation of error terms, 24–26
first-order autoregressive, 484–487
general linear, 121–127, 217–223, 510–511, 623–624
interaction, 306–313
inverse predictions, 168–170
logistic regression
mean response, 602–604
multiple, 570–577
parameters, 577–582
polytomous, 608–618
simple, 564–570
tests for, 586–601
multiple regression (*see* Multiple regression models)
nonlinear regression
building, 526–527
Gauss-Newton method, 518–524
learning curve, 533–537
least squares estimation, 515–525
logistic, 563–618
parameters, 527–533
Poisson, 618–623
normal error, X and Y random, 78–89
normal error terms, 26–33
origin, 5
overview of remedial measures, 127–129
Poisson regression, 618–623
polynomial, 219–220, 294–305
with qualitative predictor variables, 313–321
residual analysis, 102–115
scope of, 8
selection and validation, 343–375
automatic search procedures, 361–369
backward elimination, 368
criteria for model selection, 353–361
forward selection, 368
forward stepwise regression, 364–367
simple linear (*see* Simple linear regression models)
smoothing methods, 137–141
third order, 296
transformed variables, 220
validation of, 350, 369–375
Regression relation, functional form, 7–8
Regression sum of squares, 65, 260–262
Regression surface, 216, 229–230
Regression through origin, 161–165
Regression trees, 453–457
Reinforcement interaction effect, 308
Remainder sum of squares, 1366
Repeated measures design, 663–664, 894, 1127–1129
blocking of subjects in, 1153
estimation of effects, 1137–1138, 1145, 1157–1158
F test, 1130–1134, 1138–1139, 1142–1143, 1155–1157
latin square crossover design, 1198–1200
multiple pairwise testing procedure, 1138–1139
ranked data, 1138–1139
regression approach, 1161
repeated measures on both factors, 1153–1161
repeated measures on one factor, 1140–1153
residual analysis, 1134–1135, 1144, 1157
single-factor, 1129–1139
split-plot design, 1162–1163
Replicates, 120
Replication, 120, 426, 653
hidden, 816
Reproducibility, 653
Residuals, 203–204, 224–225
deleted, 395–396
departures from simple linear regression, 103
logistic regression diagnostics, 591–598
and omitted predictor variables, 112–114
outliers, 108–109
overview of tests, 114–115
properties, 102–103
in regression models, 22–24
scaled, 440–441
semistudentized, 103, 392, 591
studentized, 394
studentized deleted, 396–398
studentized Pearson, 592
variance-covariance matrix of, 203–204
Residual analysis, 102
analysis of variance, 775–781, 842–843, 1006, 1007
latin square design, 1191
nested design, 1100
randomized block design, 901–903
repeated measures design, 1134–1135, 1144, 1157
Residual dot plots, 779
Residual mean square, 25
Residual plots, 104–114, 233–234, 384–390
Residual plots against fitted values, 776–778, 780–781
Residuals, 689, 835, 1004
analysis of variance, 775–776
Residual sequence plot, 778–779
Residual sum of squares, 25
Response, 21
Response function, 764 (*See also* Regression function)
Response modeling approach, 1255
Response surface, 216, 309, 310
Response surface design, 666, 1267–1268
blocking central composite design, 1275–1276
central composite design, 1268–1276
design criteria, 1279–1281
model interpretation and visualization, 1284–1286
optimal, 1276–1283
optimal conditions, 1286–1287
rotatable central composite design, 1271–1272
Response variables, 3, 165–168, 555–563
transformations, 789–793
Restricted mixed factor effects model, 1049
Restricted randomization, 656
Retrospective studies, 667–668
Ridge regression, 431–437
Ridge trace, 434, 435, 437
Robust product design, 1244–1255
Robust regression, 437–449
Robust test, 794
ROC (receiver operating characteristic) curve, 606
Rotatable central composite design, 1271–1273
Rotatable inscribed central composite design, 1273
Roundoff errors, in normal equations calculations, 271–272
Row sum of squares, 1189
Row vector, 178
Running medians method, 137–138
Rutgers Experimental Station, 1277–1281

