Chapter 8

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:

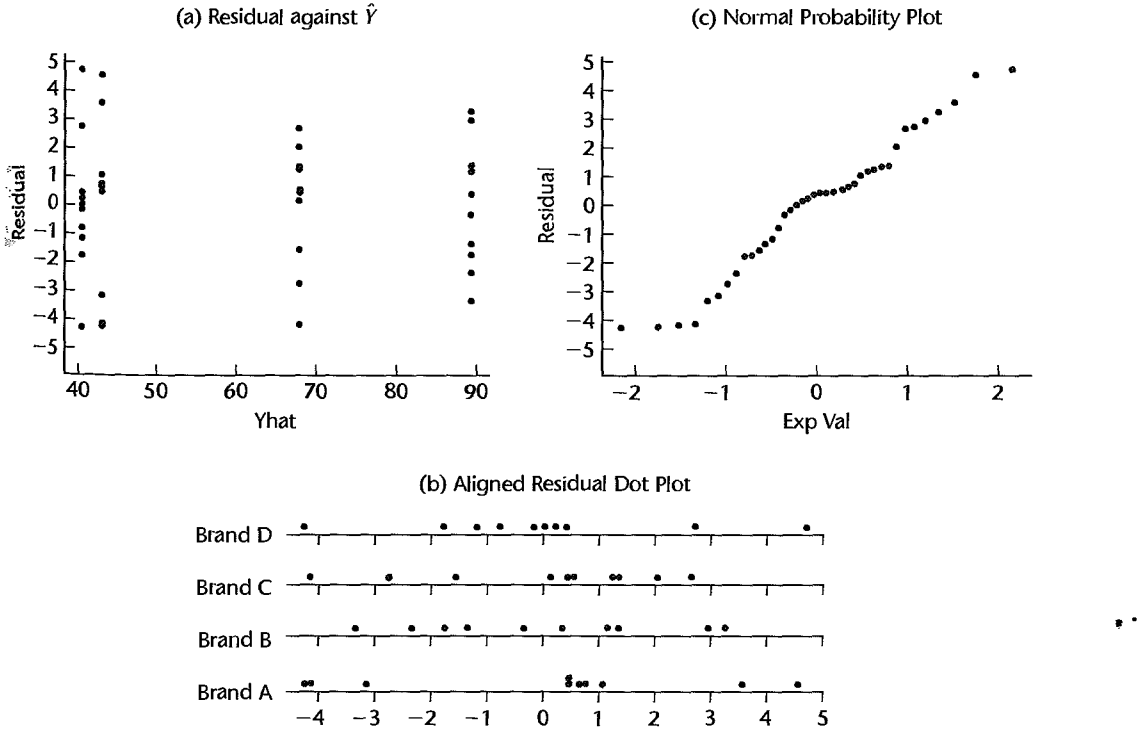
$$e_{11} = Y_{11} - \hat{Y}_{11} = Y_{11} - \bar{Y}_{1\cdot} = 43.9 - 43.14 = .76$$

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

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

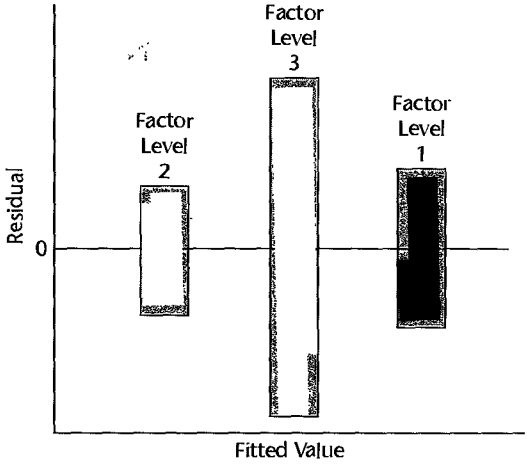
Nonconstancy of Error Variance. ANOVA model (16.2) requires that the error terms ε_{ij} have constant variance for all factor levels. When the sample sizes are not large and do not differ greatly, the appropriateness of this assumption can be studied by using the residuals, semistudentized residuals, 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.

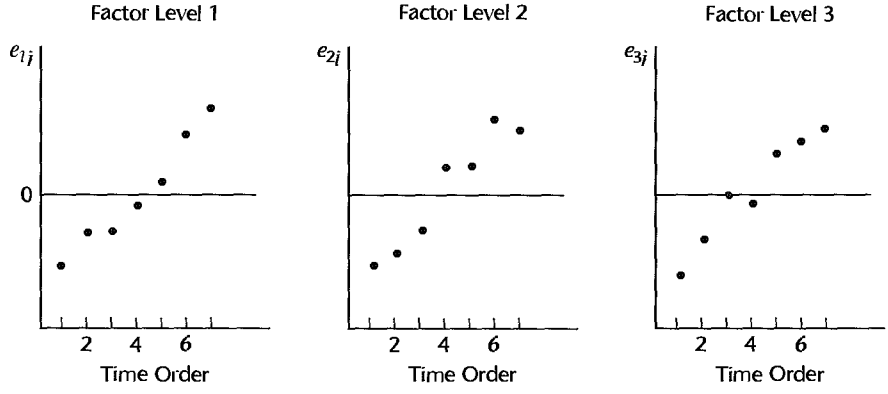
When the sample sizes for the different factor levels are large, *histograms* or boxplots of the residuals for each treatment—arranged vertically and using the same scale, like the dot plots in Figure 18.1b—are an effective means for examining the constancy of the error variance, as well as for assessing 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



GURE 18.2
totype Plot
esiduals
t Fitted
ues When
or Term
ance Is Not
onstant for
Factor
vels.



IGURE 18.3
esidual
equence Plots
or Group
nteraction
Study
llustrating
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-*

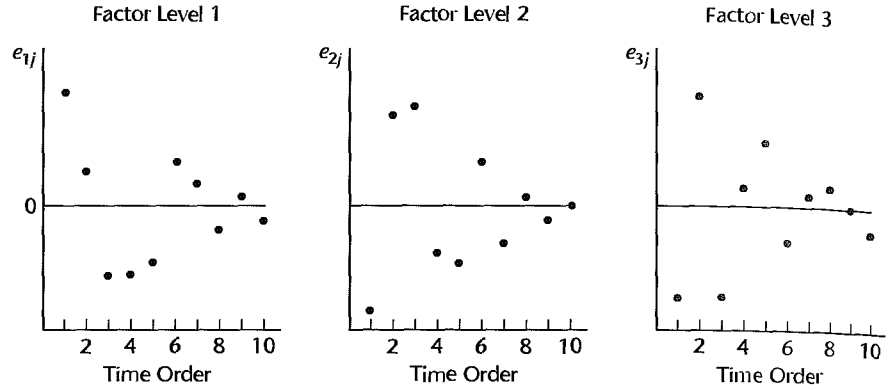


FIGURE 18.4 **Residual Sequence Plots Illustrating Decreasing Error Variance over Time.**

