CHAPTER 13

Experimental Design and Analysis of Variance

CONTENTS

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.

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

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.

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.

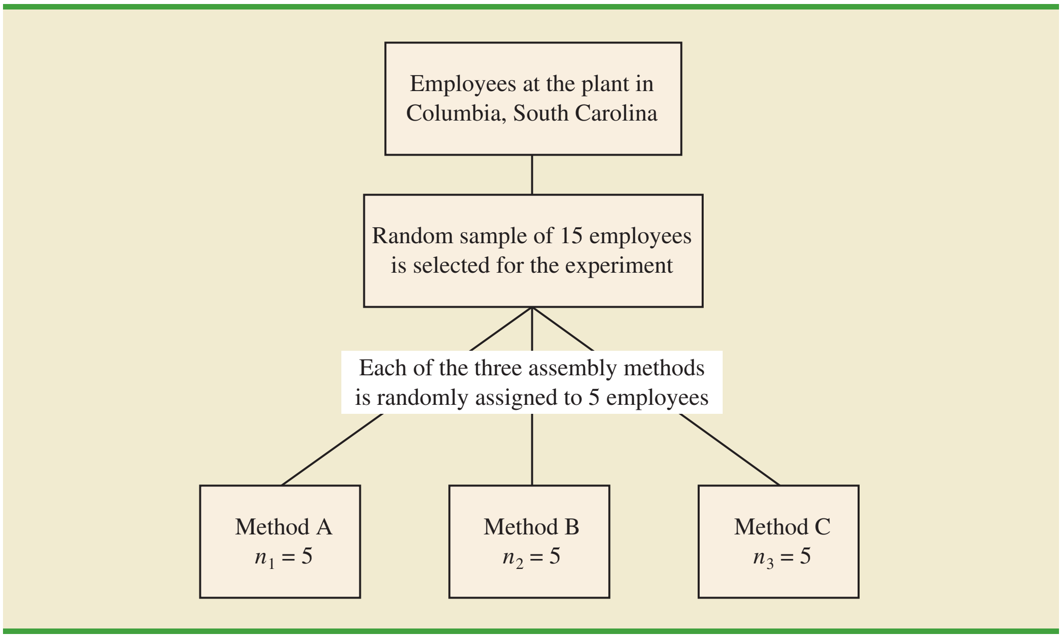
In the Chemitech experiment, assembly method is the independent variable or **factor**. Because three assembly methods correspond to this factor, we say that three treatments are associated with this experiment; each **treatment** corresponds to one of the three assembly methods. The Chemitech problem 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.

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

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.

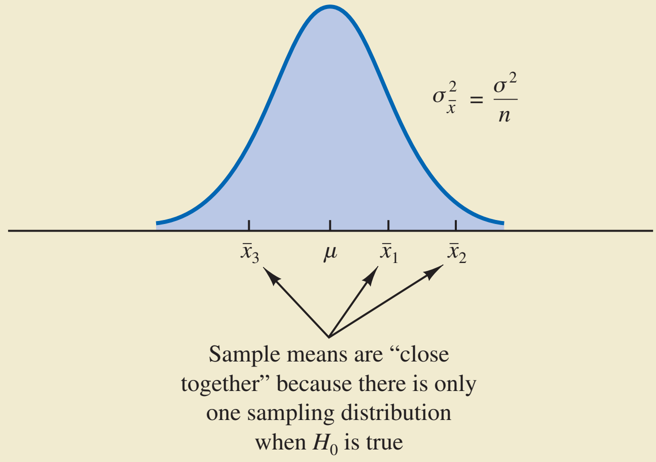
1. **For each population, the response variable is normally distributed.** Implication: In the Chemitech experiment the number of units produced per week (response variable) must be normally distributed for each assembly method.
2. **The variance of the response variable, denoted σ^2, is the same for 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 $\bar{x}$, $\sigma^2_{\bar{x}}$, is provided by the variance of the three sample means.

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

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

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

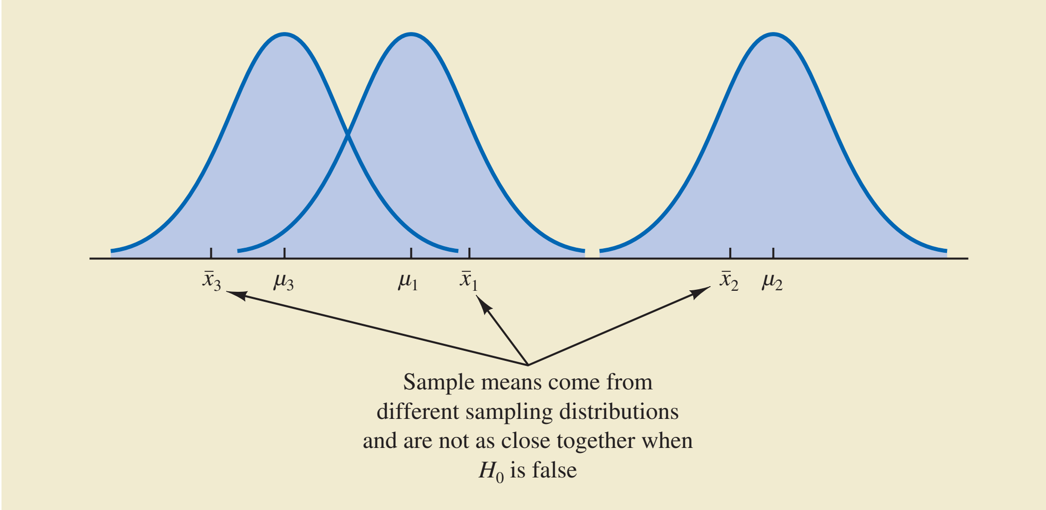
Hence,

$$\text{Estimate of } \sigma^2 = n\,(\text{Estimate of } \sigma^2_{\bar{x}}) = ns^2_{\bar{x}} = 5(52) = 260$$

The result, $ns^2_{\bar{x}} = 260$, 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

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

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

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

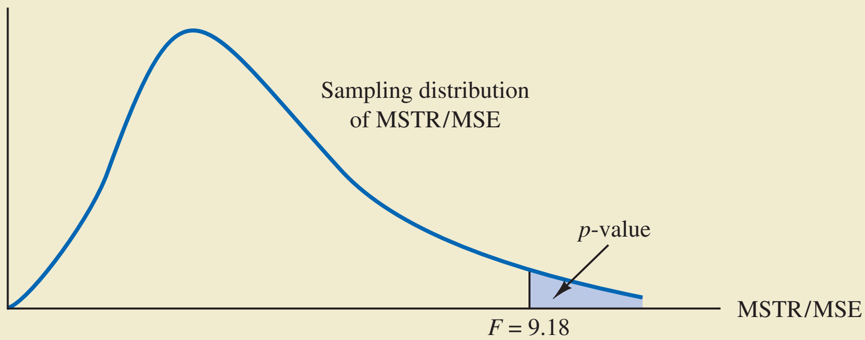
The numerator degrees of freedom is $k - 1 = 3 - 1 = 2$ and the denominator degrees of freedom is $n_T - k = 15 - 3 = 12$. Because we will only reject the null hypothesis for large values of the test statistic, the p-value is the upper tail area of the F distribution to the right of the test statistic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

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

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

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

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,

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

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

$$\mu_1 = \text{mean examination score for population 1}$$
$$\mu_2 = \text{mean examination score for population 2}$$
$$\mu_3 = \text{mean examination score for population 3}$$

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.

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

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

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

TEST STATISTIC

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

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

$t = 1.19$

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

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.

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

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

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

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

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

$$\bar{x}_i - \bar{x}_j \pm \text{LSD} \tag{13.18}$$

where

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

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 differences among treatment means. To illustrate, let us consider a stress study for air traffic controllers.

Air Traffic Controller Stress Test

A study measuring the fatigue and stress of air traffic controllers resulted in proposals for modification and redesign of the controller's work station. After consideration of several designs for the work station, three specific alternatives are selected as having the best potential for reducing controller stress. The key question is: To what extent do the three alternatives differ in terms of their effect on controller stress? To answer this question, we need to design an experiment that will provide measurements of air traffic controller stress under each alternative.

Experimental studies in business often involve experimental units that are highly heterogeneous; as a result, randomized block designs are often employed.

In a completely randomized design, a random sample of controllers would be assigned to each work station alternative. However, controllers are believed to differ substantially in their ability to handle stressful situations. What is high stress to one controller might be only moderate or even low stress to another. Hence, when considering the within-group source of variation (MSE), we must realize that this variation includes both random error and error due to individual controller differences. In fact, managers expected controller variability to be a major contributor to the MSE term.

