STATISTICS *in* PRACTICE

BURKE MARKETING SERVICES, INC.*

CINCINNATI, OHIO

Burke Marketing Services, Inc., is one of the most experienced market research firms in the industry. Burke writes more proposals, on more projects, every day than any other market research company in the world. Supported by state-of-the-art technology, Burke offers a wide variety of research capabilities, providing answers to nearly any marketing question.

In one study, a firm retained Burke to evaluate potential new versions of a children's dry cereal. To maintain confidentiality, we refer to the cereal manufacturer as the Anon Company. The four key factors that Anon's product developers thought would enhance the taste of the cereal were the following:

1. Ratio of wheat to corn in the cereal flake
2. Type of sweetener: sugar, honey, or artificial
3. Presence or absence of flavor bits with a fruit taste
4. Short or long cooking time

Burke designed an experiment to determine what effects these four factors had on cereal taste. For example, one test cereal was made with a specified ratio of wheat to corn, sugar as the sweetener, flavor bits, and a short cooking time; another test cereal was made with a different ratio of wheat to corn and the other three factors the same, and so on. Groups of children then taste-tested the cereals and stated what they thought about the taste of each.

Burke uses taste tests to provide valuable statistical information on what customers want from a product. © JLP/Sylvia Torres/CORBIS.

Analysis of variance was the statistical method used to study the data obtained from the taste tests. The results of the analysis showed the following:

- The flake composition and sweetener type were highly influential in taste evaluation.
- The flavor bits actually detracted from the taste of the cereal.
- The cooking time had no effect on the taste.

This information helped Anon identify the factors that would lead to the best-tasting cereal.

The experimental design employed by Burke and the subsequent analysis of variance were helpful in making a product design recommendation. In this chapter, we will see how such procedures are carried out.

*The authors are indebted to Dr. Ronald Tatham of Burke Marketing Services for providing this Statistics in Practice.

In Chapter 1 we stated that statistical studies can be classified as either experimental or observational. In an experimental statistical study, an experiment is conducted to generate the data. An experiment begins with identifying a variable of interest. Then one or more other variables, thought to be related, are identified and controlled, and data are collected about how those variables influence the variable of interest.

In an observational study, data are usually obtained through sample surveys and not a controlled experiment. Good design principles are still employed, but the rigorous controls associated with an experimental statistical study are often not possible. For instance, in a study of the relationship between smoking and lung cancer the researcher cannot assign a smoking habit to subjects. The researcher is restricted to simply observing the effects of smoking on people who already smoke and the effects of not smoking on people who do not already smoke.

Sir Ronald Alymer Fisher (1890–1962) invented the branch of statistics known as experimental design. In addition to being accomplished in statistics, he was a noted scientist in the field of genetics.

In this chapter we introduce three types of experimental designs: a completely randomized design, a randomized block design, and a factorial experiment. For each design we show how a statistical procedure called analysis of variance (ANOVA) can be used to analyze the data available. ANOVA can also be used to analyze the data obtained through an observation a study. For instance, we will see that the ANOVA procedure used for a completely randomized experimental design also works for testing the equality of three or more population means when data are obtained through an observational study. In the following chapters we will see that ANOVA plays a key role in analyzing the results of regression studies involving both experimental and observational data.

In the first section, we introduce the basic principles of an experimental study and show how they are employed in a completely randomized design. In the second section, we then show how ANOVA can be used to analyze the data from a completely randomized experimental design. In later sections we discuss multiple comparison procedures and two other widely used experimental designs, the randomized block design and the factorial experiment.

13.1 An Introduction to Experimental Design and Analysis of Variance

Cause-and-effect relationships can be difficult to establish in observational studies; such relationships are easier to establish in experimental studies.

As an example of an experimental statistical study, let us consider the problem facing Chemitech, Inc. Chemitech developed a new filtration system for municipal water supplies. The components for the new filtration system will be purchased from several suppliers, and Chemitech will assemble the components at its plant in Columbia, South Carolina. The industrial engineering group is responsible for determining the best assembly method for the new filtration system. After considering a variety of possible approaches, the group narrows the alternatives to three: method A, method B, and method C. These methods differ in the sequence of steps used to assemble the system. Managers at Chemitech want to determine which assembly method can produce the greatest number of filtration systems per week.

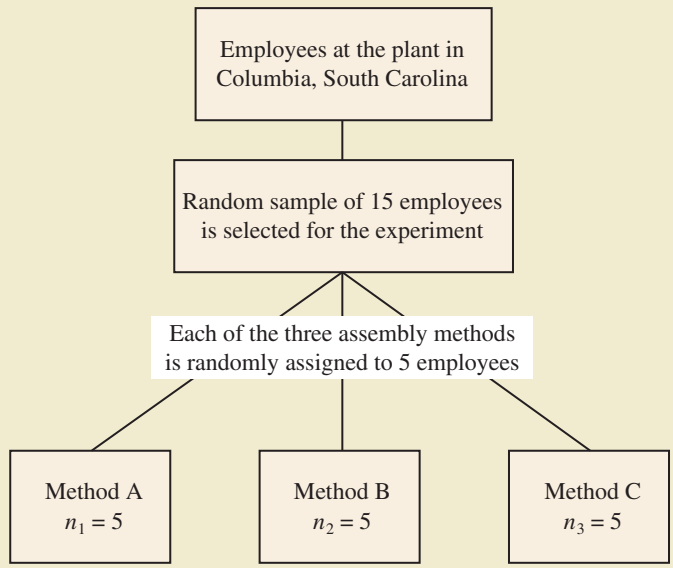
In the Chemitech experiment, assembly method is the independent variable or **factor**. Because three assembly methods correspond to this factor, we say that three treatments are associated with this experiment; each **treatment** corresponds to one of the three assembly methods. The Chemitech problem is an example of a **single-factor experiment**; it involves one qualitative factor (method of assembly). More complex experiments may consist of multiple factors; some factors may be qualitative and others may be quantitative.

The three assembly methods or treatments define the three populations of interest for the Chemitech experiment. One population is all Chemitech employees who use assembly method A, another is those who use method B, and the third is those who use method C. Note that for each population the dependent or **response variable** is the number of filtration systems assembled per week, and the primary statistical objective of the experiment is to determine whether the mean number of units produced per week is the same for all three populations (methods).

Randomization is the process of assigning the treatments to the experimental units at random. Prior to the work of Sir R. A. Fisher, treatments were assigned on a systematic or subjective basis.

Suppose a random sample of three employees is selected from all assembly workers at the Chemitech production facility. In experimental design terminology, the three randomly selected workers are the **experimental units**. The experimental design that we will use for the Chemitech problem is called a **completely randomized design**. This type of design requires that each of the three assembly methods or treatments be assigned randomly to one of the experimental units or workers. For example, method A might be randomly assigned to the second worker, method B to the first worker, and method C to the third worker. The concept of *randomization,* as illustrated in this example, is an important principle of all experimental designs.

FIGURE 13.1 COMPLETELY RANDOMIZED DESIGN FOR EVALUATING THE CHEMITECH ASSEMBLY METHOD EXPERIMENT



Note that this experiment would result in only one measurement or number of units assembled for each treatment. To obtain additional data for each assembly method, we must repeat or replicate the basic experimental process. Suppose, for example, that instead of selecting just three workers at random we selected 15 workers and then randomly assigned each of the three treatments to 5 of the workers. Because each method of assembly is assigned to 5 workers, we say that five replicates have been obtained. The process of *replication* is another important principle of experimental design. Figure 13.1 shows the completely randomized design for the Chemitech experiment.