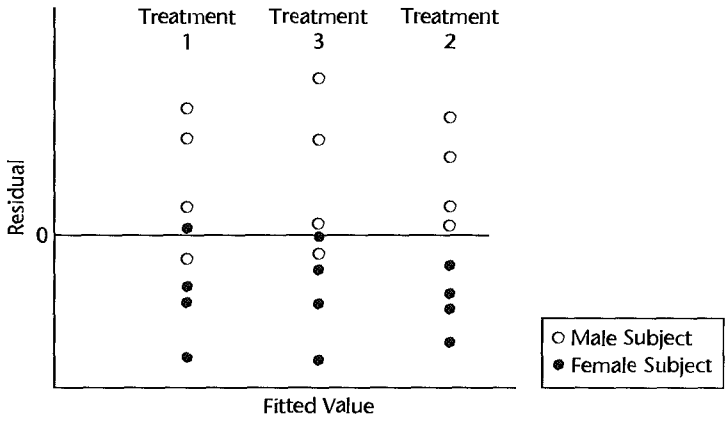


FIGURE 18.5 **Residual Plot against Fitted Values Illustrating Omission of Important Explanatory Variable.**

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:

$$\begin{aligned} &H_0\text{: } \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_r^2 \\ &H_a\text{: not all } \sigma_i^2 \text{ are equal} \end{aligned} \tag{18.7}$$

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:

$$\begin{aligned} &\text{If } H^* \leq H(1-\alpha; r, df), \text{ conclude } H_0 \\ &\text{If } H^* > H(1-\alpha; r, df), \text{ conclude } H_a \end{aligned} \tag{18.9}$$

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: not all σ_i^2 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

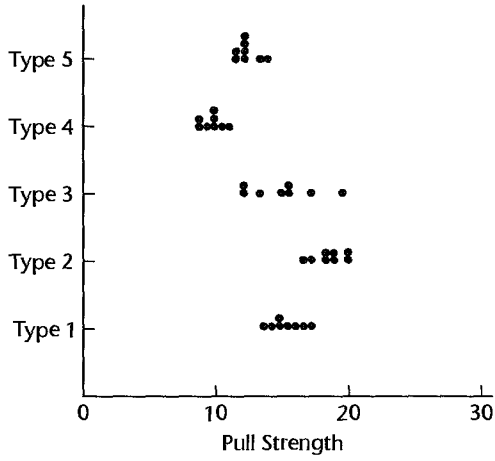


FIGURE 18.6 Dot Plots of Pull Strengths—ABT Electronics Example.

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:

$$\begin{aligned} &H_0\text{: } \mu_1 = \mu_2 = \cdots = \mu_5 \\ &H_a\text{: not all } \mu_i \text{ are equal} \end{aligned} \tag{18.16}$$

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:

$$\text{If } \sigma_i^2 \text{ proportional to } \mu_i: \quad Y' = \sqrt{Y} \quad \text{or} \quad Y' = \sqrt{Y} + \sqrt{Y+1} \tag{18.20}$$

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

E 18.5
e between
puter
ures at
tions (in
)—
ô-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	$\frac{s_i^2}{\bar{Y}_{i\cdot}}$	$\frac{s_i}{\bar{Y}_{i\cdot}}$	$\frac{s_i}{\bar{Y}_{i\cdot}^2}$
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		

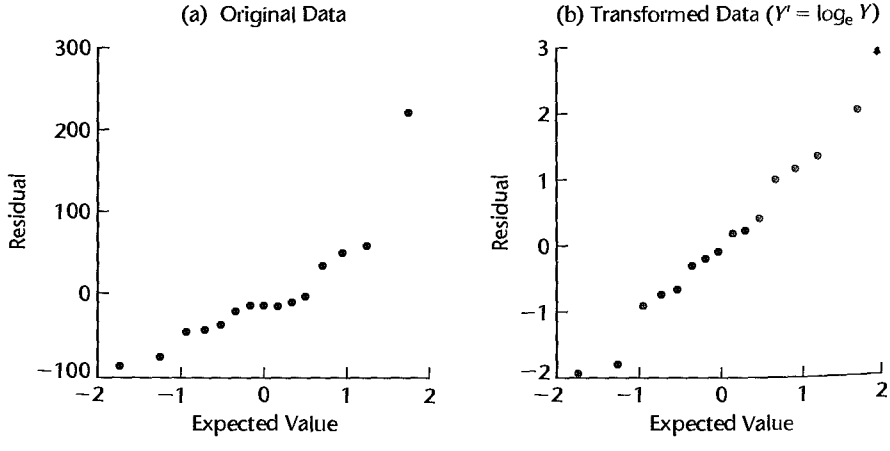


FIGURE 18.8 Normal Probability Plots for Original and Transformed Data—Servo-Data Example.

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.

18.7 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: not all μ_i are equal

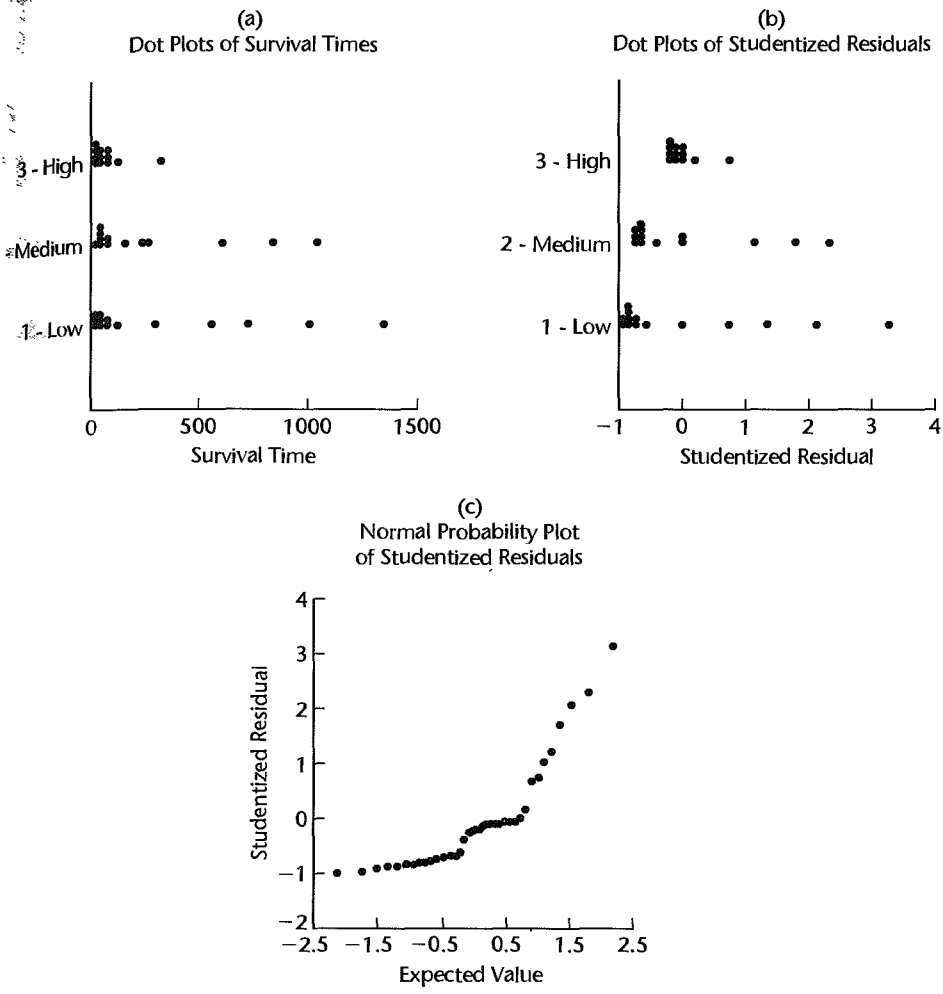
A SYSTAT dot plot of the data by mismatch category is provided in Figure 18.9a. The plot suggests that average survival time may decrease with higher degree of mismatch. An initial fit of analysis of variance model (16.2) was made and the studentized residuals were