Blocking in experimental design is similar to stratification in sampling.

One way to separate the effect of the individual differences is to use a randomized block design. Such a design will identify the variability stemming from individual controller differences and remove it from the MSE term. The randomized block design calls for a single sample of controllers. Each controller in the sample is tested with each of the three work station alternatives. In experimental design terminology, the work station is the *factor of interest* and the controllers are the *blocks.* The three treatments or populations associated with the work station factor correspond to the three work station alternatives. For simplicity, we refer to the work station alternatives as system A, system B, and system C.

The *randomized* aspect of the randomized block design is the random order in which the treatments (systems) are assigned to the controllers. If every controller were to test the three systems in the same order, any observed difference in systems might be due to the order of the test rather than to true differences in the systems.

To provide the necessary data, the three work station alternatives were installed at the Cleveland Control Center in Oberlin, Ohio. Six controllers were selected at random and assigned to operate each of the systems. A follow-up interview and a medical examination of each controller participating in the study provided a measure of the stress for each controller on each system. The data are reported in Table 13.5.

Table 13.6 is a summary of the stress data collected. In this table we include column totals (treatments) and row totals (blocks) as well as some sample means that will be helpful in

TABLE 13.5 A RANDOMIZED BLOCK DESIGN FOR THE AIR TRAFFIC CONTROLLER STRESS TEST

WEB file AirTraffic

		Treatments		
		System A	**System B**	**System C**
Blocks	**Controller 1**	15	15	18
	Controller 2	14	14	14
	Controller 3	10	11	15
	Controller 4	13	12	17
	Controller 5	16	13	16
	Controller 6	13	13	13

TABLE 13.6 SUMMARY OF STRESS DATA FOR THE AIR TRAFFIC CONTROLLER STRESS TEST

		Treatments			Row or Block Totals	Block Means
		System A	System B	System C		
Blocks	Controller 1	15	15	18	48	$\bar{x}_{1\cdot} = 48/3 = 16.0$
	Controller 2	14	14	14	42	$\bar{x}_{2\cdot} = 42/3 = 14.0$
	Controller 3	10	11	15	36	$\bar{x}_{3\cdot} = 36/3 = 12.0$
	Controller 4	13	12	17	42	$\bar{x}_{4\cdot} = 42/3 = 14.0$
	Controller 5	16	13	16	45	$\bar{x}_{5\cdot} = 45/3 = 15.0$
	Controller 6	13	13	13	39	$\bar{x}_{6\cdot} = 39/3 = 13.0$
Column or Treatment Totals		81	78	93	252	$\bar{\bar{x}} = \frac{252}{18} = 14.0$
Treatment Means		$\bar{x}_{\cdot 1} = \frac{81}{6} = 13.5$	$\bar{x}_{\cdot 2} = \frac{78}{6} = 13.0$	$\bar{x}_{\cdot 3} = \frac{93}{6} = 15.5$		

making the sum of squares computations for the ANOVA procedure. Because lower stress values are viewed as better, the sample data seem to favor system B with its mean stress rating of 13. However, the usual question remains: Do the sample results justify the conclusion that the population mean stress levels for the three systems differ? That is, are the differences statistically significant? An analysis of variance computation similar to the one performed for the completely randomized design can be used to answer this statistical question.

ANOVA Procedure

The ANOVA procedure for the randomized block design requires us to partition the sum of squares total (SST) into three groups: sum of squares due to treatments (SSTR), sum of squares due to blocks (SSBL), and sum of squares due to error (SSE). The formula for this partitioning follows.

$$\text{SST} = \text{SSTR} + \text{SSBL} + \text{SSE} \tag{13.21}$$

This sum of squares partition is summarized in the ANOVA table for the randomized block design as shown in Table 13.7. The notation used in the table is

$$k = \text{the number of treatments}$$
$$b = \text{the number of blocks}$$
$$n_T = \text{the total sample size } (n_T = kb)$$

Note that the ANOVA table also shows how the $n_T - 1$ total degrees of freedom are partitioned such that $k - 1$ degrees of freedom go to treatments, $b - 1$ go to blocks, and $(k - 1)(b - 1)$ go to the error term. The mean square column shows the sum of squares divided by the degrees of freedom, and F = MSTR/MSE is the F ratio used to test for a significant difference among the treatment means. The primary contribution of the randomized block design is that, by including blocks, we remove the individual controller differences from the MSE term and obtain a more powerful test for the stress differences in the three work station alternatives.

TABLE 13.7 ANOVA TABLE FOR THE RANDOMIZED BLOCK DESIGN WITH k TREATMENTS AND b BLOCKS

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

Computations and Conclusions

To compute the F statistic needed to test for a difference among treatment means with a randomized block design, we need to compute MSTR and MSE. To calculate these two mean squares, we must first compute SSTR and SSE; in doing so, we will also compute SSBL and SST. To simplify the presentation, we perform the calculations in four steps. In addition to k, b, and n_T as previously defined, the following notation is used.

x_{ij} = value of the observation corresponding to treatment j in block i
$\bar{x}_{\cdot j}$ = sample mean of the jth treatment
$\bar{x}_{i\cdot}$ = sample mean for the ith block
$\bar{\bar{x}}$ = overall sample mean

Step 1. Compute the total sum of squares (SST).

$$\text{SST} = \sum_{i=1}^{b}\sum_{j=1}^{k}(x_{ij} - \bar{\bar{x}})^2 \qquad \textbf{(13.22)}$$

Step 2. Compute the sum of squares due to treatments (SSTR).

$$\text{SSTR} = b\sum_{j=1}^{k}(\bar{x}_{\cdot j} - \bar{\bar{x}})^2 \qquad \textbf{(13.23)}$$

Step 3. Compute the sum of squares due to blocks (SSBL).

$$\text{SSBL} = k\sum_{i=1}^{b}(\bar{x}_{i\cdot} - \bar{\bar{x}})^2 \qquad \textbf{(13.24)}$$

Step 4. Compute the sum of squares due to error (SSE).

$$\text{SSE} = \text{SST} - \text{SSTR} - \text{SSBL} \qquad \textbf{(13.25)}$$

For the air traffic controller data in Table 13.6, these steps lead to the following sums of squares.

Step 1. $\text{SST} = (15 - 14)^2 + (15 - 14)^2 + (18 - 14)^2 + \cdots + (13 - 14)^2 = 70$

Step 2. $\text{SSTR} = 6[(13.5 - 14)^2 + (13.0 - 14)^2 + (15.5 - 14)^2] = 21$

Step 3. $\text{SSBL} = 3[(16 - 14)^2 + (14 - 14)^2 + (12 - 14)^2 + (14 - 14)^2 + (15 - 14)^2 + (13 - 14)^2] = 30$

Step 4. $\text{SSE} = 70 - 21 - 30 = 19$

TABLE 13.8 ANOVA TABLE FOR THE AIR TRAFFIC CONTROLLER STRESS TEST

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Treatments	21	2	10.5	10.5/1.9 = 5.53	.024
Blocks	30	5	6.0		
Error	19	10	1.9		
Total	70	17			

These sums of squares divided by their degrees of freedom provide the corresponding mean square values shown in Table 13.8.

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

$$F = \frac{\text{MSTR}}{\text{MSE}} = \frac{10.5}{1.9} = 5.53$$

The numerator degrees of freedom is $k - 1 = 3 - 1 = 2$ and the denominator degrees of freedom is $(k - 1)(b - 1) = (3 - 1)(6 - 1) = 10$. Because we will only reject the null hypothesis for large values of the test statistic, the p-value is the area under the F distribution to the right of $F = 5.53$. From Table 4 of Appendix B we find that with the degrees of freedom 2 and 10, $F = 5.53$ is between $F_{.025} = 5.46$ and $F_{.01} = 7.56$. As a result, the area in the upper tail, or the p-value, is between .01 and .025. Alternatively, we can use Excel or Minitab to show that the exact p-value for $F = 5.53$ is .024. With p-value $\leq \alpha = .05$, we reject the null hypothesis H_0: $\mu_1 = \mu_2 = \mu_3$ and conclude that the population mean stress levels differ for the three work station alternatives.

Some general comments can be made about the randomized block design. The experimental design described in this section is a *complete* block design; the word "complete" indicates that each block is subjected to all k treatments. That is, all controllers (blocks) were tested with all three systems (treatments). Experimental designs in which some but not all treatments are applied to each block are referred to as *incomplete* block designs. A discussion of incomplete block designs is beyond the scope of this text.