S

Saddle point, 1285
Sample size, 652–653
Sample size planning for analysis of variance:
  estimation approach, 759–761, 863–864, 1182–1183
  to find "best" treatment, 721–722, 864
  F test, 1021
  latin square design, 1193–1194
  multifactor studies, 1021–1022
  power approach, 716–723, 862–863
  random block design, 909–912, 939
  tables, 1342–1344
Sampling distribution, 44–46, 48–50, 52–54, 69–70
SAS PROC GLM, 981
SAS PROC MIXED, 1075
SAS PROC OPTEX, 1283
Satterthwaite approximate F test, 1068–1069
Satterthwaite procedure, 1043–1045
Saturated model, 588
SBC_p (Schwarz' Bayesian criterion), 359–360
Scalar, 182
Scalar matrix, 186, 187
Scaled residuals, 440–441
Scaling, 272
Scatter diagram/plot, 4, 19–21, 104–105
Scatter plot matrix, 232–233
Scheffé, Henry, 793
Scheffé joint estimation procedure, 930–931, 934
  latin square design, 1190
  nested design, 1101, 1102
  prediction of new observation, 160–161, 231
  randomized block design, 904
  repeated measures design, 1157
  single-factor analysis of variance, 761
  three-factor analysis of variance, 1015, 1017
  two-factor analysis of variance, 852, 857
Scheffé multiple comparison procedure, 794
  single-factor analysis of variance, 753–755
Schwarz' Bayesian criterion (SBC_p), 359–360
Scientific studies, statistical design of, 642
Scope of model, 8
Screening designs, 1239–1240
Second-order interaction, 996
Second-order regression model, 295–296, 297, 884
Selection and validation of models, 343–375
  automatic model selection, 582–585
  automatic search procedures, 361–369
  backward elimination, 368
  criteria for model selection, 353–361
  forward selection, 368
  forward stepwise regression, 364–367
Semistudentized residuals, 103, 392, 591, 776
Sen, P. K., 798
SENIC data set, 1348–1349
Sensitivity, 606
Sequence plot, 101, 108–109
Sequential experimental runs, 1290–1292
Serial correlation, 481
Servo-Data, Inc., 790–791
Shapiro-Wilk test, 116
Sheffield Foods Company, 1070–1076
Sigmoidal response functions, 538, 559–563
Signal-noise ratio, 1255
Simple linear regression coefficients, 11–12
  interval estimation, 45–47, 49–50, 52–55, 54–55
  least squares estimation, 15–19, 199–201
  point estimation, 21–22, 155–157
  tests for, 47–48, 50–51, 69–71
  variance-covariance matrix of, 207–208
Simple linear regression mean response:
  interval estimation, 52–55, 157–159, 208–209
  joint estimation, 157–159
  point estimation, 21–22
Simple linear regression models, 9–12
  ANOVA table, 67–68
  diagnostics for predictor variables, 100–102
  error term distribution unspecified, 9–12
  general test approach, 72–73
  interval estimation, 52–55
  joint estimation procedures, 154–161
  in matrix terms, 197–199
  normal error terms, 26–33
  through origin, 161–165
  prediction of new observation, 55–61
  regression coefficients, 11–12
  residual analysis, 102–115
  tests for coefficients, 47–48 (*See also* Regression models)
Simulated envelope, 596–598
Simultaneous estimation, 747
Simultaneous testing, 747–748
Single-blind study, 658
Single comparison procedure, 904
Single degree of freedom test, 744, 964
Single-factor ANOVA models, 681–685
  estimation of effects, 762–766
  factor effects model, 701–704
    with unweighted mean, 705–708
    with weighted mean, 709–710
  fitting of model, 685–689
  F tests, 698–701, 704
  least squares estimation, 687–689
  maximum likelihood estimation, 687–689
  model I vs. model II, 685
  partitioning of SSTO, 690–693
  residual analysis, 775–781
Single-factor study, 648
  analysis of covariance, 920–933
  estimation of effects, 737–761, 930–932
  expected mean squares, 694–698
  experimental vs. observational studies 677–679
  F tests, 716–718, 744, 795–798, 928–929
  model II, 1030–1034, 1047
  planning of sample sizes, 716–718, 718–720
  regression approach, 704–712
  repeated measures design, 1129–1139
  subsampling, 1106–1113
Single-hidden-layer, feedforward neural networks, 537
Single-layer perceptrons, 537
Singular matrix, 190
Smoothing methods, 137–141
Sparsity of effects principle, 1224
Spearman rank correlation coefficient, 87–89
Specificity, 606, 607
Spector, P., 369
Split-plot designs, 664, 1162–1163
SPSS ANOVA, 981
SPSS^x, 763
Square matrix, 178
SSE (*see* Error sum of squares)
SSTO (*see* Total sum of squares)
SSTR (*see* Treatment sum of squares)
Standard deviation, studentized statistic, 44
Standardized multiple regression model, 273
Standardized random variable, 1301
Standardized regression coefficients, 275–278
Standard latin squares, 1186
Standard normal distribution, table of cumulative probabilities, 1316
Standard normal random variable, 1302–1303
Standard order, 1211
Star points, 1269
Statement confidence coefficient, 154
Statistical computing packages, 980–981