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).

GURE 18.9 SYSTAT Diagnostic Plots—Heart Transplant Example.

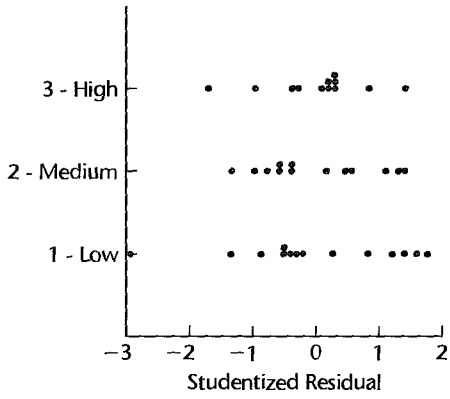


obtained for diagnostic purposes. Two residual plots are presented in Figures 18.9b and 18.9c. The dot plot of the studentized residuals in Figure 18.9b shows that the distribution of the residuals is positively skewed. It also suggests that the error variance may be smaller in the high mismatch group. 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).

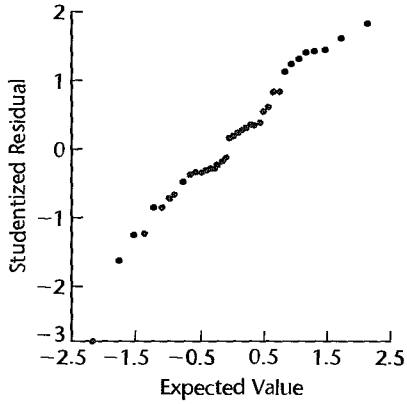
A transformation of the response variable was therefore investigated. The Box-Cox procedure led to the maximum likelihood estimate $\hat{\lambda} = .06$, which suggested the logarithmic transformation ($\lambda = 0$). The new response variable $Y' = \log_e Y$ was therefore obtained

FIGURE 18.10 Diagnostic Plots and ANOVA Table for Transformed Data—Heart Transplant Example.

(a)
Dot Plots of
Studentized Residuals



(b)
Normal Probability Plot
of Studentized Residuals



(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 $\bar{Y}_i.$ 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?

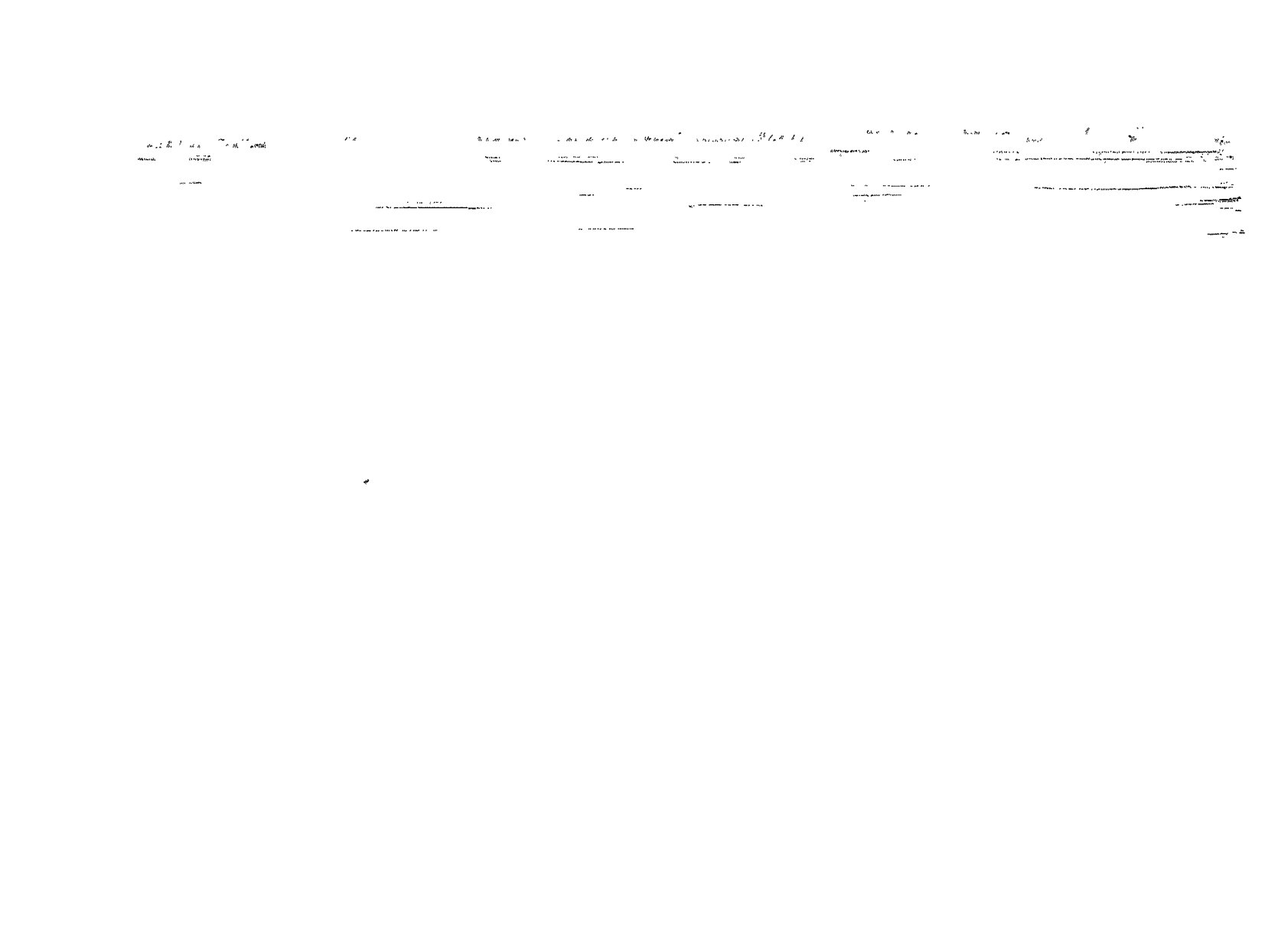
d. What proportion of the 500 test statistics obtained in part (b) is less than $F(.90; 2, 27)$? What proportion is less than $F(.99; 2, 27)$? How do these proportions agree with theoretical expectations?