Because each controller in the air traffic controller stress test was required to use all three systems, this approach guarantees a complete block design. In some cases, however, blocking is carried out with "similar" experimental units in each block. For example, assume that in a pretest of air traffic controllers, the population of controllers was divided into groups ranging from extremely high-stress individuals to extremely low-stress individuals. The blocking could still be accomplished by having three controllers from each of the stress classifications participate in the study. Each block would then consist of three controllers in the same stress group. The randomized aspect of the block design would be the random assignment of the three controllers in each block to the three systems.

Finally, note that the ANOVA table shown in Table 13.7 provides an F value to test for treatment effects but *not* for blocks. The reason is that the experiment was designed to test a single factor—work station design. The blocking based on individual stress differences was conducted to remove such variation from the MSE term. However, the study was not designed to test specifically for individual differences in stress.

Some analysts compute F = MSB/MSE and use that statistic to test for significance of the blocks. Then they use the result as a guide to whether the same type of blocking would be desired in future experiments. However, if individual stress difference is to be a factor in the study, a different experimental design should be used. A test of significance on blocks should not be performed as a basis for a conclusion about a second factor.

NOTES AND COMMENTS

The error degrees of freedom are less for a randomized block design than for a completely randomized design because $b - 1$ degrees of freedom are lost for the b blocks. If n is small, the potential effects due to blocks can be masked because of the loss of error degrees of freedom; for large n, the effects are minimized.

Exercises

Methods

SELF test

21. Consider the experimental results for the following randomized block design. Make the calculations necessary to set up the analysis of variance table.

		Treatments		
		A	**B**	**C**
Blocks	**1**	10	9	8
	2	12	6	5
	3	18	15	14
	4	20	18	18
	5	8	7	8

Use $\alpha = .05$ to test for any significant differences.

22. The following data were obtained for a randomized block design involving five treatments and three blocks: SST = 430, SSTR = 310, SSBL = 85. Set up the ANOVA table and test for any significant differences. Use $\alpha = .05$.

23. An experiment has been conducted for four treatments with eight blocks. Complete the following analysis of variance table.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Treatments	900			
Blocks	400			
Error				
Total	1800			

Use $\alpha = .05$ to test for any significant differences.

Applications

24. An automobile dealer conducted a test to determine if the time in minutes needed to complete a minor engine tune-up depends on whether a computerized engine analyzer or an electronic analyzer is used. Because tune-up time varies among compact, intermediate, and full-sized cars, the three types of cars were used as blocks in the experiment. The data obtained follow.

		Analyzer	
		Computerized	Electronic
Car	Compact	50	42
	Intermediate	55	44
	Full-sized	63	46

Use $\alpha = .05$ to test for any significant differences.

25. Prices for vitamins and other health supplements increased over the past several years, and the prices charged by different retail outlets often vary a great deal. The following data show the prices for 13 products at four retail outlets in Rochester, New York (*Democrat and Chronicle,* February 13, 2005).

WEB file
Vitamins

Item	CVS	Kmart	Rite-Aid	Wegmans
Caltrate +D (600 mg/60 tablets)	8.49	5.99	7.99	5.99
Centrum (130 tablets)	9.49	9.47	9.89	7.97
Cod liver oil (100 gel tablets)	2.66	2.59	1.99	2.69
Fish oil (1,000 mg/60 tablets)	6.19	4.99	4.99	5.99
Flintstones Children's (60 tablets)	7.69	5.99	5.99	6.29
Folic acid (400 mcg/250 tablets)	2.19	2.49	3.74	2.69
One-a-Day Maximum (100 tablets)	8.99	7.49	6.99	6.99
One-a-Day Scooby (50 tablets)	7.49	5.99	6.49	5.47
Poly-Vi-Sol (drops, 50 ml)	9.99	8.49	9.99	8.37
Vitamin B-12 (100 mcg/100 tablets)	3.59	1.99	1.99	1.79
Vitamin C (500 mg/100 tablets)	2.99	2.49	1.99	2.39
Vitamin E (200 IU/100 tablets)	4.69	3.49	2.99	3.29
Zinc (50 mg/100 tablets)	2.66	2.59	3.99	2.79

Use $\alpha = .05$ to test for any significant difference in the mean price for the four retail outlets.

26. The Scholastic Aptitude Test (SAT) contains three parts: critical reading, mathematics, and writing. Each part is scored on an 800-point scale. Information on test scores for the 2009 version of the SAT is available at the College Board website. A sample of SAT scores for six students follows.

WEB file
SATScores

Student	Critical Reading	Mathematics	Writing
1	526	534	530
2	594	590	586
3	465	464	445
4	561	566	553
5	436	478	430
6	430	458	420

a. Using a .05 level of significance, do students perform differently on the three portions of the SAT?
b. Which portion of the test seems to give the students the most trouble? Explain.

27. A study reported in the *Journal of the American Medical Association* investigated the cardiac demands of heavy snow shoveling. Ten healthy men underwent exercise testing with a treadmill and a cycle ergometer modified for arm cranking. The men then cleared two tracts of heavy, wet snow by using a lightweight plastic snow shovel and an electric snow thrower. Each subject's heart rate, blood pressure, oxygen uptake, and perceived exertion during snow removal were compared with the values obtained during treadmill

and arm-crank ergometer testing. Suppose the following table gives the heart rates in beats per minute for each of the 10 subjects.

WEB file
SnowShoveling

Subject	Treadmill	Arm-Crank Ergometer	Snow Shovel	Snow Thrower
1	177	205	180	98
2	151	177	164	120
3	184	166	167	111
4	161	152	173	122
5	192	142	179	151
6	193	172	205	158
7	164	191	156	117
8	207	170	160	123
9	177	181	175	127
10	174	154	191	109

At the .05 level of significance, test for any significant differences.

13.5 Factorial Experiment

The experimental designs we have considered thus far enable us to draw statistical conclusions about one factor. However, in some experiments we want to draw conclusions about more than one variable or factor. A **factorial experiment** is an experimental design that allows simultaneous conclusions about two or more factors. The term *factorial* is used because the experimental conditions include all possible combinations of the factors. For example, for a levels of factor A and b levels of factor B, the experiment will involve collecting data on ab treatment combinations. In this section we will show the analysis for a two-factor factorial experiment. The basic approach can be extended to experiments involving more than two factors.

As an illustration of a two-factor factorial experiment, we will consider a study involving the Graduate Management Admissions Test (GMAT), a standardized test used by graduate schools of business to evaluate an applicant's ability to pursue a graduate program in that field. Scores on the GMAT range from 200 to 800, with higher scores implying higher aptitude.

In an attempt to improve students' performance on the GMAT, a major Texas university is considering offering the following three GMAT preparation programs.

1. A three-hour review session covering the types of questions generally asked on the GMAT.
2. A one-day program covering relevant exam material, along with the taking and grading of a sample exam.
3. An intensive 10-week course involving the identification of each student's weaknesses and the setting up of individualized programs for improvement.

Hence, one factor in this study is the GMAT preparation program, which has three treatments: three-hour review, one-day program, and 10-week course. Before selecting the preparation program to adopt, further study will be conducted to determine how the proposed programs affect GMAT scores.

The GMAT is usually taken by students from three colleges: the College of Business, the College of Engineering, and the College of Arts and Sciences. Therefore, a second factor of interest in the experiment is whether a student's undergraduate college affects the GMAT score. This second factor, undergraduate college, also has three treatments: business, engineering, and arts and sciences. The factorial design for this experiment with three treatments corresponding to factor A, the preparation program, and three treatments corresponding to

TABLE 13.9 NINE TREATMENT COMBINATIONS FOR THE TWO-FACTOR GMAT EXPERIMENT

		Factor B: College		
		Business	**Engineering**	**Arts and Sciences**
Factor A: Preparation Program	**Three-hour review**	1	2	3
	One-day program	4	5	6
	10-week course	7	8	9

factor B, the undergraduate college, will have a total of $3 \times 3 = 9$ treatment combinations. These treatment combinations or experimental conditions are summarized in Table 13.9.

Assume that a sample of two students will be selected corresponding to each of the nine treatment combinations shown in Table 13.9: two business students will take the three-hour review, two will take the one-day program, and two will take the 10-week course. In addition, two engineering students and two arts and sciences students will take each of the three preparation programs. In experimental design terminology, the sample size of two for each treatment combination indicates that we have two **replications**. Additional replications and a larger sample size could easily be used, but we elect to minimize the computational aspects for this illustration.

This experimental design requires that six students who plan to attend graduate school be randomly selected from *each* of the three undergraduate colleges. Then two students from each college should be assigned randomly to each preparation program, resulting in a total of 18 students being used in the study.

