CHAPTER 11

Analysis of Variance and Design of Experiments

LEARNING OBJECTIVES

The focus of this chapter is the design of experiments and the analysis of variance, thereby enabling you to:

1. Describe an experimental design and its elements, including independent variables—both treatment and classification—and dependent variables.
2. Test a completely randomized design using a one-way analysis of variance.
3. Use multiple comparison techniques, including Tukey's honestly significant difference test and the Tukey-Kramer procedure, to test the difference in two treatment means when there is overall significant difference between treatments.
4. Test a randomized block design which includes a blocking variable to control for confounding variables.
5. Test a factorial design using a two-way analysis of variance, noting the advantages and applications of such a design and accounting for possible interaction between two treatment variables.

Lon C. Diehl/PhotoEdit

Job and Career Satisfaction of Foreign Self-Initiated Expatriates

Because of worker shortages in some industries, in a global business environment, firms around the world sometimes must compete with each other for workers. This is especially true in industries and job designations where specialty skills are required. In order to fill such needs, companies sometimes turn to self-initiated expatriates. Self-initiated expatriates are defined to be workers who are hired as individuals on a contractual basis to work in a foreign country—in contrast to individuals who are given overseas transfers by a parent organization; that is, they are "guest workers" as compared to "organizational expatriates." Some examples could be computer experts from India, China, and Japan being hired by Silicon Valley companies; American engineers working with Russian companies to extract oil and gas; or financial experts from England who are hired by Singapore companies to help manage the stock market. How satisfied are self-initiated expatriates with their jobs and their careers?

In an attempt to answer that question, suppose a study was conducted by randomly sampling self-initiated expatriates in five industries: information technology (IT), finance, education, healthcare, and consulting. Each is asked to rate his or her present job satisfaction on a 7-point Likert scale, with 7 being very satisfied and 1 being very unsatisfied. Suppose the data shown below are a portion of the study.

IT	Finance	Education	Healthcare	Consulting
5	3	2	3	6
6	4	3	2	7
5	4	3	4	5
7	5	2	3	6
	4	2	5	
		3		

Suppose in addition, self-initiated expatriates are asked to report their overall satisfaction with their career on the same 7-point scale. The ratings are broken down by the respondent's experience in the host country and age and the resultant data are shown below.

		Time in Host Country			
		<1 year	1–2 years	3–4 years	≥5 years
		3	4	3	6
	30–39	2	5	4	4
		3	3	5	5
		4	3	4	4
Age	40–49	3	4	4	6
		2	3	5	5
		4	4	5	6
	Over 50	3	4	4	5
		4	5	5	6

Managerial and Statistical Questions

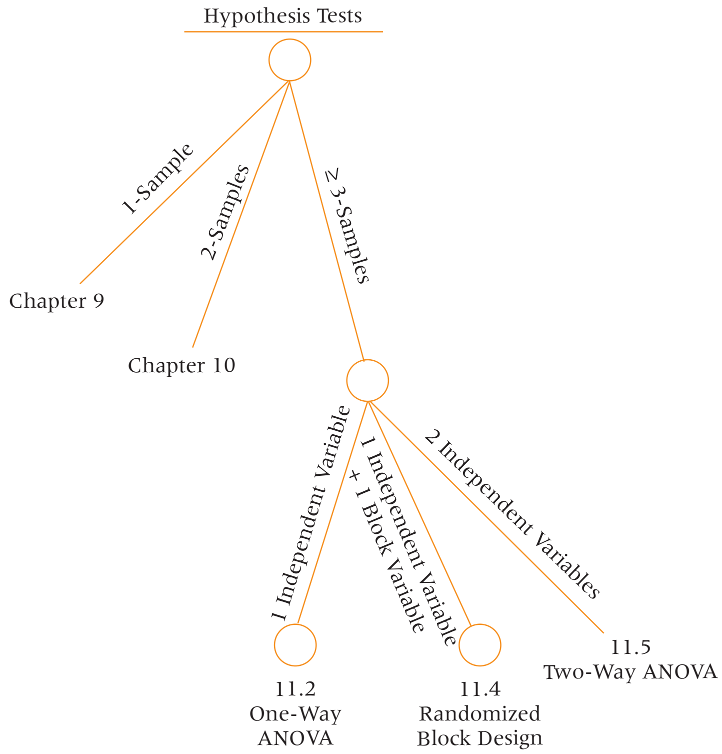
1. Is there a difference in the job satisfaction ratings of self-initiated expatriates by industry? If we were to use the *t* test for the difference of two independent population means presented in Chapter 10 to analyze these data, we would need to do 10 different *t* tests since there are five different industries. Is there a better, more parsimonious way to analyze this data? Can the analysis be done simultaneously using one technique?
2. The second table in the Decision Dilemma displays career satisfaction data broken down two different ways, age and time in country. How does a researcher analyze such data when there are two different types of groups or classifications? What if one variable, such as age, acts on another variable, such as time in the country, such that there is an interaction? That is, time in the country might matter more in one category than in another. Can this effect be measured and if so, how?