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.

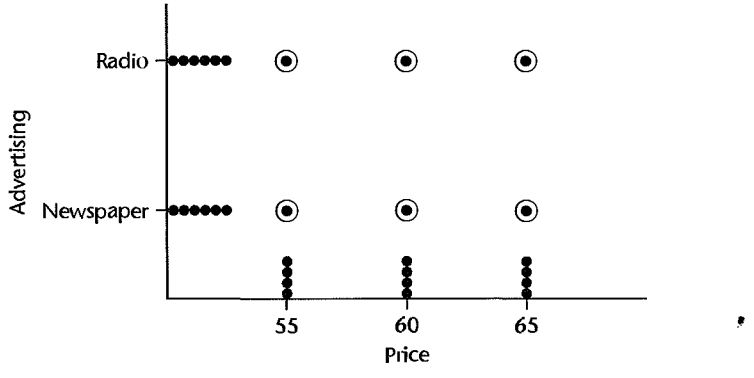


FIGURE 19.1 Experimental Layout—Example 1.

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.

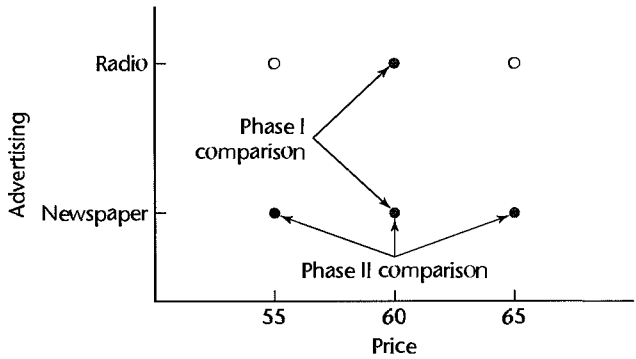


FIGURE 19.2 One-Factor-at-a-Time Approach—Example 1.

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. ■

19.2 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:

$$\mu_{ij} = \mu_{ij'} + \mu_{i'j} - \mu_{i'j'} \qquad \text{Additive factor effects} \qquad i \neq i',\ j \neq j' \tag{19.7b}$$

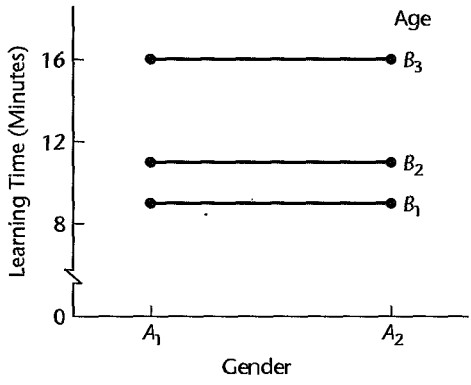


FIGURE 19.3
Age Effect but No Gender Effect, with No Interactions—Learning Example.

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—
ig
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 ($\mu_{\cdot\cdot}$)

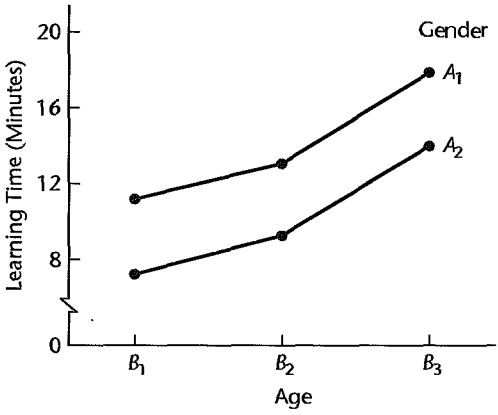
(b) Main Gender Effects (in minutes)

$$\alpha_1 = \mu_{1\cdot} - \mu_{\cdot\cdot} = 14 - 12 = 2$$
$$\alpha_2 = \mu_{2\cdot} - \mu_{\cdot\cdot} = 10 - 12 = -2$$

(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
ender Effects,
with No
Interactions—
earning
xample.



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:

$$(\alpha\beta)_{ij} = \mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{..} \tag{19.8a}$$

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.)

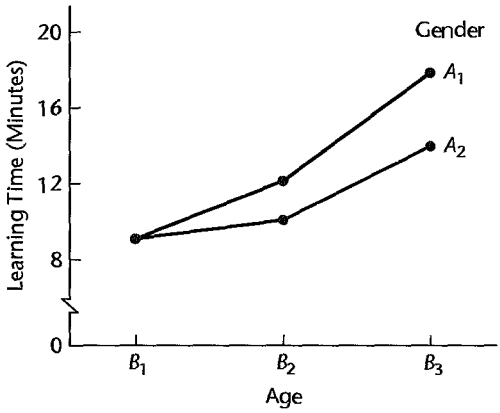


FIGURE 19.5
Age and Gender Effects, with Important Interactions—Learning Example.