Let us assume that the randomly selected students participated in the preparation programs and then took the GMAT. The scores obtained are reported in Table 13.10.

The analysis of variance computations with the data in Table 13.10 will provide answers to the following questions.

- **Main effect (factor A):** Do the preparation programs differ in terms of effect on GMAT scores?
- **Main effect (factor B):** Do the undergraduate colleges differ in terms of effect on GMAT scores?
- **Interaction effect (factors A and B):** Do students in some colleges do better on one type of preparation program whereas others do better on a different type of preparation program?

The term **interaction** refers to a new effect that we can now study because we used a factorial experiment. If the interaction effect has a significant impact on the GMAT scores,

TABLE 13.10 GMAT SCORES FOR THE TWO-FACTOR EXPERIMENT

WEB file
GMATStudy

		Factor B: College		
		Business	**Engineering**	**Arts and Sciences**
Factor A: Preparation Program	**Three-hour review**	500 580	540 460	480 400
	One-day program	460 540	560 620	420 480
	10-week course	560 600	600 580	480 410

TABLE 13.11 ANOVA TABLE FOR THE TWO-FACTOR FACTORIAL EXPERIMENT WITH r REPLICATIONS

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Factor A	SSA	$a - 1$	$\text{MSA} = \dfrac{\text{SSA}}{a - 1}$	$\dfrac{\text{MSA}}{\text{MSE}}$	
Factor B	SSB	$b - 1$	$\text{MSB} = \dfrac{\text{SSB}}{b - 1}$	$\dfrac{\text{MSB}}{\text{MSE}}$	
Interaction	SSAB	$(a - 1)(b - 1)$	$\text{MSAB} = \dfrac{\text{SSAB}}{(a - 1)(b - 1)}$	$\dfrac{\text{MSAB}}{\text{MSE}}$	
Error	SSE	$ab(r - 1)$	$\text{MSE} = \dfrac{\text{SSE}}{ab(r - 1)}$		
Total	SST	$n_T - 1$			

we can conclude that the effect of the type of preparation program depends on the undergraduate college.

ANOVA Procedure

The ANOVA procedure for the two-factor factorial experiment requires us to partition the sum of squares total (SST) into four groups: sum of squares for factor A (SSA), sum of squares for factor B (SSB), sum of squares for interaction (SSAB), and sum of squares due to error (SSE). The formula for this partitioning follows.

$$\text{SST} = \text{SSA} + \text{SSB} + \text{SSAB} + \text{SSE} \tag{13.26}$$

The partitioning of the sum of squares and degrees of freedom is summarized in Table 13.11. The following notation is used.

a = number of levels of factor A
b = number of levels of factor B
r = number of replications
n_T = total number of observations taken in the experiment; $n_T = abr$

Computations and Conclusions

To compute the F statistics needed to test for the significance of factor A, factor B, and interaction, we need to compute MSA, MSB, MSAB, and MSE. To calculate these four mean squares, we must first compute SSA, SSB, SSAB, and SSE; in doing so we will also compute SST. To simplify the presentation, we perform the calculations in five steps. In addition to a, b, r, and n_T as previously defined, the following notation is used.

x_{ijk} = observation corresponding to the kth replicate taken from treatment i of factor A and treatment j of factor B
$\bar{x}_{i.}$ = sample mean for the observations in treatment i (factor A)
$\bar{x}_{.j}$ = sample mean for the observations in treatment j (factor B)
$\bar{x}_{ij}$ = sample mean for the observations corresponding to the combination of treatment i (factor A) and treatment j (factor B)
$\bar{\bar{x}}$ = overall sample mean of all n_T observations

Step 1. Compute the total sum of squares.

$$\text{SST} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(x_{ijk} - \bar{\bar{x}})^2 \qquad \textbf{(13.27)}$$

Step 2. Compute the sum of squares for factor A.

$$\text{SSA} = br\sum_{i=1}^{a}(\bar{x}_{i\cdot} - \bar{\bar{x}})^2 \qquad \textbf{(13.28)}$$

Step 3. Compute the sum of squares for factor B.

$$\text{SSB} = ar\sum_{j=1}^{b}(\bar{x}_{\cdot j} - \bar{\bar{x}})^2 \qquad \textbf{(13.29)}$$

Step 4. Compute the sum of squares for interaction.

$$\text{SSAB} = r\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{\bar{x}})^2 \qquad \textbf{(13.30)}$$

Step 5. Compute the sum of squares due to error.

$$\text{SSE} = \text{SST} - \text{SSA} - \text{SSB} - \text{SSAB} \qquad \textbf{(13.31)}$$

Table 13.12 reports the data collected in the experiment and the various sums that will help us with the sum of squares computations. Using equations (13.27) through (13.31), we calculate the following sums of squares for the GMAT two-factor factorial experiment.

Step 1. $\text{SST} = (500 - 515)^2 + (580 - 515)^2 + (540 - 515)^2 + \cdots + (410 - 515)^2 = 82{,}450$

Step 2. $\text{SSA} = (3)(2)[(493.33 - 515)^2 + (513.33 - 515)^2 + (538.33 - 515)^2] = 6100$

Step 3. $\text{SSB} = (3)(2)[(540 - 515)^2 + (560 - 515)^2 + (445 - 515)^2] = 45{,}300$

Step 4. $\text{SSAB} = 2[(540 - 493.33 - 540 + 515)^2 + (500 - 493.33 - 560 + 515)^2 + \cdots + (445 - 538.33 - 445 + 515)^2] = 11{,}200$

Step 5. $\text{SSE} = 82{,}450 - 6100 - 45{,}300 - 11{,}200 = 19{,}850$

These sums of squares divided by their corresponding degrees of freedom provide the appropriate mean square values for testing the two main effects (preparation program and undergraduate college) and the interaction effect.

Because of the computational effort involved in any modest- to large-size factorial experiment, the computer usually plays an important role in performing the analysis of variance computations shown above and in the calculation of the p-values used to make the hypothesis testing decisions. Figure 13.6 shows the Minitab output for the analysis of variance for the GMAT two-factor factorial experiment. Let us use the Minitab output and a level of significance $\alpha = .05$ to conduct the hypothesis tests for the two-factor GMAT study. The p-value used to test for significant differences among the three preparation programs (factor A) is .299. Because the p-value $= .299$ is greater than $\alpha = .05$, there is no significant difference in the mean GMAT test scores for the three preparation programs. However, for the undergraduate college effect, the p-value $= .005$ is less than $\alpha = .05$; thus, there is a significant difference in the mean GMAT test scores among the three undergraduate colleges.

TABLE 13.12 GMAT SUMMARY DATA FOR THE TWO-FACTOR EXPERIMENT

	Treatment combination totals	Factor B: College			Row Totals	Factor A Means
		Business	**Engineering**	**Arts and Sciences**		
Factor A: Preparation Program	**Three-hour review**	500 580 1080 $\bar{x}_{11} = \frac{1080}{2} = 540$	540 460 1000 $\bar{x}_{12} = \frac{1000}{2} = 500$	480 400 880 $\bar{x}_{13} = \frac{880}{2} = 440$	2960	$\bar{x}_{1\cdot} = \frac{2960}{6} = 493.33$
	One-day program	460 540 1000 $\bar{x}_{21} = \frac{1000}{2} = 500$	560 620 1180 $\bar{x}_{22} = \frac{1180}{2} = 590$	420 480 900 $\bar{x}_{23} = \frac{900}{2} = 450$	3080	$\bar{x}_{2\cdot} = \frac{3080}{6} = 513.33$
	10-week course	560 600 1160 $\bar{x}_{31} = \frac{1160}{2} = 580$	600 580 1180 $\bar{x}_{32} = \frac{1180}{2} = 590$	480 410 890 $\bar{x}_{33} = \frac{890}{2} = 445$	3230	$\bar{x}_{3\cdot} = \frac{3230}{6} = 538.33$
	Column Totals	3240	3360	2670	9270 ← Overall total	
	Factor B Means	$\bar{x}_{\cdot 1} = \frac{3240}{6} = 540$	$\bar{x}_{\cdot 2} = \frac{3360}{6} = 560$	$\bar{x}_{\cdot 3} = \frac{2670}{6} = 445$	$\bar{\bar{x}} = \frac{9270}{18} = 515$	

FIGURE 13.6 MINITAB OUTPUT FOR THE GMAT TWO-FACTOR DESIGN

```
SOURCE          DF        SS        MS         F         P
Factor A         2      6100      3050      1.38     0.299
Factor B         2     45300     22650     10.27     0.005
Interaction      4     11200      2800      1.27     0.350
Error            9     19850      2206
Total           17     82450
```

Finally, because the p-value of .350 for the interaction effect is greater than $\alpha = .05$, there is no significant interaction effect. Therefore, the study provides no reason to believe that the three preparation programs differ in their ability to prepare students from the different colleges for the GMAT.