Source: Concepts adapted from Chay Hoon Lee, "A Study of Underemployment Among Self-Initiated Expatriates," *Journal of World Business* vol. 40, no. 2 (May 2005), pp. 172–187.

Sometimes business research entails more complicated hypothesis-testing scenarios than those presented to this point in the text. Instead of comparing the wear of tire tread for two brands of tires to determine whether there is a significant difference between the brands, as we could have done by using Chapter 10 techniques, a tire researcher may choose to compare three, four, or even more brands of tires at the same time. In addition, the researcher may want to include different levels of quality of tires in the experiment, such as low-quality, medium-quality, and high-quality tires. Tests may be conducted under

FIGURE 11.1

Branch of the Tree Diagram Taxonomy of Inference Techniques



varying conditions of temperature, precipitation, or road surface. Such experiments involve selecting and analyzing more than two samples of data.

Figure III-1 of the Introduction to Unit III displays the Tree Diagram Taxonomy of Inferential Techniques, organizing the techniques by usage and number of samples. The entire right side of the tree diagram taxonomy contains various hypothesis-testing techniques. Techniques for testing hypotheses using a *single* sample are presented in Chapter 9; and techniques for testing hypotheses about the differences in two populations using *two* samples are presented in Chapter 10. The far right branch of the tree diagram taxonomy contains techniques for analyzing *three or more* samples. This branch, shown in Figure 11.1, represents the techniques presented in Chapter 11.

11.1 INTRODUCTION TO DESIGN OF EXPERIMENTS

An **experimental design** is *a plan and a structure to test hypotheses in which the researcher either controls or manipulates one or more variables.* It contains independent and dependent variables. In an experimental design, an **independent variable** may be either a treatment variable or a classification variable. A **treatment variable** is *a variable the experimenter controls or modifies in the experiment.* A **classification variable** is *some characteristic of the experimental subject that was present prior to the experiment and is not a result of the experimenter's manipulations or control.* Independent variables are sometimes also referred to as **factors**. Wal-Mart executives might sanction an in-house study to compare daily sales volumes for a given size store in four different demographic settings: (1) inner-city stores (large city), (2) suburban stores (large city), (3) stores in a medium-sized city, and (4) stores in a small town. Managers might also decide to compare sales on the five different weekdays (Monday through Friday). In this study, the independent variables are store demographics and day of the week.

A finance researcher might conduct a study to determine whether there is a significant difference in application fees for home loans in five geographic regions of the

United States and might include three different types of lending organizations. In this study, the independent variables are geographic region and types of lending organizations. Or suppose a manufacturing organization produces a valve that is specified to have an opening of 6.37 centimeters. Quality controllers within the company might decide to test to determine how the openings for produced valves vary among four different machines on three different shifts. This experiment includes the independent variables of type of machine and work shift.