Statistical design of scientific studies, 642
Statistical estimation, 1305–1306
Statistical relation, 2, 3–5
Steichen Bakeries, 1241–1244
Stem-and-leaf models, 101–102, 108, 110
Stem-and-leaf plots, 779
Stepwise Regression Methods, 364–368
Stepwise regression selection procedures, 364–368, 583–584
Structural empty cell, 967
Studentized deleted residuals, 396–398, 776–777
Studentized Pearson residuals, 592
Studentized range, 746–747
Studentized range distribution, tables of percentiles, 1333–1335
Studentized residuals, 394, 776
Studentized statistic, 44, 58
Subjects, 1127
  blocking, 1153
Subsampling:
  randomized block design, 1367, 1369–1370
  single-factor study, 1106–1113
  in three stages, 1113–1114
Sufficient estimator, 1305
Summation operator, 1297
Sums of squares, 25, 225
  for blocks, 898
  in matrix notation, 204–205
  nested design, 1094–1095
  partitioning, 63–66
  pooling, in two-factor analysis of variance, 861–862
  quadratic forms, 205–206
  rule for finding, 1359–1361
  for subjects, 1130
Supplemental variables, 344, 347, 919
Suppressor variable, 286
SYGRAPH, 101, 102, 104
Symmetric matrix, 185
Symmetry (of probit response function), 560
Synergistic interaction type, 308
SYSTAT, 19–20, 981

T

Taguchi, G., 1244
Taylor series expansion, 518
t distribution, 1304
  Bonferroni procedure, 159, 160
  table of percentiles, 1317–1318
Testing, 738
Tests:
  for constancy of error variance, 116–119, 780–785
  for constancy of variance, 115
  factor level means, 704
  family of, 154–155
  goodness of fit, 586–590
  lack of fit test, 119–127
  for normality, 115
  for outliers, 115
  for randomness, 114 (*See also F* test; *t* test)
Third-order regression model, 296
Three-dimensional plots, 1284–1286
Three-dimensional scatter plots, 233
Three-factor interactions, 996
  interpretation of, 998–999
  test for, 1011–1012
Three-factor study:
  ANOVA model, 992–998
  ANOVA partitioning, 1008–1009
  estimation of effects, 1013–1017, 1069–1070
  evaluation of appropriateness, 1005–1007
  expected mean squares, 1009
  fitting of model, 1003–1005
  F tests, 1009–1010, 1067–1068
  model II, 1066
  model III, 1066–1067
  nested design, 1114–1119
  regression approach, 1019–1020
  residual analysis, 1006, 1007
  unequal sample sizes, 1019–1021, 1070–1077
Tidwell, P. W, 236
Time series data, 319–321, 481 (*See also* Autocorrelation)
Total deviation, 65
Total mean squared error, 357–359
Total sum of squares, 63–66, 690–693
  partitioning, 836–838, 1008–1009
Total uncorrected sum of squares, 67
Total variance, 1033
Training sample, 372
Training the network, 542
Transformable interactions, 826–827
Transformations of variables, 85–87, 129–137, 220, 236, 490–492, 562, 789–793
Transpose of matrix, 178–179
Treatment, 13, 649–652
Treatment combination, 649
Treatment effects (analysis of covariance), 922–923, 928–929, 940, 1180–1182
Treatment means, 764, 817, 853, 1018–1019
  differences of, 1002
  of equal importance, 1091
  estimation, 884–886
  multiple comparisons, 856–861
  of unequal importance, 970–980
Treatment means plot, 820
Treatment mean square, 694
Treatment pattern sum of squares, 1199
Treatment sum of squares, 691, 837–838
Trial, 4
t test, 287–288
  equivalence of *F* test, 71
  power of, 50–51
  power value charts, 1327–1328
Tukey joint estimation procedure:
  balanced incomplete block design, 1182
  latin square design, 1190, 1191
  nested design, 1101, 1102
  randomized block design, 904
  repeated measures design, 1148, 1157
  three-factor analysis of variance, 1015, 1017
  two-factor analysis of variance, 850–851, 856
Tukey-Kramer procedure, 751
Tukey multiple comparison procedure, single-factor analysis of variance, 746–753
Tukey one degree of freedom test, 887
Tukey test for additivity:
  latin square design, 1191–1192
  randomized block design, 903–904
  two-factor analysis of variance, 886–888
Tuning constants, 440
Two-factor interactions, 823, 995, 1012
  interpretation of multiple, 999–1000, 1016
  interpretation of single, 1016–1017
Two-factor studies:
  analysis of covariance, 934–935
  ANOVA model for, 829–833
  ANOVA partitioning, 836–840
  crossed, 1366–1369
  empty cells, 964–967
  estimation of effects, 848–861, 959–964, 970–980, 1055–1060
  example, 812–813
  expected mean squares, 840, 1052–1053
  fitting of model, 834–836
  F tests, 843–847, 1053–1054
  general linear test approach, 953, 972–974
  mean squares, 839–840
  model II, 1047–1049
  model III, 1049–1052
  nested design, 1091–1092
  no-interaction model, 880–886
  partitioning, 836–840
  planning sample sizes for, 862–864
    estimation approach, 863–864
    finding the "best" treatment, 864
    power approach, 862–863
  pooling sums of squares, 861–862
  regression approach, 953–959