Undergraduate college was found to be a significant factor. Checking the calculations in Table 13.12, we see that the sample means are: business students $\bar{x}_{\cdot 1} = 540$, engineering students $\bar{x}_{\cdot 2} = 560$, and arts and sciences students $\bar{x}_{\cdot 3} = 445$. Tests on individual treatment means can be conducted; yet after reviewing the three sample means, we would anticipate no difference in preparation for business and engineering graduates. However, the arts and sciences students appear to be significantly less prepared for the GMAT than students in the other colleges. Perhaps this observation will lead the university to consider other options for assisting these students in preparing for the Graduate Management Admission Test.

Exercises

Methods

SELF test

28. A factorial experiment involving two levels of factor A and three levels of factor B resulted in the following data.

		Factor B		
		Level 1	Level 2	Level 3
Factor A	Level 1	135 165	90 66	75 93
	Level 2	125 95	127 105	120 136

Test for any significant main effects and any interaction. Use $\alpha = .05$.

29. The calculations for a factorial experiment involving four levels of factor A, three levels of factor B, and three replications resulted in the following data: SST = 280, SSA = 26, SSB = 23, SSAB = 175. Set up the ANOVA table and test for any significant main effects and any interaction effect. Use $\alpha = .05$.

Applications

30. A mail-order catalog firm designed a factorial experiment to test the effect of the size of a magazine advertisement and the advertisement design on the number of catalog requests received (data in thousands). Three advertising designs and two different size advertisements were considered. The data obtained follow. Use the ANOVA procedure for

factorial designs to test for any significant effects due to type of design, size of advertisement, or interaction. Use $\alpha = .05$.

		Size of Advertisement	
		Small	**Large**
	A	8	12
		12	8
Design	**B**	22	26
		14	30
	C	10	18
		18	14

31. An amusement park studied methods for decreasing the waiting time (minutes) for rides by loading and unloading riders more efficiently. Two alternative loading/unloading methods have been proposed. To account for potential differences due to the type of ride and the possible interaction between the method of loading and unloading and the type of ride, a factorial experiment was designed. Use the following data to test for any significant effect due to the loading and unloading method, the type of ride, and interaction. Use $\alpha = .05$.

	Type of Ride		
	Roller Coaster	**Screaming Demon**	**Log Flume**
Method 1	41	52	50
	43	44	46
Method 2	49	50	48
	51	46	44

32. As part of a study designed to compare hybrid and similarly equipped conventional vehicles, *Consumer Reports* tested a variety of classes of hybrid and all-gas model cars and sport utility vehicles (SUVs). The following data show the miles-per-gallon rating *Consumer Reports* obtained for two hybrid small cars, two hybrid midsize cars, two hybrid small SUVs, and two hybrid midsized SUVs; also shown are the miles per gallon obtained for eight similarly equipped conventional models (*Consumer Reports,* October 2008).

WEB file
HybridTest

Make/Model	Class	Type	MPG
Honda Civic	Small Car	Hybrid	37
Honda Civic	Small Car	Conventional	28
Toyota Prius	Small Car	Hybrid	44
Toyota Corolla	Small Car	Conventional	32
Chevrolet Malibu	Midsize Car	Hybrid	27
Chevrolet Malibu	Midsize Car	Conventional	23
Nissan Altima	Midsize Car	Hybrid	32
Nissan Altima	Midsize Car	Conventional	25
Ford Escape	Small SUV	Hybrid	27
Ford Escape	Small SUV	Conventional	21
Saturn Vue	Small SUV	Hybrid	28
Saturn Vue	Small SUV	Conventional	22
Lexus RX	Midsize SUV	Hybrid	23
Lexus RX	Midsize SUV	Conventional	19
Toyota Highlander	Midsize SUV	Hybrid	24
Toyota Highlander	Midsize SUV	Conventional	18

At the $\alpha = .05$ level of significance, test for significant effects due to class, type, and interaction.

33. A study reported in *The Accounting Review* examined the separate and joint effects of two levels of time pressure (low and moderate) and three levels of knowledge (naive, declarative, and procedural) on key word selection behavior in tax research. Subjects were given a tax case containing a set of facts, a tax issue, and a key word index consisting of 1336 key words. They were asked to select the key words they believed would refer them to a tax authority relevant to resolving the tax case. Prior to the experiment, a group of tax experts determined that the text contained 19 relevant key words. Subjects in the naive group had little or no declarative or procedural knowledge, subjects in the declarative group had significant declarative knowledge but little or no procedural knowledge, and subjects in the procedural group had significant declarative knowledge and procedural knowledge. Declarative knowledge consists of knowledge of both the applicable tax rules and the technical terms used to describe such rules. Procedural knowledge is knowledge of the rules that guide the tax researcher's search for relevant key words. Subjects in the low time pressure situation were told they had 25 minutes to complete the problem, an amount of time which should be "more than adequate" to complete the case; subjects in the moderate time pressure situation were told they would have "only" 11 minutes to complete the case. Suppose 25 subjects were selected for each of the six treatment combinations and the sample means for each treatment combination are as follows (standard deviations are in parentheses).

		Knowledge		
		Naive	**Declarative**	**Procedural**
Time Pressure	**Low**	1.13 (1.12)	1.56 (1.33)	2.00 (1.54)
	Moderate	0.48 (0.80)	1.68 (1.36)	2.86 (1.80)

Use the ANOVA procedure to test for any significant differences due to time pressure, knowledge, and interaction. Use a .05 level of significance. Assume that the total sum of squares for this experiment is 327.50.

Summary

In this chapter we showed how analysis of variance can be used to test for differences among means of several populations or treatments. We introduced the completely randomized design, the randomized block design, and the two-factor factorial experiment. The completely randomized design and the randomized block design are used to draw conclusions about differences in the means of a single factor. The primary purpose of blocking in the randomized block design is to remove extraneous sources of variation from the error term. Such blocking provides a better estimate of the true error variance and a better test to determine whether the population or treatment means of the factor differ significantly.

We showed that the basis for the statistical tests used in analysis of variance and experimental design is the development of two independent estimates of the population variance σ^2. In the single-factor case, one estimator is based on the variation between the treatments; this estimator provides an unbiased estimate of σ^2 only if the means $\mu_1, \mu_2, \ldots, \mu_k$ are all equal. A second estimator of σ^2 is based on the variation of the observations within each sample; this estimator will always provide an unbiased estimate of σ^2. By computing the ratio of these two estimators (the F statistic) we developed a rejection rule for determining whether to reject the null hypothesis that the population or treatment means are equal. In all the experimental designs considered, the partitioning of the sum of squares and degrees of freedom into their various sources enabled us to compute the appropriate values for the analysis of variance calculations and tests. We also showed how Fisher's LSD procedure and the Bonferroni adjustment can be used to perform pairwise comparisons to determine which means are different.

Glossary

Factor Another word for the independent variable of interest.
Treatments Different levels of a factor.
Single-factor experiment An experiment involving only one factor with k populations or treatments.
Response variable Another word for the dependent variable of interest.
Experimental units The objects of interest in the experiment.
ANOVA table A table used to summarize the analysis of variance computations and results. It contains columns showing the source of variation, the sum of squares, the degrees of freedom, the mean square, and the F value(s).
Partitioning The process of allocating the total sum of squares and degrees of freedom to the various components.
Multiple comparison procedures Statistical procedures that can be used to conduct statistical comparisons between pairs of population means.
Comparisonwise Type I error rate The probability of a Type I error associated with a single pairwise comparison.
Experimentwise Type I error rate The probability of making a Type I error on at least one of several pairwise comparisons.
Completely randomized design An experimental design in which the treatments are randomly assigned to the experimental units.
Blocking The process of using the same or similar experimental units for all treatments. The purpose of blocking is to remove a source of variation from the error term and hence provide a more powerful test for a difference in population or treatment means.
Randomized block design An experimental design employing blocking.
Factorial experiment An experimental design that allows simultaneous conclusions about two or more factors.
Replications The number of times each experimental condition is repeated in an experiment.
Interaction The effect produced when the levels of one factor interact with the levels of another factor in influencing the response variable.