Data Collection

Once we are satisfied with the experimental design, we proceed by collecting and analyzing the data. In the Chemitech case, the employees would be instructed in how to perform the assembly method assigned to them and then would begin assembling the new filtration systems using that method. After this assignment and training, the number of units assembled by each employee during one week is as shown in Table 13.1. The sample means, sample variances, and sample standard deviations for each assembly method are also provided. Thus, the sample mean number of units produced using method A is 62; the sample mean using method B is 66; and the sample mean using method C is 52. From these data, method B appears to result in higher production rates than either of the other methods.

The real issue is whether the three sample means observed are different enough for us to conclude that the means of the populations corresponding to the three methods of assembly are different. To write this question in statistical terms, we introduce the following notation.

μ_1 = mean number of units produced per week using method A
μ_2 = mean number of units produced per week using method B
μ_3 = mean number of units produced per week using method C

TABLE 13.1 NUMBER OF UNITS PRODUCED BY 15 WORKERS

WEB file
Chemitech

	Method		
	A	B	C
	58	58	48
	64	69	57
	55	71	59
	66	64	47
	67	68	49
Sample mean	62	66	52
Sample variance	27.5	26.5	31.0
Sample standard deviation	5.244	5.148	5.568

Although we will never know the actual values of μ_1, μ_2, and μ_3, we want to use the sample means to test the following hypotheses.

If H_0 is rejected, we cannot conclude that all population means are different. Rejecting H_0 means that at least two population means have different values.

$$H_0: \mu_1 = \mu_2 = \mu_3$$
$$H_a: \text{Not all population means are equal}$$

As we will demonstrate shortly, analysis of variance (ANOVA) is the statistical procedure used to determine whether the observed differences in the three sample means are large enough to reject H_0.

Assumptions for Analysis of Variance

Three assumptions are required to use analysis of variance.

If the sample sizes are equal, analysis of variance is not sensitive to departures from the assumption of normally distributed populations.

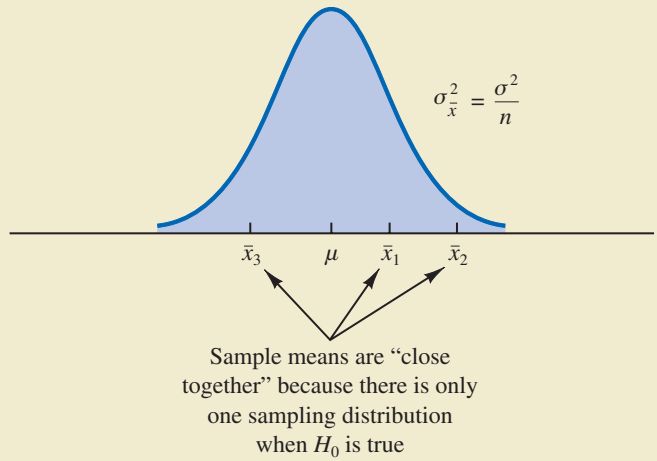
1. **For each population, the response variable is normally distributed.** Implication: In the Chemitech experiment the number of units produced per week (response variable) must be normally distributed for each assembly method.
2. **The variance of the response variable, denoted σ^2, is the same for all of the populations.** Implication: In the Chemitech experiment, the variance of the number of units produced per week must be the same for each assembly method.
3. **The observations must be independent.** Implication: In the Chemitech experiment, the number of units produced per week for each employee must be independent of the number of units produced per week for any other employee.

Analysis of Variance: A Conceptual Overview

If the means for the three populations are equal, we would expect the three sample means to be close together. In fact, the closer the three sample means are to one another, the more evidence we have for the conclusion that the population means are equal. Alternatively, the more the sample means differ, the more evidence we have for the conclusion that the population means are not equal. In other words, if the variability among the sample means is "small," it supports H_0; if the variability among the sample means is "large," it supports H_a.

If the null hypothesis, $H_0: \mu_1 = \mu_2 = \mu_3$, is true, we can use the variability among the sample means to develop an estimate of σ^2. First, note that if the assumptions for analysis

FIGURE 13.2 SAMPLING DISTRIBUTION OF $\bar{x}$ GIVEN H_0 IS TRUE



of variance are satisfied, each sample will have come from the same normal distribution with mean μ and variance σ^2. Recall from Chapter 7 that the sampling distribution of the sample mean $\bar{x}$ for a simple random sample of size n from a normal population will be normally distributed with mean μ and variance σ^2/n. Figure 13.2 illustrates such a sampling distribution.

Thus, if the null hypothesis is true, we can think of each of the three sample means, $\bar{x}_1 = 62$, $\bar{x}_2 = 66$, and $\bar{x}_3 = 52$ from Table 13.1, as values drawn at random from the sampling distribution shown in Figure 13.2. In this case, the mean and variance of the three $\bar{x}$ values can be used to estimate the mean and variance of the sampling distribution. When the sample sizes are equal, as in the Chemitech experiment, the best estimate of the mean of the sampling distribution of $\bar{x}$ is the mean or average of the sample means. Thus, in the Chemitech experiment, an estimate of the mean of the sampling distribution of $\bar{x}$ is $(62 + 66 + 52)/3 = 60$. We refer to this estimate as the *overall sample mean.* An estimate of the variance of the sampling distribution of $\bar{x}$, $\sigma_{\bar{x}}^2$, is provided by the variance of the three sample means.

$$s_{\bar{x}}^2 = \frac{(62 - 60)^2 + (66 - 60)^2 + (52 - 60)^2}{3 - 1} = \frac{104}{2} = 52$$

Because $\sigma_{\bar{x}}^2 = \sigma^2/n$, solving for σ^2 gives

$$\sigma^2 = n\sigma_{\bar{x}}^2$$

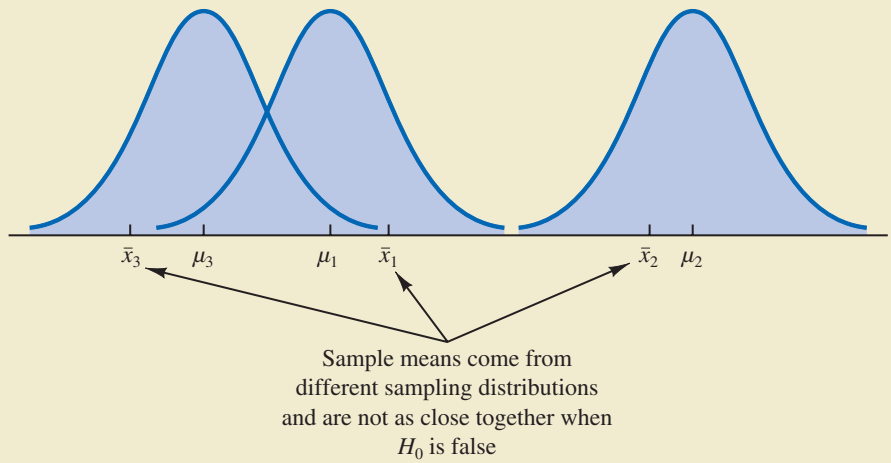
Hence,

$$\text{Estimate of } \sigma^2 = n \text{ (Estimate of } \sigma_{\bar{x}}^2) = ns_{\bar{x}}^2 = 5(52) = 260$$

The result, $ns_{\bar{x}}^2 = 260$, is referred to as the *between-treatments* estimate of σ^2.