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:

$$\sum_i (\alpha\beta)_{ij} = a\mu_{\cdot j} - a\mu_{\cdot\cdot} - a(\mu_{\cdot j} - \mu_{\cdot\cdot}) = 0$$

■

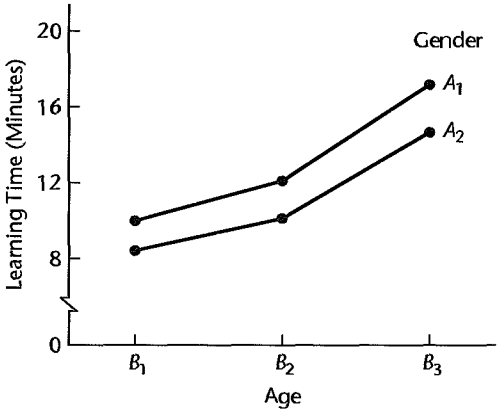
Important and Unimportant Interactions

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
ender 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 separately. This 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 level 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 ability 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):

$$\mu_{ij} = \mu_{..}\alpha_i\beta_j \qquad \text{Multiplicative factor effects} \qquad \textbf{(19.10)}$$

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):

$$\log \mu_{ij} = \log \mu_{..} + \log \alpha_i + \log \beta_j \qquad \textbf{(19.11)}$$

This result can be restated equivalently as follows:

$$\mu'_{ij} = \mu'_{..} + \alpha'_i + \beta'_j \qquad \textbf{(19.11a)}$$

where:

$$\mu'_{ij} = \log \mu_{ij}$$
$$\mu'_{..} = \log \mu_{..}$$
$$\alpha'_i = \log \alpha_i$$
$$\beta'_j = \log \beta_j$$

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:

$$\mu_{ij} = \alpha_i + \beta_j + 2\sqrt{\alpha_i}\sqrt{\beta_j} \qquad \text{Multiplicative interactions} \qquad \textbf{(19.12)}$$

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

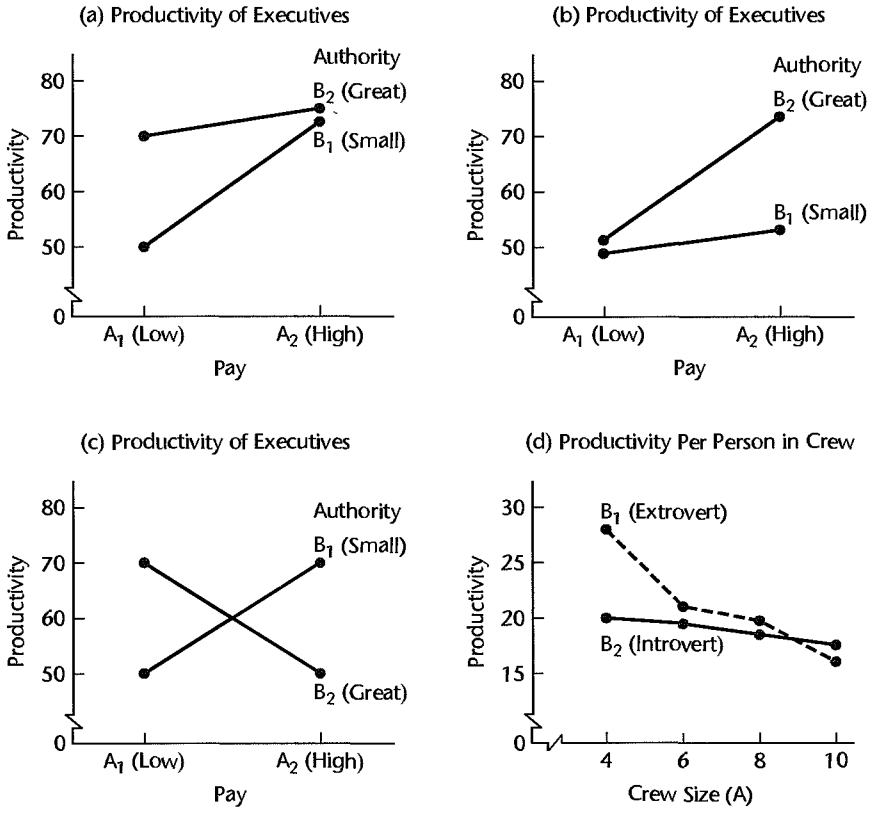
(d) Productivity per Person in Crew

	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:

$$\mathbf{Y} = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{111} \\ \varepsilon_{112} \\ \varepsilon_{121} \\ \varepsilon_{122} \\ \varepsilon_{211} \\ \varepsilon_{212} \\ \varepsilon_{221} \\ \varepsilon_{222} \end{bmatrix} \tag{19.19}$$

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):

$$(\alpha\beta)_{ij} = \mu_{ij} - (\mu_{\cdot\cdot} + \alpha_i + \beta_j)$$

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		Row Total	Display Height Average
	B_1 (regular)	B_2 (wide)		
A_1 (bottom)	47 (Y_{111}) 43 (Y_{112})	46 (Y_{121}) 40 (Y_{122})		
Total	90 ($Y_{11\cdot}$)	86 ($Y_{12\cdot}$)	176 ($Y_{1\cdot\cdot}$)	
Average	45 ($\bar{Y}_{11\cdot}$)	43 ($\bar{Y}_{12\cdot}$)		44 ($\bar{Y}_{1\cdot\cdot}$)
A_2 (middle)	62 (Y_{211}) 68 (Y_{212})	67 (Y_{221}) 71 (Y_{222})		
Total	130 ($Y_{21\cdot}$)	138 ($Y_{22\cdot}$)	268 ($Y_{2\cdot\cdot}$)	
Average	65 ($\bar{Y}_{21\cdot}$)	69 ($\bar{Y}_{22\cdot}$)		67 ($\bar{Y}_{2\cdot\cdot}$)
A_3 (top)	41 (Y_{311}) 39 (Y_{312})	42 (Y_{321}) 46 (Y_{322})		
Total	80 ($Y_{31\cdot}$)	88 ($Y_{32\cdot}$)	168 ($Y_{3\cdot\cdot}$)	
Average	40 ($\bar{Y}_{31\cdot}$)	44 ($\bar{Y}_{32\cdot}$)		42 ($\bar{Y}_{3\cdot\cdot}$)
Column total	300 ($Y_{\cdot 1\cdot}$)	312 ($Y_{\cdot 2\cdot}$)	612 ($Y_{\cdot\cdot\cdot}$)	
Display width average	50 ($\bar{Y}_{\cdot 1\cdot}$)	52 ($\bar{Y}_{\cdot 2\cdot}$)		51 ($\bar{Y}_{\cdot\cdot\cdot}$)

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:

$$e_{ijk} = Y_{ijk} - \bar{Y}_{ij\cdot} \tag{19.35}$$

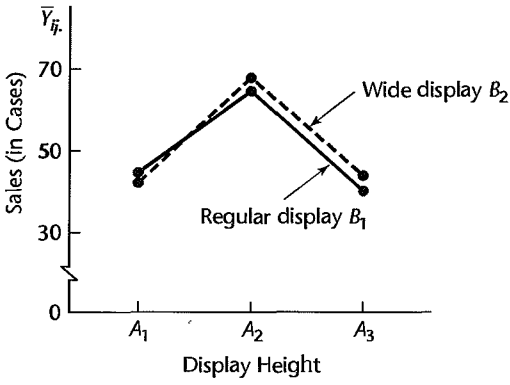


FIGURE 19.8 Estimated Treatment Means Plot—Castle Bakery Example.