Two-factor studies (*cont.*)
repeated measures design, 1153–1161
residual analysis, 842–843
strategy for analysis, 847–858
Tukey test for additivity, 886–888
unequal sample sizes, 951–964
Two-level factorial design, 665–666, 1210–1212
center point replications, 1222–1223, 1243
estimation of effects, 1212–1214
F test, 1214–1215
incomplete block designs, 1240–1244
normal probability plot, 1222
Pareto plot, 1219–1220
pooling of interactions, 1218–1219
unreplicated, 1216–1223
Two-level fractional factorial design, 1223–1224
confounding, 1224–1227
defining relation, 1227–1228
half-fraction, 1229
projection property, 1232
quarter-fraction, 1229–1231
resolution, 1231–1232
setting a fraction of highest resolution, 1232–1239
smaller-fraction design, 1229–1231
Two-sided test, 47, 51
Two-variable conditioning plots, 451–452

U

Unbalanced nested design, 1104–1106
Unbiased condition, 43–44
Unbiased estimator, 1305
Uncontrolled variables, 919 (*See also* Supplemental variables)
Unequal error variances, transformations, 132–134
Unequal sample sizes in analysis of variance, 951–964
estimation of effects, 959–964, 970–980, 1020–1021
testing of effects, 953–959, 1019–1020
three-factor studies, 1070–1077
Unequal treatment importance, 970–980
Uniform precision central composite design, 1273, 1275
Unimportant interactions, 824–826
University admissions data set, 1351
Unmeasurable mean, 1226
Unrestricted mixed factor effects model, 1049, 1050
Unweighted mean, 702
factor effects model with, 705–708

V

Validation of regression model, 350, 369–375
Validation set, 372
Variables:
relations between, 2–5
transformations, 129–137
Variable metric method, 543
Variance, 52–54
of error terms, 9, 24–26, 27–28, 43–44
of prediction error, 57–59
of random variable, 1299–1300
of residuals, 102
tests for constancy of error variance, 115–119, 234, 780–785
Variance analysis (*see* Analysis of variance)
Variance components, 1055–1056
Variance-covariance matrix:
of random vector, 194–196
of regression coefficients, 207–208, 227–228
of residuals, 203–204
Variance function, 1271
Variance inflation factor, 406–410, 434–435
Variance operator, 1299
V criterion, 1280–1281
Vector, 178
with all elements 0, 187
with all elements unity, 187
random, 193–196

W

Wald test, 578
Watts, D. G., 529
Website developer data set, 1352
Weighted least squares method, 128, 421–431
ANOVA models, 786–789
Weighted mean, 703
factor effects model with, 709–710
Weight function, 439–441
Whole plots, 1162, 1163
Within-class matching, 669
Within-subjects sum of squares, 1131
Working-Hotelling joint estimation of mean responses:
confidence band, 61–62
multiple regression, 230
simple linear regression, 158–159

X

χ^2 distribution, 1303
table of percentiles, 1319
χ levels, 170–171
χ values, random, 78–89

Y

Y values, random, 78–89

Z

z' transformation, 85
Zero vector, 187