The between-treatments estimate of σ^2 is based on the assumption that the null hypothesis is true. In this case, each sample comes from the same population, and there is only

FIGURE 13.3 SAMPLING DISTRIBUTIONS OF $\bar{x}$ GIVEN H_0 IS FALSE



one sampling distribution of $\bar{x}$. To illustrate what happens when H_0 is false, suppose the population means all differ. Note that because the three samples are from normal populations with different means, they will result in three different sampling distributions. Figure 13.3 shows that in this case, the sample means are not as close together as they were when H_0 was true. Thus, $s_{\bar{x}}^2$ will be larger, causing the between-treatments estimate of σ^2 to be larger. In general, when the population means are not equal, the between-treatments estimate will overestimate the population variance σ^2.

The variation within each of the samples also has an effect on the conclusion we reach in analysis of variance. When a simple random sample is selected from each population, each of the sample variances provides an unbiased estimate of σ^2. Hence, we can combine or pool the individual estimates of σ^2 into one overall estimate. The estimate of σ^2 obtained in this way is called the *pooled* or *within-treatments* estimate of σ^2. Because each sample variance provides an estimate of σ^2 based only on the variation within each sample, the within-treatments estimate of σ^2 is not affected by whether the population means are equal. When the sample sizes are equal, the within-treatments estimate of σ^2 can be obtained by computing the average of the individual sample variances. For the Chemitech experiment we obtain

$$\text{Within-treatments estimate of } \sigma^2 = \frac{27.5 + 26.5 + 31.0}{3} = \frac{85}{3} = 28.33$$

In the Chemitech experiment, the between-treatments estimate of σ^2 (260) is much larger than the within-treatments estimate of σ^2 (28.33). In fact, the ratio of these two estimates is $260/28.33 = 9.18$. Recall, however, that the between-treatments approach provides a good estimate of σ^2 only if the null hypothesis is true; if the null hypothesis is false, the between-treatments approach overestimates σ^2. The within-treatments approach provides a good estimate of σ^2 in either case. Thus, if the null hypothesis is true, the two estimates will be similar and their ratio will be close to 1. If the null hypothesis is false, the between-treatments estimate will be larger than the within-treatments estimate, and their ratio will be large. In the next section we will show how large this ratio must be to reject H_0.

In summary, the logic behind ANOVA is based on the development of two independent estimates of the common population variance σ^2. One estimate of σ^2 is based on the variability among the sample means themselves, and the other estimate of σ^2 is based on the variability of the data within each sample. By comparing these two estimates of σ^2, we will be able to determine whether the population means are equal.

NOTES AND COMMENTS

1. Randomization in experimental design is the analog of probability sampling in an observational study.
2. In many medical experiments, potential bias is eliminated by using a double-blind experimental design. With this design, neither the physician applying the treatment nor the subject knows which treatment is being applied. Many other types of experiments could benefit from this type of design.
3. In this section we provided a conceptual overview of how analysis of variance can be used to test for the equality of k population means for a completely randomized experimental design. We will see that the same procedure can also be used to test for the equality of k population means for an observational or nonexperimental study.
4. In Sections 10.1 and 10.2 we presented statistical methods for testing the hypothesis that the means of two populations are equal. ANOVA can also be used to test the hypothesis that the means of two populations are equal. In practice, however, analysis of variance is usually not used except when dealing with three or more population means.

13.2 Analysis of Variance and the Completely Randomized Design

In this section we show how analysis of variance can be used to test for the equality of k population means for a completely randomized design. The general form of the hypotheses tested is

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$
$$H_a: \text{Not all population means are equal}$$

where

$$\mu_j = \text{mean of the } j\text{th population}$$

We assume that a simple random sample of size n_j has been selected from each of the k populations or treatments. For the resulting sample data, let

$$x_{ij} = \text{value of observation } i \text{ for treatment } j$$
$$n_j = \text{number of observations for treatment } j$$
$$\bar{x}_j = \text{sample mean for treatment } j$$
$$s_j^2 = \text{sample variance for treatment } j$$
$$s_j = \text{sample standard deviation for treatment } j$$

The formulas for the sample mean and sample variance for treatment j are as follow.

$$\bar{x}_j = \frac{\sum_{i=1}^{n_j} x_{ij}}{n_j} \tag{13.1}$$

$$s_j^2 = \frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1} \tag{13.2}$$

The overall sample mean, denoted $\bar{\bar{x}}$, is the sum of all the observations divided by the total number of observations. That is,

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}}{n_T} \tag{13.3}$$

where

$$n_T = n_1 + n_2 + \cdots + n_k \tag{13.4}$$

If the size of each sample is n, $n_T = kn$; in this case equation (13.3) reduces to

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}}{kn} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} x_{ij}/n}{k} = \frac{\sum_{j=1}^{k} \bar{x}_j}{k} \tag{13.5}$$

In other words, whenever the sample sizes are the same, the overall sample mean is just the average of the k sample means.

Because each sample in the Chemitech experiment consists of $n = 5$ observations, the overall sample mean can be computed by using equation (13.5). For the data in Table 13.1 we obtained the following result.

$$\bar{\bar{x}} = \frac{62 + 66 + 52}{3} = 60$$

If the null hypothesis is true ($\mu_1 = \mu_2 = \mu_3 = \mu$), the overall sample mean of 60 is the best estimate of the population mean μ.

Between-Treatments Estimate of Population Variance

In the preceding section, we introduced the concept of a between-treatments estimate of σ^2 and showed how to compute it when the sample sizes were equal. This estimate of σ^2 is called the *mean square due to treatments* and is denoted MSTR. The general formula for computing MSTR is

$$\text{MSTR} = \frac{\sum_{j=1}^{k} n_j (\bar{x}_j - \bar{\bar{x}})^2}{k - 1} \tag{13.6}$$

The numerator in equation (13.6) is called the *sum of squares due to treatments* and is denoted SSTR. The denominator, $k - 1$, represents the degrees of freedom associated with SSTR. Hence, the mean square due to treatments can be computed using the following formula.

MEAN SQUARE DUE TO TREATMENTS

$$\text{MSTR} = \frac{\text{SSTR}}{k - 1} \tag{13.7}$$

where

$$\text{SSTR} = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2 \tag{13.8}$$

If H_0 is true, MSTR provides an unbiased estimate of σ^2. However, if the means of the k populations are not equal, MSTR is not an unbiased estimate of σ^2; in fact, in that case, MSTR should overestimate σ^2.