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:

$$SSTO = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}...)^2 \tag{19.37a}$$

$$SSTR = n \sum_i \sum_j (\bar{Y}_{ij\cdot} - \bar{Y}...)^2 \tag{19.37b}$$

$$SSE = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij\cdot})^2 = \sum_i \sum_j \sum_k e_{ijk}^2 \tag{19.37c}$$

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 $\bar{Y}_{ij\cdot} - \bar{Y}...$ in terms of components reflecting the factor A main effect, the factor B main effect, and the AB interaction effect:

$$\underbrace{\bar{Y}_{ij\cdot} - \bar{Y}...}_{\substack{\text{Deviation of estimated treatment mean around overall mean}}} = \underbrace{\bar{Y}_{i\cdot\cdot} - \bar{Y}...}_{\substack{A \text{ main effect}}} + \underbrace{\bar{Y}_{\cdot j\cdot} - \bar{Y}...}_{\substack{B \text{ main effect}}} + \underbrace{\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}...}_{\substack{AB \text{ interaction effect}}} \tag{19.38}$$

When we square (19.38) and sum over all treatments and over the n cases associated with each estimated treatment mean $\bar{Y}_{ij\cdot}$, all cross-product terms drop out and we obtain:

$$SSTR = SSA + SSB + SSAB \tag{19.39}$$

where:

$$SSA = nb \sum_i (\bar{Y}_{i\cdot\cdot} - \bar{Y}...)^2 \tag{19.39a}$$

$$SSB = na \sum_j (\bar{Y}_{\cdot j\cdot} - \bar{Y}...)^2 \tag{19.39b}$$

$$SSAB = n \sum_i \sum_j (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}...)^2 \tag{19.39c}$$

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 $\bar{Y}_{i\cdot\cdot}$. 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 $\bar{Y}_{\cdot j\cdot}$. Finally, *SSAB*, called the *AB interaction sum of squares*, measures the variability of the estimated interactions $\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot}$ 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 $\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}$, but one degree of freedom is lost because the deviations are subject to one restriction, i.e., $\sum(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot}) = 0$. 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:

$$\sum_i (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot}) = 0 \qquad j = 1, \ldots, b$$

There are a additional restrictions since:

$$\sum_j (\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot}) = 0 \qquad i = 1, \ldots, a$$

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	$SSA = nb\sum(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$a-1$	$MSA = \dfrac{SSA}{a-1}$	$\sigma^2 + bn\dfrac{\sum(\mu_{i\cdot} - \mu_{\cdot\cdot})^2}{a-1}$
Factor B	$SSB = na\sum(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$	$b-1$	$MSB = \dfrac{SSB}{b-1}$	$\sigma^2 + an\dfrac{\sum(\mu_{\cdot j} - \mu_{\cdot\cdot})^2}{b-1}$
AB interactions	$SSAB = n\sum\sum(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$	$(a-1)(b-1)$	$MSAB = \dfrac{SSAB}{(a-1)(b-1)}$	$\sigma^2 + n\dfrac{\sum\sum(\mu_{ij} - \mu_{i\cdot} - \mu_{\cdot j} + \mu_{\cdot\cdot})^2}{(a-1)(b-1)}$
Error	$SSE = \sum\sum\sum(Y_{ijk} - \bar{Y}_{ij\cdot})^2$	$ab(n-1)$	$MSE = \dfrac{SSE}{ab(n-1)}$	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

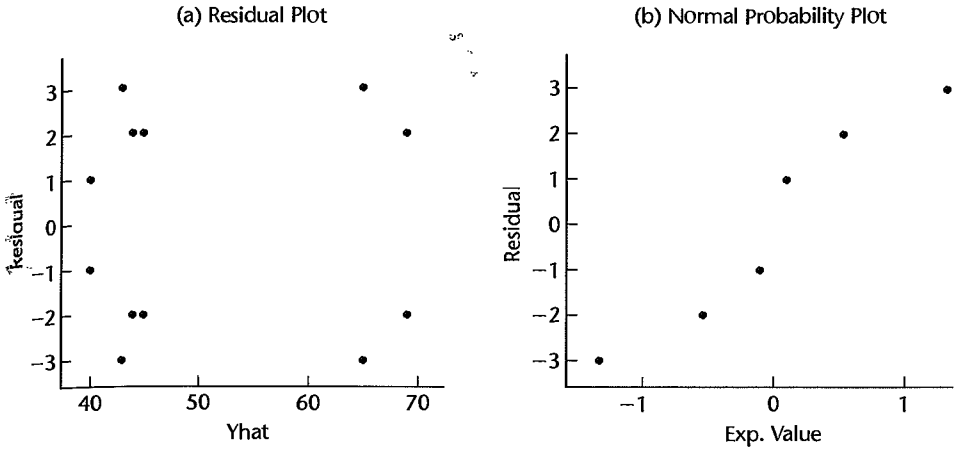
Before undertaking formal inference procedures, we need to evaluate the appropriateness of two-factor ANOVA model (19.23). No new problems arise here. The residuals in (19.35):

$$e_{ijk} = Y_{ijk} - \bar{Y}_{ij.}$$

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 $\mu_{\cdot j}$ 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):

$$Y_{ijk} = \mu_{\cdot\cdot} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk} \qquad \text{Full model} \tag{19.54}$$

Fitting this full model leads to the fitted values $\hat{Y}_{ijk} = \bar{Y}_{ij\cdot}$ and the error sum of squares:

$$SSE(F) = \sum\sum\sum(Y_{ijk} - \hat{Y}_{ijk})^2 = \sum\sum\sum(Y_{ijk} - \bar{Y}_{ij\cdot})^2 = SSE \tag{19.55}$$

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.