Key Formulas

Completely Randomized Design

Sample Mean for Treatment j

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

Sample Variance for Treatment j

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

Overall Sample Mean

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

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

Mean Square Due to Treatments

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

Sum of Squares Due to Treatments

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

Mean Square Due to Error

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

Sum of Squares Due to Error

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

Test Statistic for the Equality of *k* Population Means

$$F = \frac{\text{MSTR}}{\text{MSE}} \tag{13.12}$$

Total Sum of Squares

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

Partitioning of Sum of Squares

$$\text{SST} = \text{SSTR} + \text{SSE} \tag{13.14}$$

Multiple Comparison Procedures

Test Statistic for Fisher's LSD Procedure

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

Fisher's LSD

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

Randomized Block Design

Total Sum of Squares

$$\text{SST} = \sum_{i=1}^{b}\sum_{j=1}^{k}(x_{ij} - \bar{\bar{x}})^2 \tag{13.22}$$

Sum of Squares Due to Treatments

$$\text{SSTR} = b\sum_{j=1}^{k}(\bar{x}_{\cdot j} - \bar{\bar{x}})^2 \tag{13.23}$$

Sum of Squares Due to Blocks

$$\text{SSBL} = k\sum_{i=1}^{b}(\bar{x}_{i\cdot} - \bar{\bar{x}})^2 \tag{13.24}$$

Sum of Squares Due to Error

$$\text{SSE} = \text{SST} - \text{SSTR} - \text{SSBL} \tag{13.25}$$

Factorial Experiment

Total Sum of Squares

$$\text{SST} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(x_{ijk} - \bar{\bar{x}})^2 \tag{13.27}$$

Sum of Squares for Factor A

$$\text{SSA} = br\sum_{i=1}^{a}(\bar{x}_{i\cdot} - \bar{\bar{x}})^2 \tag{13.28}$$

Sum of Squares for Factor B

$$\text{SSB} = ar\sum_{j=1}^{b}(\bar{x}_{\cdot j} - \bar{\bar{x}})^2 \tag{13.29}$$

Sum of Squares for Interaction

$$\text{SSAB} = r\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij} - \bar{x}_{i\cdot} - \bar{x}_{\cdot j} + \bar{\bar{x}})^2 \tag{13.30}$$

Sum of Squares for Error

$$\text{SSE} = \text{SST} - \text{SSA} - \text{SSB} - \text{SSAB} \tag{13.31}$$

Supplementary Exercises

34. In a completely randomized experimental design, three brands of paper towels were tested for their ability to absorb water. Equal-size towels were used, with four sections of towels tested per brand. The absorbency rating data follow. At a .05 level of significance, does there appear to be a difference in the ability of the brands to absorb water?

Brand		
x	*y*	*z*
91	99	83
100	96	88
88	94	89
89	99	76

35. A study reported in the *Journal of Small Business Management* concluded that self-employed individuals do not experience higher job satisfaction than individuals who are not self-employed. In this study, job satisfaction is measured using 18 items, each of which is rated using a Likert-type scale with 1–5 response options ranging from strong agreement to strong disagreement. A higher score on this scale indicates a higher degree of job satisfaction. The sum of the ratings for the 18 items, ranging from 18–90, is used as the measure of job satisfaction. Suppose that this approach was used to measure the job satisfaction for lawyers, physical therapists, cabinetmakers, and systems analysts. The results obtained for a sample of 10 individuals from each profession follow.

WEB file

SatisJob

Lawyer	Physical Therapist	Cabinetmaker	Systems Analyst
44	55	54	44
42	78	65	73
74	80	79	71
42	86	69	60
53	60	79	64
50	59	64	66
45	62	59	41
48	52	78	55
64	55	84	76
38	50	60	62

At the $\alpha = .05$ level of significance, test for any difference in the job satisfaction among the four professions.

36. *Money* magazine reports percentage returns and expense ratios for stock and bond funds. The following data are the expense ratios for 10 midcap stock funds, 10 small-cap stock funds, 10 hybrid stock funds, and 10 specialty stock funds (*Money,* March 2003).

WEB file

Funds

Midcap	Small-Cap	Hybrid	Specialty
1.2	2.0	2.0	1.6
1.1	1.2	2.7	2.7
1.0	1.7	1.8	2.6
1.2	1.8	1.5	2.5
1.3	1.5	2.5	1.9
1.8	2.3	1.0	1.5
1.4	1.9	0.9	1.6
1.4	1.3	1.9	2.7
1.0	1.2	1.4	2.2
1.4	1.3	0.3	0.7

Use $\alpha = .05$ to test for any significant difference in the mean expense ratio among the four types of stock funds.

37. The U.S. Census Bureau computes quarterly vacancy and homeownership rates by state and metropolitan statistical area. Each metropolitan statistical area (MSA) has at least one urbanized area of 50,000 or more inhabitants. The following data are the rental vacancy rates (%) for MSAs in four geographic regions of the United States for the first quarter of 2008 (U.S. Census Bureau website, January 2009).

WEB file
RentalVacancy

Midwest	Northeast	South	West
16.2	2.7	16.6	7.9
10.1	11.5	8.5	6.6
8.6	6.6	12.1	6.9
12.3	7.9	9.8	5.6
10.0	5.3	9.3	4.3
16.9	10.7	9.1	15.2
16.9	8.6	5.6	5.7
5.4	5.5	9.4	4.0
18.1	12.7	11.6	12.3
11.9	8.3	15.6	3.6
11.0	6.7	18.3	11.0
9.6	14.2	13.4	12.1
7.6	1.7	6.5	8.7
12.9	3.6	11.4	5.0
12.2	11.5	13.1	4.7
13.6	16.3	4.4	3.3
		8.2	3.4
		24.0	5.5
		12.2	
		22.6	
		12.0	
		14.5	
		12.6	
		9.5	
		10.1	

Use $\alpha = .05$ to test whether there the mean vacancy rate is the same for each geographic region.

38. Three different assembly methods have been proposed for a new product. A completely randomized experimental design was chosen to determine which assembly method results in the greatest number of parts produced per hour, and 30 workers were randomly selected and assigned to use one of the proposed methods. The number of units produced by each worker follows.

WEB file
Assembly

Method		
A	B	C
97	93	99
73	100	94
93	93	87
100	55	66
73	77	59
91	91	75
100	85	84
86	73	72
92	90	88
95	83	86

Use these data and test to see whether the mean number of parts produced is the same with each method. Use $\alpha = .05$.

39. In a study conducted to investigate browsing activity by shoppers, each shopper was initially classified as a nonbrowser, light browser, or heavy browser. For each shopper, the study obtained a measure to determine how comfortable the shopper was in a store. Higher scores indicated greater comfort. Suppose the following data were collected.

WEB file
Browsing

Nonbrowser	Light Browser	Heavy Browser
4	5	5
5	6	7
6	5	5
3	4	7
3	7	4
4	4	6
5	6	5
4	5	7

a. Use $\alpha = .05$ to test for differences among comfort levels for the three types of browsers.
b. Use Fisher's LSD procedure to compare the comfort levels of nonbrowsers and light browsers. Use $\alpha = .05$. What is your conclusion?

40. A research firm tests the miles-per-gallon characteristics of three brands of gasoline. Because of different gasoline performance characteristics in different brands of automobiles, five brands of automobiles are selected and treated as blocks in the experiment; that is, each brand of automobile is tested with each type of gasoline. The results of the experiment (in miles per gallon) follow.

		Gasoline Brands		
		I	**II**	**III**
	A	18	21	20
	B	24	26	27
Automobiles	**C**	30	29	34
	D	22	25	24
	E	20	23	24

a. At $\alpha = .05$, is there a significant difference in the mean miles-per-gallon characteristics of the three brands of gasoline?
b. Analyze the experimental data using the ANOVA procedure for completely randomized designs. Compare your findings with those obtained in part (a). What is the advantage of attempting to remove the block effect?

41. Wegmans Food Markets and Tops Friendly Markets are the major grocery chains in the Rochester, New York, area. When Wal-Mart opened a Supercenter in one of the Rochester suburbs, experts predicted that Wal-Mart would undersell both local stores. The *Democrat and Chronicle* obtained the price data in the following table for a 15-item market basket (*Democrat and Chronicle,* March 17, 2002).