For the Chemitech data in Table 13.1, we obtain the following results.

$$\text{SSTR} = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2 = 5(62 - 60)^2 + 5(66 - 60)^2 + 5(52 - 60)^2 = 520$$

$$\text{MSTR} = \frac{\text{SSTR}}{k - 1} = \frac{520}{2} = 260$$

Within-Treatments Estimate of Population Variance

Earlier, we introduced the concept of a within-treatments estimate of σ^2 and showed how to compute it when the sample sizes were equal. This estimate of σ^2 is called the *mean square due to error* and is denoted MSE. The general formula for computing MSE is

$$\text{MSE} = \frac{\sum_{j=1}^{k}(n_j - 1)s_j^2}{n_T - k} \tag{13.9}$$

The numerator in equation (13.9) is called the *sum of squares due to error* and is denoted SSE. The denominator of MSE is referred to as the degrees of freedom associated with SSE. Hence, the formula for MSE can also be stated as follows.

MEAN SQUARE DUE TO ERROR

$$\text{MSE} = \frac{\text{SSE}}{n_T - k} \tag{13.10}$$

where

$$\text{SSE} = \sum_{j=1}^{k}(n_j - 1)s_j^2 \tag{13.11}$$

Note that MSE is based on the variation within each of the treatments; it is not influenced by whether the null hypothesis is true. Thus, MSE always provides an unbiased estimate of σ^2.

For the Chemitech data in Table 13.1 we obtain the following results.

$$\text{SSE} = \sum_{j=1}^{k}(n_j - 1)s_j^2 = (5 - 1)27.5 + (5 - 1)26.5 + (5 - 1)31 = 340$$

$$\text{MSE} = \frac{\text{SSE}}{n_T - k} = \frac{340}{15 - 3} = \frac{340}{12} = 28.33$$

Comparing the Variance Estimates: The F Test

An introduction to the F distribution and the use of the F distribution table were presented in Section 11.2.

If the null hypothesis is true, MSTR and MSE provide two independent, unbiased estimates of σ^2. Based on the material covered in Chapter 11 we know that for normal populations, the sampling distribution of the ratio of two independent estimates of σ^2 follows an F distribution. Hence, if the null hypothesis is true and the ANOVA assumptions are valid, the sampling distribution of MSTR/MSE is an F distribution with numerator degrees of freedom equal to $k - 1$ and denominator degrees of freedom equal to $n_T - k$. In other words, if the null hypothesis is true, the value of MSTR/MSE should appear to have been selected from this F distribution.

However, if the null hypothesis is false, the value of MSTR/MSE will be inflated because MSTR overestimates σ^2. Hence, we will reject H_0 if the resulting value of MSTR/MSE appears to be too large to have been selected from an F distribution with $k - 1$ numerator degrees of freedom and $n_T - k$ denominator degrees of freedom. Because the decision to reject H_0 is based on the value of MSTR/MSE, the test statistic used to test for the equality of k population means is as follows.

TEST STATISTIC FOR THE EQUALITY OF k POPULATION MEANS

$$F = \frac{\text{MSTR}}{\text{MSE}} \qquad \textbf{(13.12)}$$

The test statistic follows an F distribution with $k - 1$ degrees of freedom in the numerator and $n_T - k$ degrees of freedom in the denominator.

Let us return to the Chemitech experiment and use a level of significance $\alpha = .05$ to conduct the hypothesis test. The value of the test statistic is

$$F = \frac{\text{MSTR}}{\text{MSE}} = \frac{260}{28.33} = 9.18$$

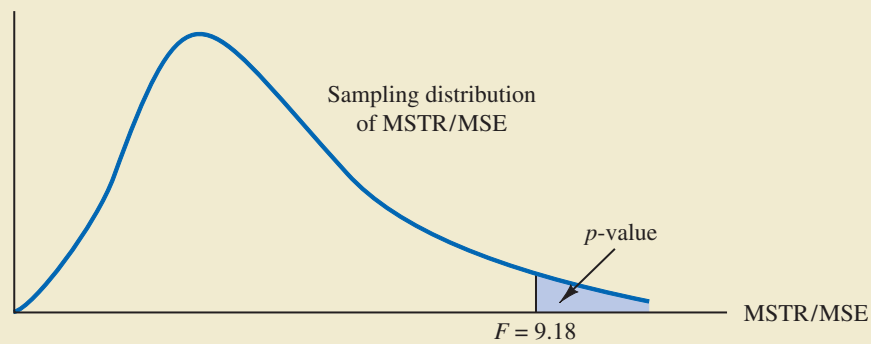
The numerator degrees of freedom is $k - 1 = 3 - 1 = 2$ and the denominator degrees of freedom is $n_T - k = 15 - 3 = 12$. Because we will only reject the null hypothesis for large values of the test statistic, the p-value is the upper tail area of the F distribution to the right of the test statistic $F = 9.18$. Figure 13.4 shows the sampling distribution of $F = \text{MSTR/MSE}$, the value of the test statistic, and the upper tail area that is the p-value for the hypothesis test.

From Table 4 of Appendix B we find the following areas in the upper tail of an F distribution with 2 numerator degrees of freedom and 12 denominator degrees of freedom.

Area in Upper Tail	**.10**	**.05**	**.025**	**.01**
***F* Value ($df_1 = 2, df_2 = 12$)**	2.81	3.89	5.10	6.93

$F = 9.18$

FIGURE 13.4 COMPUTATION OF p-VALUE USING THE SAMPLING DISTRIBUTION OF MSTR/MSE



Appendix F shows how to compute p-values using Minitab or Excel.

Because $F = 9.18$ is greater than 6.93, the area in the upper tail at $F = 9.18$ is less than .01. Thus, the p-value is less than .01. Minitab or Excel can be used to show that the exact p-value is .004. With p-value $\leq \alpha = .05$, H_0 is rejected. The test provides sufficient evidence to conclude that the means of the three populations are not equal. In other words, analysis of variance supports the conclusion that the population mean number of units produced per week for the three assembly methods are not equal.

As with other hypothesis testing procedures, the critical value approach may also be used. With $\alpha = .05$, the critical F value occurs with an area of .05 in the upper tail of an F distribution with 2 and 12 degrees of freedom. From the F distribution table, we find $F_{.05} = 3.89$. Hence, the appropriate upper tail rejection rule for the Chemitech experiment is

$$\text{Reject } H_0 \text{ if } F \geq 3.89$$

With $F = 9.18$, we reject H_0 and conclude that the means of the three populations are not equal. A summary of the overall procedure for testing for the equality of k population means follows.

TEST FOR THE EQUALITY OF k POPULATION MEANS

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$
$$H_a: \text{Not all population means are equal}$$

TEST STATISTIC

$$F = \frac{\text{MSTR}}{\text{MSE}}$$

REJECTION RULE

p-value approach:	Reject H_0 if p-value $\leq \alpha$
Critical value approach:	Reject H_0 if $F \geq F_\alpha$

where the value of F_α is based on an F distribution with $k - 1$ numerator degrees of freedom and $n_T - k$ denominator degrees of freedom.

ANOVA Table

The results of the preceding calculations can be displayed conveniently in a table referred to as the analysis of variance or **ANOVA table**. The general form of the ANOVA table for a completely randomized design is shown in Table 13.2; Table 13.3 is the corresponding ANOVA table for the Chemitech experiment. The sum of squares associated with the source of variation referred to as "Total" is called the total sum of squares (SST). Note that the results for the Chemitech experiment suggest that SST = SSTR + SSE, and that the degrees of freedom associated with this total sum of squares is the sum of the degrees of freedom associated with the sum of squares due to treatments and the sum of squares due to error.