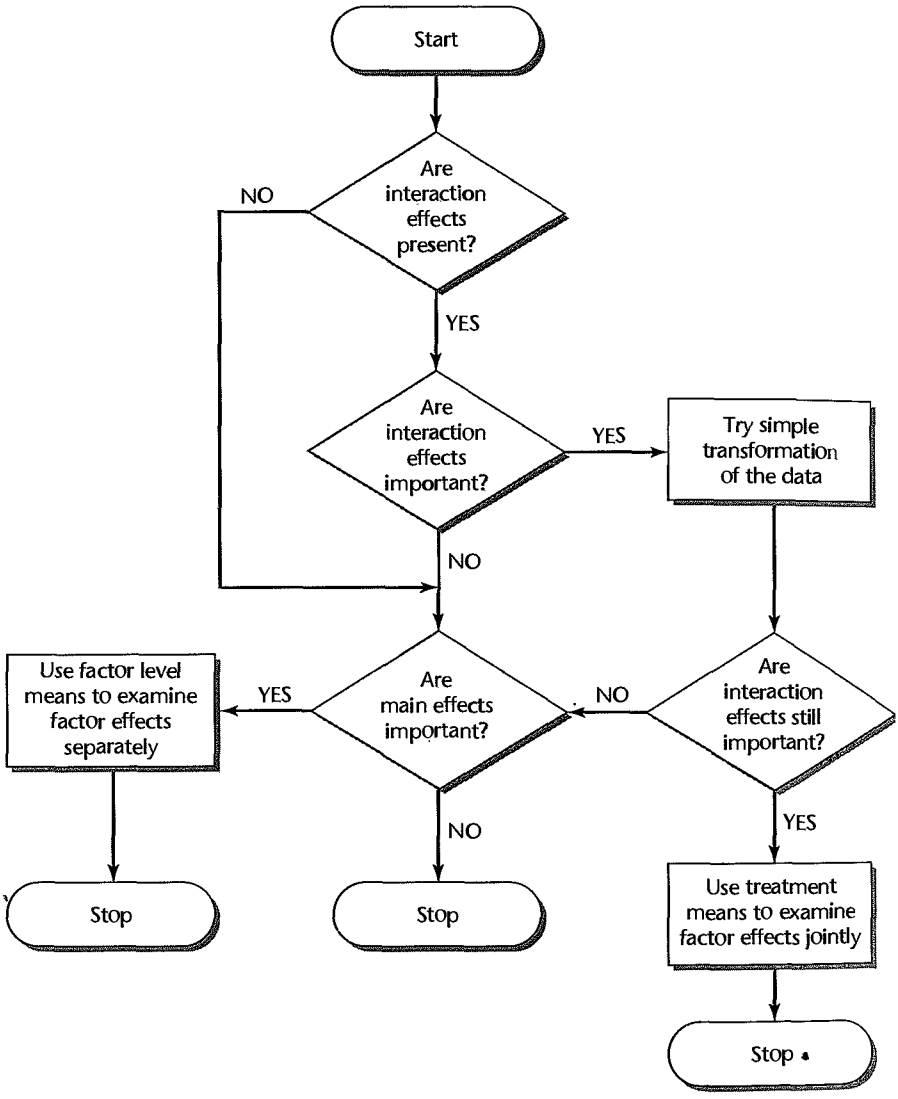


FIGURE 19.11
Strategy for Analysis of Two-Factor Studies.

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:

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

$$\hat{\mu}_{\cdot j} = \bar{Y}_{\cdot j}$$

where $\bar{Y}_{i\cdot\cdot}$ and $\bar{Y}_{\cdot j\cdot}$ are defined in (19.27d) and (19.27f), respectively. The variance of $\bar{Y}_{i\cdot\cdot}$ is:

$$\sigma^2\{\bar{Y}_{i\cdot\cdot}\} = \frac{\sigma^2}{bn} \tag{19.58a}$$

since $\bar{Y}_{i\cdot\cdot}$ contains bn independent observations, each with variance σ^2. Similarly, we have:

$$\sigma^2\{\bar{Y}_{\cdot j\cdot}\} = \frac{\sigma^2}{an} \tag{19.58b}$$

Unbiased estimators of these variances are obtained by replacing σ^2 with MSE:

$$s^2\{\bar{Y}_{i\cdot\cdot}\} = \frac{MSE}{bn} \tag{19.59a}$$

$$s^2\{\bar{Y}_{\cdot j\cdot}\} = \frac{MSE}{an} \tag{19.59b}$$

Confidence limits for $\mu_{i\cdot}$ and $\mu_{\cdot j}$ utilize, as usual, the t distribution:

$$\bar{Y}_{i\cdot\cdot} \pm t[1-\alpha/2;\,(n-1)ab]s\{\bar{Y}_{i\cdot\cdot}\} \tag{19.60a}$$

$$\bar{Y}_{\cdot j\cdot} \pm t[1-\alpha/2;\,(n-1)ab]s\{\bar{Y}_{\cdot j\cdot}\} \tag{19.60b}$$

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 $\mu_{i\cdot}$:

$$L = \sum c_i \mu_{i\cdot} \qquad \text{where } \sum c_i = 0 \tag{19.61}$$

is estimated unbiasedly by:

$$\hat{L} = \sum c_i \bar{Y}_{i\cdot\cdot} \tag{19.62}$$

Because of the independence of the $\bar{Y}_{i\cdot\cdot}$, the variance of this estimator is:

$$\sigma^2\{\hat{L}\} = \sum c_i^2 \sigma^2\{\bar{Y}_{i\cdot\cdot}\} = \frac{\sigma^2}{bn}\sum c_i^2 \tag{19.63}$$

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 $\mu_{\cdot j}$:

$$L = \sum c_j \mu_{\cdot j} \qquad \text{where } \sum c_j = 0 \tag{19.66}$$

we use the estimator:

$$\hat{L} = \sum c_j \bar{Y}_{\cdot j\cdot} \tag{19.67}$$

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:

$$D = \mu_{\cdot j} - \mu_{\cdot j'} \tag{19.74}$$

$$\hat{D} = \bar{Y}_{\cdot j} - \bar{Y}_{\cdot j'} \tag{19.75}$$

$$s^2\{\hat{D}\} = \frac{2MSE}{an} \tag{19.76}$$

$$T = \frac{1}{\sqrt{2}} q[1-\alpha; b, (n-1)ab] \tag{19.77}$$

$$q^* = \frac{\sqrt{2}\hat{D}}{s\{\hat{D}\}}; \qquad \text{If } |q^*| > q[1-\alpha; b, (n-1)ab], \text{ conclude } H_a \tag{19.78}$$

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

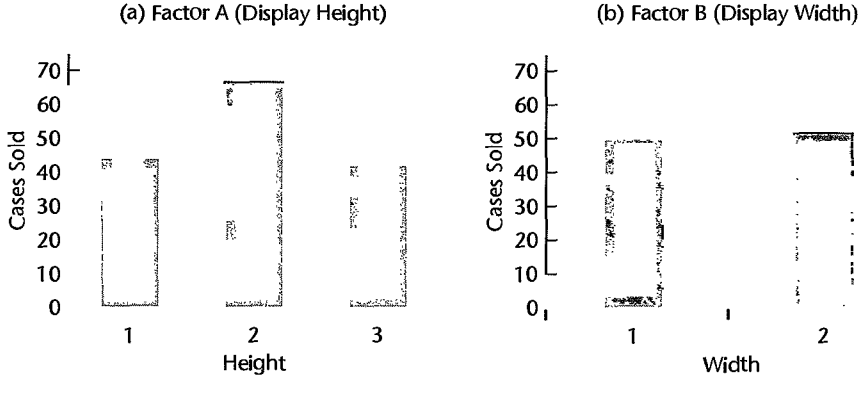


FIGURE 19.12 Bar Graphs of Estimated Factor Level Means—Castle Bakery Example.