WEB file
MarketBasket

Item	Tops	Wal-Mart	Wegmans
Bananas (1 lb.)	0.49	0.48	0.49
Campbell's soup (10.75 oz.)	0.60	0.54	0.77
Chicken breasts (3 lbs.)	10.47	8.61	8.07
Colgate toothpaste (6.2 oz.)	1.99	2.40	1.97
Large eggs (1 dozen)	1.59	0.88	0.79
Heinz ketchup (36 oz.)	2.59	1.78	2.59
Jell-O (cherry, 3 oz.)	0.67	0.42	0.65
Jif peanut butter (18 oz.)	2.29	1.78	2.09
Milk (fat free, 1/2 gal.)	1.34	1.24	1.34
Oscar Meyer hotdogs (1 lb.)	3.29	1.50	3.39
Ragu pasta sauce (1 lb., 10 oz.)	2.09	1.50	1.25
Ritz crackers (1 lb.)	3.29	2.00	3.39
Tide detergent (liquid, 100 oz.)	6.79	5.24	5.99
Tropicana orange juice (1/2 gal.)	2.50	2.50	2.50
Twizzlers (strawberry, 1 lb.)	1.19	1.27	1.69

At the .05 level of significance, test for any significant difference in the mean price for the 15-item shopping basket for the three stores.

42. The U.S. Department of Housing and Urban Development provides data that show the fair market monthly rent for metropolitan areas. The following data show the fair market monthly rent ($) in 2005 for 1-bedroom, 2-bedroom, and 3-bedroom apartments for five metropolitan areas (*The New York Times Almanac,* 2006).

	Boston	Miami	San Diego	San Jose	Washington
1 Bedroom	1077	775	975	1107	1045
2 Bedrooms	1266	929	1183	1313	1187
3 Bedrooms	1513	1204	1725	1889	1537

At the .05 level of significance, test whether the mean fair market monthly rent is the same for each metropolitan area.

43. A factorial experiment was designed to test for any significant differences in the time needed to perform English to foreign language translations with two computerized language translators. Because the type of language translated was also considered a significant factor, translations were made with both systems for three different languages: Spanish, French, and German. Use the following data for translation time in hours.

	Language		
	Spanish	**French**	**German**
System 1	8	10	12
	12	14	16
System 2	6	14	16
	10	16	22

Test for any significant differences due to language translator, type of language, and interaction. Use $\alpha = .05$.

44. A manufacturing company designed a factorial experiment to determine whether the number of defective parts produced by two machines differed and if the number of defective parts produced also depended on whether the raw material needed by each machine was

loaded manually or by an automatic feed system. The following data give the numbers of defective parts produced. Use $\alpha = .05$ to test for any significant effect due to machine, loading system, and interaction.

	Loading System	
	Manual	**Automatic**
Machine 1	30	30
	34	26
Machine 2	20	24
	22	28

Case Problem 1 Wentworth Medical Center

As part of a long-term study of individuals 65 years of age or older, sociologists and physicians at the Wentworth Medical Center in upstate New York investigated the relationship between geographic location and depression. A sample of 60 individuals, all in reasonably good health, was selected; 20 individuals were residents of Florida, 20 were residents of New York, and 20 were residents of North Carolina. Each of the individuals sampled was given a standardized test to measure depression. The data collected follow; higher test scores indicate higher levels of depression. These data are contained in the file Medical1.

A second part of the study considered the relationship between geographic location and depression for individuals 65 years of age or older who had a chronic health condition such as arthritis, hypertension, and/or heart ailment. A sample of 60 individuals with such conditions was identified. Again, 20 were residents of Florida, 20 were residents of New York, and 20 were residents of North Carolina. The levels of depression recorded for this study follow. These data are contained in the file named Medical2.

WEB file Medical1

WEB file Medical2

Data from Medical1			Data from Medical2		
Florida	**New York**	**North Carolina**	**Florida**	**New York**	**North Carolina**
3	8	10	13	14	10
7	11	7	12	9	12
7	9	3	17	15	15
3	7	5	17	12	18
8	8	11	20	16	12
8	7	8	21	24	14
8	8	4	16	18	17
5	4	3	14	14	8
5	13	7	13	15	14
2	10	8	17	17	16
6	6	8	12	20	18
2	8	7	9	11	17
6	12	3	12	23	19
6	8	9	15	19	15
9	6	8	16	17	13
7	8	12	15	14	14
5	5	6	13	9	11
4	7	3	10	14	12
7	7	8	11	13	13
3	8	11	17	11	11

Managerial Report

1. Use descriptive statistics to summarize the data from the two studies. What are your preliminary observations about the depression scores?
2. Use analysis of variance on both data sets. State the hypotheses being tested in each case. What are your conclusions?
3. Use inferences about individual treatment means where appropriate. What are your conclusions?

Case Problem 2 Compensation for Sales Professionals

Suppose that a local chapter of sales professionals in the greater San Francisco area conducted a survey of its membership to study the relationship, if any, between the years of experience and salary for individuals employed in inside and outside sales positions. On the survey, respondents were asked to specify one of three levels of years of experience: low (1–10 years), medium (11–20 years), and high (21 or more years). A portion of the data obtained follow. The complete data set, consisting of 120 observations, is contained in the file named SalesSalary.

WEB file
SalesSalary

Observation	Salary $	Position	Experience
1	53938	Inside	Medium
2	52694	Inside	Medium
3	70515	Outside	Low
4	52031	Inside	Medium
5	62283	Outside	Low
6	57718	Inside	Low
7	79081	Outside	High
8	48621	Inside	Low
9	72835	Outside	High
10	54768	Inside	Medium
.	.	.	.
.	.	.	.
.	.	.	.
115	58080	Inside	High
116	78702	Outside	Medium
117	83131	Outside	Medium
118	57788	Inside	High
119	53070	Inside	Medium
120	60259	Outside	Low

Managerial Report

1. Use descriptive statistics to summarize the data.
2. Develop a 95% confidence interval estimate of the mean annual salary for all salespersons, regardless of years of experience and type of position.
3. Develop a 95% confidence interval estimate of the mean salary for inside salespersons.
4. Develop a 95% confidence interval estimate of the mean salary for outside salespersons.
5. Use analysis of variance to test for any significant differences due to position. Use a .05 level of significance, and for now, ignore the effect of years of experience.

6. Use analysis of variance to test for any significant differences due to years of experience. Use a .05 level of significance, and for now, ignore the effect of position.
7. At the .05 level of significance test for any significant differences due to position, years of experience, and interaction.

Appendix 13.1 Analysis of Variance with Minitab

Completely Randomized Design

WEB file
Chemitech

In Section 13.2 we showed how analysis of variance could be used to test for the equality of k population means using data from a completely randomized design. To illustrate how Minitab can be used for this type of experimental design, we show how to test whether the mean number of units produced per week is the same for each assembly method in the Chemitech experiment introduced in Section 13.1. The sample data are entered into the first three columns of a Minitab worksheet; column 1 is labeled A, column 2 is labeled B, and column 3 is labeled C. The following steps produce the Minitab output in Figure 13.5.

Step 1. Select the **Stat** menu
Step 2. Choose **ANOVA**
Step 3. Choose **One-way (Unstacked)**
Step 4. When the One-way Analysis of Variance dialog box appears:
Enter C1-C3 in the **Responses (in separate columns)** box
Click **OK**

Randomized Block Design

WEB file
AirTraffic

In Section 13.4 we showed how analysis of variance could be used to test for the equality of k population means using the data from a randomized block design. To illustrate how Minitab can be used for this type of experimental design, we show how to test whether the mean stress levels for air traffic controllers are the same for three work stations using the data in Table 13.5. The blocks (controllers), treatments (system), and stress level scores shown in Table 13.5 are entered into columns C1, C2, and C3 of a Minitab worksheet, respectively. The following steps produce the Minitab output corresponding to the ANOVA table shown in Table 13.8.

The treatments are entered in the Row factor box and the blocks are entered in the Column factor box.

Step 1. Select the **Stat** menu
Step 2. Choose **ANOVA**
Step 3. Choose **Two-way**
Step 4. When the Two-way Analysis of Variance dialog box appears:
Enter C3 in the **Response** box
Enter C2 in the **Row factor** box
Enter C1 in the **Column factor** box
Select **Fit Additive Model**
Click **OK**

Factorial Experiment

WEB file
GMATStudy

In Section 13.5 we showed how analysis of variance could be used to test for the equality of k population means using data from a factorial experiment. To illustrate how Minitab can be used for this type of experimental design, we show how to analyze the data for the two-factor GMAT experiment introduced in that section. The GMAT scores shown in Table 13.11 are entered into column 1 of a Minitab worksheet; column 1 is labeled Score, column 2 is labeled Program, and column 3 is labeled College. The following steps produce the Minitab output corresponding to the ANOVA table shown in Figure 13.6.