We point out that SST divided by its degrees of freedom $n_T - 1$ is nothing more than the overall sample variance that would be obtained if we treated the entire set of 15 observations as one data set. With the entire data set as one sample, the formula for computing the total sum of squares, SST, is

$$\text{SST} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{\bar{x}})^2 \qquad \textbf{(13.13)}$$

It can be shown that the results we observed for the analysis of variance table for the Chemitech experiment also apply to other problems. That is,

$$\text{SST} = \text{SSTR} + \text{SSE} \qquad \textbf{(13.14)}$$

Analysis of variance can be thought of as a statistical procedure for partitioning the total sum of squares into separate components.

In other words, SST can be partitioned into two sums of squares: the sum of squares due to treatments and the sum of squares due to error. Note also that the degrees of freedom corresponding to SST, $n_T - 1$, can be partitioned into the degrees of freedom corresponding to SSTR, $k - 1$, and the degrees of freedom corresponding to SSE, $n_T - k$. The analysis of variance can be viewed as the process of **partitioning** the total sum of squares and the degrees of freedom into their corresponding sources: treatments and error. Dividing the sum of squares by the appropriate degrees of freedom provides the variance estimates, the F value, and the p-value used to test the hypothesis of equal population means.

TABLE 13.2 ANOVA TABLE FOR A COMPLETELY RANDOMIZED DESIGN

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Treatments	SSTR	$k - 1$	$\text{MSTR} = \dfrac{\text{SSTR}}{k-1}$	$\dfrac{\text{MSTR}}{\text{MSE}}$	
Error	SSE	$n_T - k$	$\text{MSE} = \dfrac{\text{SSE}}{n_T - k}$		
Total	SST	$n_T - 1$			

TABLE 13.3 ANALYSIS OF VARIANCE TABLE FOR THE CHEMITECH EXPERIMENT

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Treatments	520	2	260.00	9.18	.004
Error	340	12	28.33		
Total	860	14			

FIGURE 13.5 MINITAB OUTPUT FOR THE CHEMITECH EXPERIMENT ANALYSIS OF VARIANCE

```
Source     DF        SS        MS        F        P
Factor      2     520.0     260.0     9.18    0.004
Error      12     340.0      28.3
Total      14     860.0

S = 5.323     R-Sq = 60.47%     R-Sq(adj) = 53.88%

                                 Individual 95% CIs For Mean Based on
                                 Pooled StDev
 Level      N      Mean    StDev  ---+---------+---------+---------+------
 A          5    62.000    5.244                (-------*-------)
 B          5    66.000    4.148                        (------*-------)
 C          5    52.000    5.568  (------*-------)
                                  ---+---------+---------+---------+------
Pooled StDev = 5.323               49.0      56.0      63.0      70.0
```

Computer Results for Analysis of Variance

Using statistical computer packages, analysis of variance computations with large sample sizes or a large number of populations can be performed easily. Appendixes 13.1 – 13.3 show the steps required to use Minitab, Excel, and StatTools to perform the analysis of variance computations. In Figure 13.5 we show output for the Chemitech experiment obtained using Minitab. The first part of the computer output contains the familiar ANOVA table format. Comparing Figure 13.5 with Table 13.3, we see that the same information is available, although some of the headings are slightly different. The heading Source is used for the source of variation column, Factor identifies the treatments row, and the sum of squares and degrees of freedom columns are interchanged.

Note that following the ANOVA table the computer output contains the respective sample sizes, the sample means, and the standard deviations. In addition, Minitab provides a figure that shows individual 95% confidence interval estimates of each population mean. In developing these confidence interval estimates, Minitab uses MSE as the estimate of σ^2. Thus, the square root of MSE provides the best estimate of the population standard deviation σ. This estimate of σ on the computer output is Pooled StDev; it is equal to 5.323. To provide an illustration of how these interval estimates are developed, we will compute a 95% confidence interval estimate of the population mean for method A.

From our study of interval estimation in Chapter 8, we know that the general form of an interval estimate of a population mean is

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \tag{13.15}$$

where s is the estimate of the population standard deviation σ. Because the best estimate of σ is provided by the Pooled StDev, we use a value of 5.323 for s in expression (13.15). The degrees of freedom for the t value is 12, the degrees of freedom associated with the error sum of squares. Hence, with $t_{.025} = 2.179$ we obtain

$$62 \pm 2.179 \frac{5.323}{\sqrt{5}} = 62 \pm 5.19$$

Thus, the individual 95% confidence interval for method A goes from $62 - 5.19 = 56.81$ to $62 + 5.19 = 67.19$. Because the sample sizes are equal for the Chemitech experiment, the individual confidence intervals for methods B and C are also constructed by adding and subtracting 5.19 from each sample mean. Thus, in the figure provided by Minitab we see that the widths of the confidence intervals are the same.

Testing for the Equality of k Population Means: An Observational Study

We have shown how analysis of variance can be used to test for the equality of k population means for a completely randomized experimental design. It is important to understand that ANOVA can also be used to test for the equality of three or more population means using data obtained from an observational study. As an example, let us consider the situation at National Computer Products, Inc. (NCP).

NCP manufactures printers and fax machines at plants located in Atlanta, Dallas, and Seattle. To measure how much employees at these plants know about quality management, a random sample of six employees was selected from each plant and the employees selected were given a quality awareness examination. The examination scores for these 18 employees are shown in Table 13.4. The sample means, sample variances, and sample standard deviations for each group are also provided. Managers want to use these data to test the hypothesis that the mean examination score is the same for all three plants.

We define population 1 as all employees at the Atlanta plant, population 2 as all employees at the Dallas plant, and population 3 as all employees at the Seattle plant. Let

μ_1 = mean examination score for population 1
μ_2 = mean examination score for population 2
μ_3 = mean examination score for population 3

Although we will never know the actual values of μ_1, μ_2, and μ_3, we want to use the sample results to test the following hypotheses.

H_0: $\mu_1 = \mu_2 = \mu_3$
H_a: Not all population means are equal

Note that the hypothesis test for the NCP observational study is exactly the same as the hypothesis test for the Chemitech experiment. Indeed, the same analysis of variance

TABLE 13.4 EXAMINATION SCORES FOR 18 EMPLOYEES

WEB file NCP

	Plant 1 Atlanta	Plant 2 Dallas	Plant 3 Seattle
	85	71	59
	75	75	64
	82	73	62
	76	74	69
	71	69	75
	85	82	67
Sample mean	79	74	66
Sample variance	34	20	32
Sample standard deviation	5.83	4.47	5.66

Exercise 8 will ask you to analyze the NCP data using the analysis of variance procedure.

methodology we used to analyze the Chemitech experiment can also be used to analyze the data from the NCP observational study.

Even though the same ANOVA methodology is used for the analysis, it is worth noting how the NCP observational statistical study differs from the Chemitech experimental statistical study. The individuals who conducted the NCP study had no control over how the plants were assigned to individual employees. That is, the plants were already in operation and a particular employee worked at one of the three plants. All that NCP could do was to select a random sample of 6 employees from each plant and administer the quality awareness examination. To be classified as an experimental study, NCP would have had to be able to randomly select 18 employees and then assign the plants to each employee in a random fashion.

NOTES AND COMMENTS

1. The overall sample mean can also be computed as a weighted average of the k sample means.

$$\bar{\bar{x}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n_T}$$

In problems where the sample means are provided, this formula is simpler than equation (13.3) for computing the overall mean.