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):

$$q(.95; 3, 6) = 4.34$$

$$T = \frac{4.34}{\sqrt{2}} = 3.07$$

$$Ts\{\hat{D}_1\} = Ts\{\hat{D}_3\} = 3.07(2.27) = 7.0$$

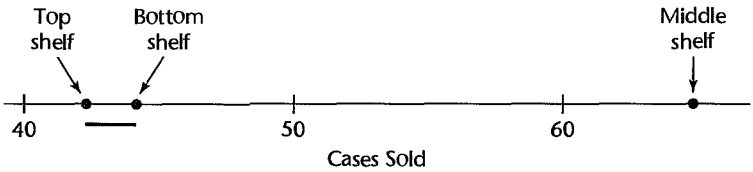
We therefore find the following confidence intervals for the two pairwise comparisons of the shelf height factor level means:

$$16 = 23 - 7.0 \leq \mu_{2\cdot} - \mu_{1\cdot} \leq 23 + 7.0 = 30$$

$$18 = 25 - 7.0 \leq \mu_{2\cdot} - \mu_{3\cdot} \leq 25 + 7.0 = 32$$

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:

$$t^* = \frac{\hat{D}}{s\{\hat{D}\}}; \qquad \text{If } |t^*| > t[1-\alpha/2g;\,(n-1)ab], \text{ conclude } H_a \qquad \textbf{(19.91)}$$

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:

$$L = \sum\sum c_{ij}\mu_{ij} \qquad \text{where } \sum\sum c_{ij} = 0 \qquad \textbf{(19.92)}$$

are:

$$\hat{L} \pm Ss\{\hat{L}\} \qquad \textbf{(19.93)}$$

where:

$$\hat{L} = \sum\sum c_{ij}\bar{Y}_{ij\cdot} \qquad \textbf{(19.93a)}$$

$$s^2\{\hat{L}\} = \frac{MSE}{n}\sum\sum c_{ij}^2 \qquad \textbf{(19.93b)}$$

$$S^2 = (ab-1)F[1-\alpha;\,ab-1,\,(n-1)ab] \qquad \textbf{(19.93c)}$$

The test statistic and associated decision rule for all simultaneous Scheffé tests of the form:

$$\begin{aligned} H_0&: L = 0 \\ H_a&: L \neq 0 \end{aligned} \qquad \textbf{(19.94)}$$

are as follows when the family significance level is controlled at α:

$$F^* = \frac{\hat{L}^2}{(ab-1)s^2\{\hat{L}\}}; \qquad \text{If } F^* > F[1-\alpha;\,ab-1,\,(n-1)ab], \text{ conclude } H_a \qquad \textbf{(19.94a)}$$

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:

$$t^* = \frac{\hat{L}}{s\{\hat{L}\}}; \qquad \text{If } |t^*| > t[1-\alpha/2g;\,(n-1)ab], \text{ conclude } H_a \qquad \textbf{(19.95)}$$

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+.

GURE 19.13
ots of
timated
tment
eans—
athematics
arning
xample.

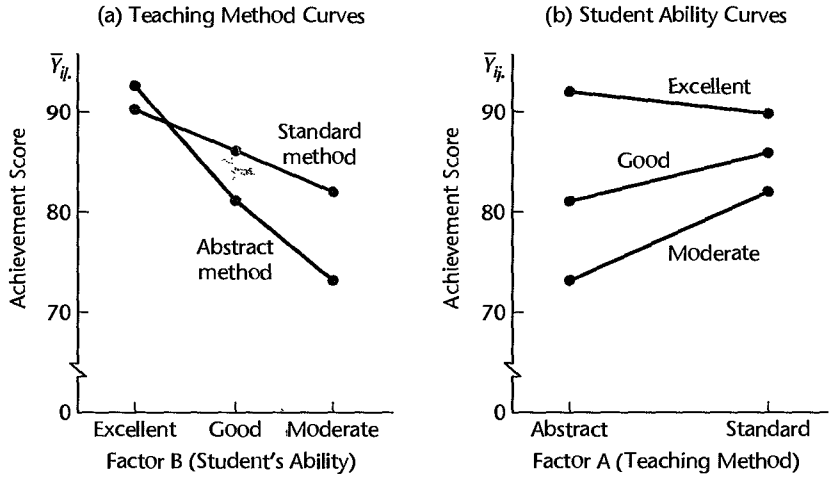


Figure 19.13 suggests that the interactions are important: students with excellent quantitative ability are but little affected by teaching method (perhaps doing slightly better with the abstract method); students with good or moderate abilities learn much better with the standard teaching method. Hence, 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:

$$Y_{ijk} = \mu.. + \alpha_i + \beta_j + \varepsilon_{ijk} \qquad \text{Revised full model} \qquad \textbf{(19.96)}$$

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:

$$SSE(F) = SSE + SSAB \qquad \textbf{(19.97)}$$

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:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{n\sum\sum(\alpha\beta)_{ij}^2}{(a-1)(b-1)+1}} = \frac{1}{\sigma}\sqrt{\frac{n\sum\sum(\mu_{ij}-\mu_{i\cdot}-\mu_{\cdot j}+\mu_{\cdot\cdot})^2}{(a-1)(b-1)+1}} \qquad (19.98a)$$

$$\nu_1 = (a-1)(b-1) \qquad \nu_2 = ab(n-1)$$

Test for *A* main effects:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{nb\sum\alpha_i^2}{a}} = \frac{1}{\sigma}\sqrt{\frac{nb\sum(\mu_{i\cdot}-\mu_{\cdot\cdot})^2}{a}} \qquad (19.98b)$$

$$\nu_1 = a-1 \qquad \nu_2 = ab(n-1)$$

Test for *B* main effects:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{na\sum\beta_j^2}{b}} = \frac{1}{\sigma}\sqrt{\frac{na\sum(\mu_{\cdot j}-\mu_{\cdot\cdot})^2}{b}} \qquad (19.98c)$$

$$\nu_1 = b-1 \qquad \nu_2 = ab(n-1)$$

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 $\bar{Y}_{ij}$. 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 B (gender of owner)	
Factor A (age)		$j = 1$ 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 B (ingredient 2)		
Factor A (ingredient 1)		$j = 1$ 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 B (weight gain)					
Factor A (duration)		$j = 1$ 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 *B* (years of experience)		
Factor *A* (type 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 $D = \mu_{\cdot 1} - \mu_{\cdot 2}$ 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 $\bar{Y}_{i\cdot\cdot\cdot}$. 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 $\bar{Y}_{i\cdot\cdot\cdot}$. 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.