Step 1. Select the **Stat** menu
Step 2. Choose **ANOVA**
Step 3. Choose **Two-way**
Step 4. When the Two-way Analysis of Variance dialog box appears:
Enter C1 in the **Response** box
Enter C2 in the **Row factor** box
Enter C3 in the **Column factor** box
Click **OK**

Appendix 13.2 Analysis of Variance with Excel

Completely Randomized Design

WEB file Chemitech

In Section 13.2 we showed how analysis of variance could be used to test for the equality of *k* population means using data from a completely randomized design. To illustrate how Excel can be used to test for the equality of *k* population means for this type of experimental design, we show how to test whether the mean number of units produced per week is the same for each assembly method in the Chemitech experiment introduced in Section 13.1. The sample data are entered into worksheet rows 2 to 6 of columns A, B, and C as shown in Figure 13.7. The following steps are used to obtain the output shown in cells A8:G22; the ANOVA portion of this output corresponds to the ANOVA table shown in Table 13.3.

Step 1. Click the **Data** tab on the Ribbon
Step 2. In the **Analysis** group, click **Data Analysis**
Step 3. Choose **Anova: Single Factor** from the list of Analysis Tools
Click **OK**
Step 4. When the Anova: Single Factor dialog box appears:
Enter A1:C6 in **Input Range** box
Select **Columns**
Select **Labels in First Row**
Select **Output Range** and enter A8 in the box
Click **OK**

Randomized Block Design

WEB file AirTraffic

In Section 13.4 we showed how analysis of variance could be used to test for the equality of *k* population means using data from a randomized block design. To illustrate how Excel can be used for this type of experimental design, we show how to test whether the mean stress levels for air traffic controllers are the same for three work stations. The stress level scores shown in Table 13.5 are entered into worksheet rows 2 to 7 of columns B, C, and D as shown in Figure 13.8. The cells in rows 2 to 7 of column A contain the number of each controller (1, 2, 3, 4, 5, 6). The following steps produce the Excel output shown in cells A9:G30. The ANOVA portion of this output corresponds to the Minitab output shown in Table 13.8.

Step 1. Click the **Data** tab on the Ribbon
Step 2. In the **Analysis** group, click **Data Analysis**

FIGURE 13.7 EXCEL SOLUTION FOR THE CHEMITECH EXPERIMENT

	A	B	C	D	E	F	G	H
1	**Method A**	**Method B**	**Method C**					
2	58	58	48					
3	64	69	57					
4	55	71	59					
5	66	64	47					
6	67	68	49					
7								
8	Anova: Single Factor							
9								
10	SUMMARY							
11	*Groups*	*Count*	*Sum*	*Average*	*Variance*			
12	Method A	5	310	62	27.5			
13	Method B	5	330	66	26.5			
14	Method C	5	260	52	31			
15								
16								
17	ANOVA							
18	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	
19	Between Groups	520	2	260	9.1765	0.0038	3.8853	
20	Within Groups	340	12	28.3333				
21								
22	Total	860	14					
23								
24								

Step 3. Choose **Anova: Two-Factor Without Replication** from the list of Analysis Tools
Click **OK**

Step 4. When the Anova: Two-Factor Without Replication dialog box appears:
Enter A1:D7 in **Input Range** box
Select **Labels**
Select **Output Range** and enter A9 in the box
Click **OK**

Factorial Experiment

WEB file GMATStudy

In Section 13.5 we showed how analysis of variance could be used to test for the equality of *k* population means using data from a factorial experiment. To illustrate how Excel can be used for this type of experimental design, we show how to analyze the data for the two-factor GMAT experiment introduced in that section. The GMAT scores shown in Table 13.10 are entered into worksheet rows 2 to 7 of columns B, C, and D as shown in Figure 13.9. The following steps are used to obtain the output shown in cells A9:G44; the ANOVA portion of this output corresponds to the Minitab output in Figure 13.6.

Step 1. Click the **Data** tab on the Ribbon

Step 2. In the **Analysis** group, click **Data Analysis**

Step 3. Choose **Anova: Two-Factor With Replication** from the list of Analysis Tools
Click **OK**

Step 4. When the Anova: Two-Factor With Replication dialog box appears:
Enter A1:D7 in **Input Range** box
Enter 2 in **Rows per sample** box

FIGURE 13.8 EXCEL SOLUTION FOR THE AIR TRAFFIC CONTROLLER STRESS TEST

	A	B	C	D	E	F	G	H
1	**Controller**	**System A**	**System B**	**System C**				
2	1	15	15	18				
3	2	14	14	14				
4	3	10	11	15				
5	4	13	12	17				
6	5	16	13	16				
7	6	13	13	13				
8								
9	Anova: Two-Factor Without Replication							
10								
11	*SUMMARY*	*Count*	*Sum*	*Average*	*Variance*			
12	1	3	48	16	3			
13	2	3	42	14	0			
14	3	3	36	12	7			
15	4	3	42	14	7			
16	5	3	45	15	3			
17	6	3	39	13	0			
18								
19	System A	6	81	13.5	4.3			
20	System B	6	78	13	2			
21	System C	6	93	15.5	3.5			
22								
23								
24	ANOVA							
25	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	
26	Rows	30	5	6	3.16	0.0574	3.33	
27	Columns	21	2	10.5	5.53	0.0242	4.10	
28	Error	19	10	1.9				
29								
30	Total	70	17					
31								

Select **Output Range** and enter A9 in the box
Click **OK**

Appendix 13.3 Analysis of a Completely Randomized Design Using StatTools

In this appendix we show how StatTools can be used to test for the equality of k population means for a completely randomized design. We use the Chemitech data in Table 13.1 to illustrate. Begin by using the Data Set Manager to create a StatTools data set for these data using the procedure described in the appendix in Chapter 1. The following steps can be used to test for the equality of the three population means.

WEB file
Chemitech

Step 1. Click the **StatTools** tab on the Ribbon
Step 2. In the **Analyses** group, click **Statistical Inference**
Step 3. Choose the **One-Way ANOVA** option

FIGURE 13.9 EXCEL SOLUTION FOR THE TWO-FACTOR GMAT EXPERIMENT

	A	B	C	D	E	F	G	H
1		**Business**	**Engineering**	**Arts and Sciences**				
2	**3-hour review**	500	540	480				
3		580	460	400				
4	**1-day program**	460	560	420				
5		540	620	480				
6	**10-week course**	560	600	480				
7		600	580	410				
8								
9	Anova: Two-Factor With Replication							
10								
11	SUMMARY	Business	Engineering	Arts and Sciences	Total			
12	*3-hour review*							
13	Count	2	2	2	6			
14	Sum	1080	1000	880	2960			
15	Average	540	500	440	493.33333			
16	Variance	3200	3200	3200	3946.6667			
17								
18	*1-day program*							
19	Count	2	2	2	6			
20	Sum	1000	1180	900	3080			
21	Average	500	590	450	513.33333			
22	Variance	3200	1800	1800	5386.6667			
23								
24	*10-week course*							
25	Count	2	2	2	6			
26	Sum	1160	1180	890	3230			
27	Average	580	590	445	538.33333			
28	Variance	800	200	2450	5936.6667			
29								
30	*Total*							
31	Count	6	6	6				
32	Sum	3240	3360	2670				
33	Average	540	560	445				
34	Variance	2720	3200	1510				
35								
36								
37	ANOVA							
38	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	
39	Sample	6100	2	3050	1.38	0.2994	4.26	
40	Columns	45300	2	22650	10.27	0.0048	4.26	
41	Interaction	11200	4	2800	1.27	0.3503	3.63	
42	Within	19850	9	2205.5556				
43								
44	Total	82450	17					
45								

Step 4. When the StatTools-One-Way ANOVA dialog box appears:

In the **Variables** section:

Click the **Format button** and select **Unstacked**

Select **Method A**

Select **Method B**

Select **Method C**

Select 95% in the **Confidence Level** box

Click **OK**

Note that in step 4 we selected the Unstacked option after clicking the Format button. The Unstacked option means that the data for the three treatments appear in separate columns of the worksheet. In a stacked format, only two columns would be used. For example, the data could have been organized as follows:

	A	B	C
1	**Method**	**Units Produced**	
2	Method A	58	
3	Method A	64	
4	Method A	55	
5	Method A	66	
6	Method A	67	
7	Method B	58	
8	Method B	69	
9	Method B	71	
10	Method B	64	
11	Method B	68	
12	Method C	48	
13	Method C	57	
14	Method C	59	
15	Method C	47	
16	Method C	49	
17			

Data are frequently recorded in a stacked format. For stacked data, simply select the Stacked option after clicking the Format button.