2. If each sample consists of n observations, equation (13.6) can be written as

$$\text{MSTR} = \frac{n\sum_{j=1}^{k}(\bar{x}_j - \bar{\bar{x}})^2}{k-1} = n\left[\frac{\sum_{j=1}^{k}(\bar{x}_j - \bar{\bar{x}})^2}{k-1}\right] = ns_{\bar{x}}^2$$

Note that this result is the same as presented in Section 13.1 when we introduced the concept of the between-treatments estimate of σ^2. Equation (13.6) is simply a generalization of this result to the unequal sample-size case.

3. If each sample has n observations, $n_T = kn$; thus, $n_T - k = k(n-1)$, and equation (13.9) can be rewritten as

$$\text{MSE} = \frac{\sum_{j=1}^{k}(n-1)s_j^2}{k(n-1)} = \frac{(n-1)\sum_{j=1}^{k}s_j^2}{k(n-1)} = \frac{\sum_{j=1}^{k}s_j^2}{k}$$

In other words, if the sample sizes are the same, MSE is just the average of the k sample variances. Note that it is the same result we used in Section 13.1 when we introduced the concept of the within-treatments estimate of σ^2.

Exercises

Methods

SELF test

1. The following data are from a completely randomized design.

	Treatment		
	A	B	C
	162	142	126
	142	156	122
	165	124	138
	145	142	140
	148	136	150
	174	152	128
Sample mean	156	142	134
Sample variance	164.4	131.2	110.4

a. Compute the sum of squares between treatments.
b. Compute the mean square between treatments.

c. Compute the sum of squares due to error.
d. Compute the mean square due to error.
e. Set up the ANOVA table for this problem.
f. At the $\alpha = .05$ level of significance, test whether the means for the three treatments are equal.

2. In a completely randomized design, seven experimental units were used for each of the five levels of the factor. Complete the following ANOVA table.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*-value
Treatments	300				
Error					
Total	460				

3. Refer to exercise 2.
   a. What hypotheses are implied in this problem?
   b. At the $\alpha = .05$ level of significance, can we reject the null hypothesis in part (a)? Explain.
4. In an experiment designed to test the output levels of three different treatments, the following results were obtained: SST = 400, SSTR = 150, $n_T = 19$. Set up the ANOVA table and test for any significant difference between the mean output levels of the three treatments. Use $\alpha = .05$.
5. In a completely randomized design, 12 experimental units were used for the first treatment, 15 for the second treatment, and 20 for the third treatment. Complete the following analysis of variance. At a .05 level of significance, is there a significant difference between the treatments?

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	*F*	*p*-value
Treatments	1200				
Error					
Total	1800				

6. Develop the analysis of variance computations for the following completely randomized design. At $\alpha = .05$, is there a significant difference between the treatment means?

WEB file
Exer6

	Treatment		
	A	B	C
	136	107	92
	120	114	82
	113	125	85
	107	104	101
	131	107	89
	114	109	117
	129	97	110
	102	114	120
		104	98
		89	106
$\bar{x}_j$	119	107	100
s_j^2	146.86	96.44	173.78

Applications

7. Three different methods for assembling a product were proposed by an industrial engineer. To investigate the number of units assembled correctly with each method, 30 employees were randomly selected and randomly assigned to the three proposed methods in such a way that each method was used by 10 workers. The number of units assembled correctly was recorded, and the analysis of variance procedure was applied to the resulting data set. The following results were obtained: SST = 10,800; SSTR = 4560.
   a. Set up the ANOVA table for this problem.
   b. Use $\alpha = .05$ to test for any significant difference in the means for the three assembly methods.
8. Refer to the NCP data in Table 13.4. Set up the ANOVA table and test for any significant difference in the mean examination score for the three plants. Use $\alpha = .05$.
9. To study the effect of temperature on yield in a chemical process, five batches were produced at each of three temperature levels. The results follow. Construct an analysis of variance table. Use a .05 level of significance to test whether the temperature level has an effect on the mean yield of the process.

Temperature		
50° C	**60° C**	**70° C**
34	30	23
24	31	28
36	34	28
39	23	30
32	27	31

10. Auditors must make judgments about various aspects of an audit on the basis of their own direct experience, indirect experience, or a combination of the two. In a study, auditors were asked to make judgments about the frequency of errors to be found in an audit. The judgments by the auditors were then compared to the actual results. Suppose the following data were obtained from a similar study; lower scores indicate better judgments.

WEB file
AudJudg

Direct	Indirect	Combination
17.0	16.6	25.2
18.5	22.2	24.0
15.8	20.5	21.5
18.2	18.3	26.8
20.2	24.2	27.5
16.0	19.8	25.8
13.3	21.2	24.2

Use $\alpha = .05$ to test to see whether the basis for the judgment affects the quality of the judgment. What is your conclusion?

11. Four different paints are advertised as having the same drying time. To check the manufacturer's claims, five samples were tested for each of the paints. The time in minutes until the paint was dry enough for a second coat to be applied was recorded. The following data were obtained.

WEB file
Paint

Paint 1	Paint 2	Paint 3	Paint 4
128	144	133	150
137	133	143	142
135	142	137	135
124	146	136	140
141	130	131	153

At the $\alpha = .05$ level of significance, test to see whether the mean drying time is the same for each type of paint.

12. The *Consumer Reports* Restaurant Customer Satisfaction Survey is based upon 148,599 visits to full-service restaurant chains (Consumer Reports website). One of the variables in the study is meal price, the average amount paid per person for dinner and drinks, minus the tip. Suppose a reporter for the *Sun Coast Times* thought that it would be of interest to her readers to conduct a similar study for restaurants located on the Grand Strand section in Myrtle Beach, South Carolina. The reporter selected a sample of eight seafood restaurants, eight Italian restaurants, and eight steakhouses. The following data show the meal prices ($) obtained for the 24 restaurants sampled. Use $\alpha = .05$ to test whether there is a significant difference among the mean meal price for the three types of restaurants.

WEB file
GrandStrand

Italian	Seafood	Steakhouse
$12	$16	$24
13	18	19
15	17	23
17	26	25
18	23	21
20	15	22
17	19	27
24	18	31

13.3 Multiple Comparison Procedures

When we use analysis of variance to test whether the means of k populations are equal, rejection of the null hypothesis allows us to conclude only that the population means are *not all equal.* In some cases we will want to go a step further and determine where the differences among means occur. The purpose of this section is to show how **multiple comparison procedures** can be used to conduct statistical comparisons between pairs of population means.

Fisher's LSD

Suppose that analysis of variance provides statistical evidence to reject the null hypothesis of equal population means. In this case, Fisher's least significant difference (LSD) procedure can be used to determine where the differences occur. To illustrate the use of Fisher's LSD procedure in making pairwise comparisons of population means, recall the Chemitech experiment introduced in Section 13.1. Using analysis of variance, we concluded that the mean number of units produced per week are not the same for the three assembly methods. In this case, the follow-up question is: We believe the assembly methods differ, but where do the differences occur? That is, do the means of populations 1 and 2 differ? Or those of populations 1 and 3? Or those of populations 2 and 3?