Whether an independent variable can be manipulated by the researcher depends on the concept being studied. Independent variables such as work shift, gender of employee, geographic region, type of machine, and quality of tire are classification variables with conditions that existed prior to the study. The business researcher cannot change the characteristic of the variable, so he or she studies the phenomenon being explored under several conditions of the various aspects of the variable. As an example, the valve experiment is conducted under the conditions of all three work shifts.

However, some independent variables can be manipulated by the researcher. For example, in the well-known Hawthorne studies of the Western Electric Company in the 1920s in Illinois, the amount of light in production areas was varied to determine the effect of light on productivity. In theory, this independent variable could be manipulated by the researcher to allow any level of lighting. Other examples of independent variables that can be manipulated include the amount of bonuses offered workers, level of humidity, and temperature. These are examples of treatment variables.

Each independent variable has two or more levels, or classifications. **Levels**, or **classifications**, of independent variables are *the subcategories of the independent variable used by the researcher in the experimental design*. For example, the different demographic settings listed for the Wal-Mart study are four levels, or classifications, of the independent variable store demographics: (1) inner-city store, (2) suburban store, (3) store in a medium-sized city, and (4) store in small town. In the valve experiment, four levels or classifications of machines within the independent variable machine type are used: machine 1, machine 2, machine 3, and machine 4.

The other type of variable in an experimental design is a **dependent variable**. A dependent variable is *the response to the different levels of the independent variables*. It is the measurement taken under the conditions of the experimental design that reflect the effects of the independent variable(s). In the Wal-Mart study, the dependent variable is the dollar amount of daily total sales. For the study on loan application fees, the fee charged for a loan application is probably the dependent variable. In the valve experiment, the dependent variable is the size of the opening of the valve.

Experimental designs in this chapter are analyzed statistically by a group of techniques referred to as **analysis of variance**, or **ANOVA**. The analysis of variance concept begins with the notion that individual items being studied, such as employees, machine-produced products, district offices, hospitals, and so on, are not all the same. Note the measurements for the openings of 24 valves randomly selected from an assembly line that are given in Table 11.1. The mean opening is 6.34 centimeters (cm). Only one of the 24 valve openings is actually the mean. Why do the valve openings vary? The total sum of squares of deviation of these valve openings around the mean is .3915 cm^2. Why is this value not zero? Using various types of experimental designs, we can explore some possible reasons for this variance with analysis of variance techniques. As we explore each of the experimental designs and their associated analysis, note that the statistical technique is attempting to "break down" the total variance among the objects being studied into possible causes. In the case

TABLE 11.1

Valve Opening Measurements (in cm) for 24 Valves Produced on an Assembly Line

6.26	6.19	6.33	6.26	6.50
6.19	6.44	6.22	6.54	6.23
6.29	6.40	6.23	6.29	6.58
6.27	6.38	6.58	6.31	6.34
6.21	6.19	6.36	6.56	

$\bar{x} = 6.34$ Total Sum of Squares Deviation $= \text{SST} = \sum(x_i - \bar{x})^2 = .3915$

of the valve openings, this variance of measurements might be due to such variables as machine, operator, shift, supplier, and production conditions, among others.

Many different types of experimental designs are available to researchers. In this chapter, we will present and discuss three specific types of experimental designs: completely randomized design, randomized block design, and factorial experiments.

11.1 PROBLEMS

11.1 Some New York Stock Exchange analysts believe that 24-hour trading on the stock exchange is the wave of the future. As an initial test of this idea, the New York Stock Exchange opened two after-hour "crossing sections" in the early 1990s and studied the results of these extra-hour sessions for one year.

a. State an independent variable that could have been used for this study.

b. List at least two levels, or classifications, for this variable.

c. Give a dependent variable for this study.