In Chapter 10 we presented a statistical procedure for testing the hypothesis that the means of two populations are equal. With a slight modification in how we estimate the

population variance, Fisher's LSD procedure is based on the t test statistic presented for the two-population case. The following table summarizes Fisher's LSD procedure.

FISHER'S LSD PROCEDURE

$$H_0: \mu_i = \mu_j$$
$$H_a: \mu_i \neq \mu_j$$

TEST STATISTIC

$$t = \frac{\bar{x}_i - \bar{x}_j}{\sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \qquad \textbf{(13.16)}$$

REJECTION RULE

p-value approach:	Reject H_0 if p-value $\leq \alpha$
Critical value approach:	Reject H_0 if $t \leq -t_{\alpha/2}$ or $t \geq t_{\alpha/2}$

where the value of $t_{\alpha/2}$ is based on a t distribution with $n_T - k$ degrees of freedom.

Let us now apply this procedure to determine whether there is a significant difference between the means of population 1 (method A) and population 2 (method B) at the $\alpha = .05$ level of significance. Table 13.1 showed that the sample mean is 62 for method A and 66 for method B. Table 13.3 showed that the value of MSE is 28.33; it is the estimate of σ^2 and is based on 12 degrees of freedom. For the Chemitech data the value of the test statistic is

$$t = \frac{62 - 66}{\sqrt{28.33\left(\frac{1}{5} + \frac{1}{5}\right)}} = -1.19$$

Because we have a two-tailed test, the p-value is two times the area under the curve for the t distribution to the left of $t = -1.19$. Using Table 2 in Appendix B, the t distribution table for 12 degrees of freedom provides the following information.

Area in Upper Tail	**.20**	**.10**	**.05**	**.025**	**.01**	**.005**
***t* Value (12 *df*)**	.873	1.356	1.782	2.179	2.681	3.055

$t = 1.19$

The t distribution table only contains positive t values. Because the t distribution is symmetric, however, we can find the area under the curve to the right of $t = 1.19$ and double it to find the p-value corresponding to $t = -1.19$. We see that $t = 1.19$ is between .20 and .10. Doubling these amounts, we see that the p-value must be between .40 and .20. Excel or Minitab can be used to show that the exact p-value is .2571. Because the p-value is greater than $\alpha = .05$, we cannot reject the null hypothesis. Hence, we cannot conclude that the population mean number of units produced per week for method A is different from the population mean for method B.

Appendix F shows how to compute p-values using Excel or Minitab.

Many practitioners find it easier to determine how large the difference between the sample means must be to reject H_0. In this case the test statistic is $\bar{x}_i - \bar{x}_j$, and the test is conducted by the following procedure.

FISHER'S LSD PROCEDURE BASED ON THE TEST STATISTIC $\bar{x}_i - \bar{x}_j$

$$H_0: \mu_i = \mu_j$$
$$H_a: \mu_i \neq \mu_j$$

TEST STATISTIC

$$\bar{x}_i - \bar{x}_j$$

REJECTION RULE AT A LEVEL OF SIGNIFICANCE α

$$\text{Reject } H_0 \text{ if } |\bar{x}_i - \bar{x}_j| \geq \text{LSD}$$

where

$$\text{LSD} = t_{\alpha/2}\sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \qquad \textbf{(13.17)}$$

For the Chemitech experiment the value of LSD is

$$\text{LSD} = 2.179\sqrt{28.33\left(\frac{1}{5} + \frac{1}{5}\right)} = 7.34$$

Note that when the sample sizes are equal, only one value for LSD is computed. In such cases we can simply compare the magnitude of the difference between any two sample means with the value of LSD. For example, the difference between the sample means for population 1 (method A) and population 3 (method C) is $62 - 52 = 10$. This difference is greater than LSD = 7.34, which means we can reject the null hypothesis that the population mean number of units produced per week for method A is equal to the population mean for method C. Similarly, with the difference between the sample means for populations 2 and 3 of $66 - 52 = 14 > 7.34$, we can also reject the hypothesis that the population mean for method B is equal to the population mean for method C. In effect, our conclusion is that methods A and B both differ from method C.

Fisher's LSD can also be used to develop a confidence interval estimate of the difference between the means of two populations. The general procedure follows.

CONFIDENCE INTERVAL ESTIMATE OF THE DIFFERENCE BETWEEN TWO POPULATION MEANS USING FISHER'S LSD PROCEDURE

$$\bar{x}_i - \bar{x}_j \pm \text{LSD} \qquad \textbf{(13.18)}$$

where

$$\text{LSD} = t_{\alpha/2}\sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \qquad \textbf{(13.19)}$$

and $t_{\alpha/2}$ is based on a t distribution with $n_T - k$ degrees of freedom.

If the confidence interval in expression (13.18) includes the value zero, we cannot reject the hypothesis that the two population means are equal. However, if the confidence interval does not include the value zero, we conclude that there is a difference between the population means. For the Chemitech experiment, recall that LSD = 7.34 (corresponding to $t_{.025} = 2.179$). Thus, a 95% confidence interval estimate of the difference between the means of populations 1 and 2 is $62 - 66 \pm 7.34 = -4 \pm 7.34 = -11.34$ to 3.34; because this interval includes zero, we cannot reject the hypothesis that the two population means are equal.

Type I Error Rates

We began the discussion of Fisher's LSD procedure with the premise that analysis of variance gave us statistical evidence to reject the null hypothesis of equal population means. We showed how Fisher's LSD procedure can be used in such cases to determine where the differences occur. Technically, it is referred to as a *protected* or *restricted* LSD test because it is employed only if we first find a significant F value by using analysis of variance. To see why this distinction is important in multiple comparison tests, we need to explain the difference between a *comparisonwise* Type I error rate and an *experimentwise* Type I error rate.

In the Chemitech experiment we used Fisher's LSD procedure to make three pairwise comparisons.

Test 1	Test 2	Test 3
$H_0: \mu_1 = \mu_2$	$H_0: \mu_1 = \mu_3$	$H_0: \mu_2 = \mu_3$
$H_a: \mu_1 \neq \mu_2$	$H_a: \mu_1 \neq \mu_3$	$H_a: \mu_2 \neq \mu_3$

In each case, we used a level of significance of $\alpha = .05$. Therefore, for each test, if the null hypothesis is true, the probability that we will make a Type I error is $\alpha = .05$; hence, the probability that we will not make a Type I error on each test is $1 - .05 = .95$. In discussing multiple comparison procedures we refer to this probability of a Type I error ($\alpha = .05$) as the **comparisonwise Type I error rate**; comparisonwise Type I error rates indicate the level of significance associated with a single pairwise comparison.

Let us now consider a slightly different question. What is the probability that in making three pairwise comparisons, we will commit a Type I error on at least one of the three tests? To answer this question, note that the probability that we will not make a Type I error on any of the three tests is $(.95)(.95)(.95) = .8574$.[1] Therefore, the probability of making at least one Type I error is $1 - .8574 = .1426$. Thus, when we use Fisher's LSD procedure to make all three pairwise comparisons, the Type I error rate associated with this approach is not .05, but actually .1426; we refer to this error rate as the *overall* or **experimentwise Type I error rate**. To avoid confusion, we denote the experimentwise Type I error rate as α_{EW}.

The experimentwise Type I error rate gets larger for problems with more populations. For example, a problem with five populations has 10 possible pairwise comparisons. If we tested all possible pairwise comparisons by using Fisher's LSD with a comparisonwise error rate of $\alpha = .05$, the experimentwise Type I error rate would be $1 - (1 - .05)^{10} = .40$. In such cases, practitioners look to alternatives that provide better control over the experimentwise error rate.

One alternative for controlling the overall experimentwise error rate, referred to as the Bonferroni adjustment, involves using a smaller comparisonwise error rate for each test. For example, if we want to test C pairwise comparisons and want the maximum probability of

[1]The assumption is that the three tests are independent, and hence the joint probability of the three events can be obtained by simply multiplying the individual probabilities. In fact, the three tests are not independent because MSE is used in each test; therefore, the error involved is even greater than that shown.

making a Type I error for the overall experiment to be α_{EW}, we simply use a comparisonwise error rate equal to α_{EW}/C. In the Chemitech experiment, if we want to use Fisher's LSD procedure to test all three pairwise comparisons with a maximum experimentwise error rate of $\alpha_{EW} = .05$, we set the comparisonwise error rate to be $\alpha = .05/3 = .017$. For a problem with five populations and 10 possible pairwise comparisons, the Bonferroni adjustment would suggest a comparisonwise error rate of $.05/10 = .005$. Recall from our discussion of hypothesis testing in Chapter 9 that for a fixed sample size, any decrease in the probability of making a Type I error will result in an increase in the probability of making a Type II error, which corresponds to accepting the hypothesis that the two population means are equal when in fact they are not equal. As a result, many practitioners are reluctant to perform individual tests with a low comparisonwise Type I error rate because of the increased risk of making a Type II error.

Several other procedures, such as Tukey's procedure and Duncan's multiple range test, have been developed to help in such situations. However, there is considerable controversy in the statistical community as to which procedure is "best." The truth is that no one procedure is best for all types of problems.

Exercises

Methods

SELF test

13. The following data are from a completely randomized design.

	Treatment A	Treatment B	Treatment C
	32	44	33
	30	43	36
	30	44	35
	26	46	36
	32	48	40
Sample mean	30	45	36
Sample variance	6.00	4.00	6.50

a. At the $\alpha = .05$ level of significance, can we reject the null hypothesis that the means of the three treatments are equal?
b. Use Fisher's LSD procedure to test whether there is a significant difference between the means for treatments A and B, treatments A and C, and treatments B and C. Use $\alpha = .05$.
c. Use Fisher's LSD procedure to develop a 95% confidence interval estimate of the difference between the means of treatments A and B.

14. The following data are from a completely randomized design. In the following calculations, use $\alpha = .05$.

	Treatment 1	Treatment 2	Treatment 3
	63	82	69
	47	72	54
	54	88	61
	40	66	48
$\bar{x}_j$	51	77	58
s_j^2	96.67	97.34	81.99

a. Use analysis of variance to test for a significant difference among the means of the three treatments.
b. Use Fisher's LSD procedure to determine which means are different.

Applications

SELF test

15. To test whether the mean time needed to mix a batch of material is the same for machines produced by three manufacturers, the Jacobs Chemical Company obtained the following data on the time (in minutes) needed to mix the material.

Manufacturer		
1	2	3
20	28	20
26	26	19
24	31	23
22	27	22

a. Use these data to test whether the population mean times for mixing a batch of material differ for the three manufacturers. Use $\alpha = .05$.
b. At the $\alpha = .05$ level of significance, use Fisher's LSD procedure to test for the equality of the means for manufacturers 1 and 3. What conclusion can you draw after carrying out this test?

SELF test

16. Refer to exercise 15. Use Fisher's LSD procedure to develop a 95% confidence interval estimate of the difference between the means for manufacturer 1 and manufacturer 2.

17. The following data are from an experiment designed to investigate the perception of corporate ethical values among individuals specializing in marketing (higher scores indicate higher ethical values).

Marketing Managers	Marketing Research	Advertising
6	5	6
5	5	7
4	4	6
5	4	5
6	5	6
4	4	6

a. Use $\alpha = .05$ to test for significant differences in perception among the three groups.
b. At the $\alpha = .05$ level of significance, we can conclude that there are differences in the perceptions for marketing managers, marketing research specialists, and advertising specialists. Use the procedures in this section to determine where the differences occur. Use $\alpha = .05$.

18. To test for any significant difference in the number of hours between breakdowns for four machines, the following data were obtained.

Machine 1	Machine 2	Machine 3	Machine 4
6.4	8.7	11.1	9.9
7.8	7.4	10.3	12.8
5.3	9.4	9.7	12.1
7.4	10.1	10.3	10.8
8.4	9.2	9.2	11.3
7.3	9.8	8.8	11.5

a. At the $\alpha = .05$ level of significance, what is the difference, if any, in the population mean times among the four machines?
b. Use Fisher's LSD procedure to test for the equality of the means for machines 2 and 4. Use a .05 level of significance.

19. Refer to exercise 18. Use the Bonferroni adjustment to test for a significant difference between all pairs of means. Assume that a maximum overall experimentwise error rate of .05 is desired.

20. The International League of Triple-A minor league baseball consists of 14 teams organized into three divisions: North, South, and West. The following data show the average attendance for the 14 teams in the International League (The Biz of Baseball website, January 2009). Also shown are the teams' records; W denotes the number of games won, L denotes the number of games lost, and PCT is the proportion of games played that were won.

WEB file
Triple-A

Team Name	Division	W	L	PCT	Attendance
Buffalo Bisons	North	66	77	.462	8812
Lehigh Valley IronPigs	North	55	89	.382	8479
Pawtucket Red Sox	North	85	58	.594	9097
Rochester Red Wings	North	74	70	.514	6913
Scranton-Wilkes Barre Yankees	North	88	56	.611	7147
Syracuse Chiefs	North	69	73	.486	5765
Charlotte Knights	South	63	78	.447	4526
Durham Bulls	South	74	70	.514	6995
Norfolk Tides	South	64	78	.451	6286
Richmond Braves	South	63	78	.447	4455
Columbus Clippers	West	69	73	.486	7795
Indianapolis Indians	West	68	76	.472	8538
Louisville Bats	West	88	56	.611	9152
Toledo Mud Hens	West	75	69	.521	8234

a. Use $\alpha = .05$ to test for any difference in the mean attendance for the three divisions.
b. Use Fisher's LSD procedure to determine where the differences occur. Use $\alpha = .05$.

13.4 Randomized Block Design

Thus far we have considered the completely randomized experimental design. Recall that to test for a difference among treatment means, we computed an F value by using the ratio

$$F = \frac{\text{MSTR}}{\text{MSE}} \tag{13.20}$$

A completely randomized design is useful when the experimental units are homogeneous. If the experimental units are heterogeneous, **blocking** *is often used to form homogeneous groups.*

A problem can arise whenever differences due to extraneous factors (ones not considered in the experiment) cause the MSE term in this ratio to become large. In such cases, the F value in equation (13.20) can become small, signaling no difference among treatment means when in fact such a difference exists.

In this section we present an experimental design known as a **randomized block design**. Its purpose is to control some of the extraneous sources of variation by removing such variation from the MSE term. This design tends to provide a better estimate of the true error variance and leads to a more powerful hypothesis test in terms of the ability to detect