11.2 Southwest Airlines is able to keep fares low, in part because of relatively low maintenance costs on its airplanes. One of the main reasons for the low maintenance costs is that Southwest flies only one type of aircraft, the Boeing 737. However, Southwest flies three different versions of the 737. Suppose Southwest decides to conduct a study to determine whether there is a significant difference in the average annual maintenance costs for the three types of 737s used.

a. State an independent variable for such a study.

b. What are some of the levels or classifications that might be studied under this variable?

c. Give a dependent variable for this study.

11.3 A large multinational banking company wants to determine whether there is a significant difference in the average dollar amounts purchased by users of different types of credit cards. Among the credit cards being studied are MasterCard, Visa, Discover, and American Express.

a. If an experimental design were set up for such a study, what are some possible independent variables?

b. List at least three levels, or classifications, for each independent variable.

c. What are some possible dependent variables for this experiment?

11.4 Is there a difference in the family demographics of people who stay at motels? Suppose a study is conducted in which three categories of motels are used: economy motels, modestly priced chain motels, and exclusive motels. One of the dependent variables studied might be the number of children in the family of the person staying in the motel. Name three other dependent variables that might be used in this study.

11.2 THE COMPLETELY RANDOMIZED DESIGN (ONE-WAY ANOVA)

Video

Interactive Applet

One of the simplest experimental designs is the completely randomized design. In the **completely randomized design**, *subjects are assigned randomly to treatments.* The completely randomized design contains only one independent variable, with two or more treatment levels, or classifications. If only two treatment levels, or classifications, of the independent variable are present, the design is the same one used to test the difference in means of two independent populations presented in Chapter 10, which used the *t* test to analyze the data.

In this section, we will focus on completely randomized designs with three or more classification levels. Analysis of variance, or ANOVA, will be used to analyze the data that result from the treatments.

FIGURE 11.2

Completely Randomized Design

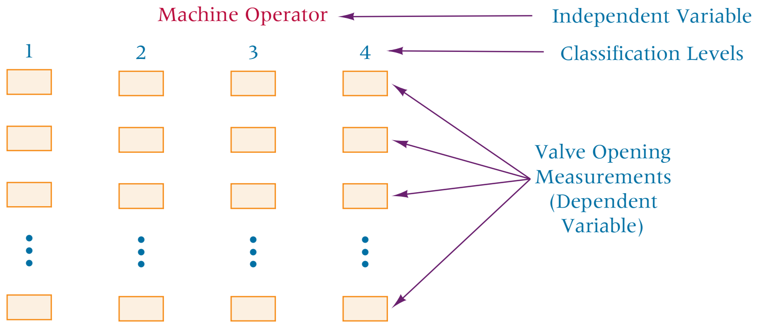


TABLE 11.2

Valve Openings by Operator

1	2	3	4
6.33	6.26	6.44	6.29
6.26	6.36	6.38	6.23
6.31	6.23	6.58	6.19
6.29	6.27	6.54	6.21
6.40	6.19	6.56	
	6.50	6.34	
	6.19	6.58	
	6.22		

A completely randomized design could be structured for a tire-quality study in which tire quality is the independent variable and the treatment levels are low, medium, and high quality. The dependent variable might be the number of miles driven before the tread fails state inspection. A study of daily sales volumes for Wal-Mart stores could be undertaken by using a completely randomized design with demographic setting as the independent variable. The treatment levels, or classifications, would be inner-city stores, suburban stores, stores in medium-sized cities, and stores in small towns. The dependent variable would be sales dollars.

As an example of a completely randomized design, suppose a researcher decides to analyze the effects of the machine operator on the valve opening measurements of valves produced in a manufacturing plant, like those shown in Table 11.1. The independent variable in this design is machine operator. Suppose further that four different operators operate the machines. These four machine operators are the levels of treatment, or classification, of the independent variable. The dependent variable is the opening measurement of the valve. Figure 11.2 shows the structure of this completely randomized design. Is there a significant difference in the mean valve openings of 24 valves produced by the four operators? Table 11.2 contains the valve opening measurements for valves produced under each operator.

One-Way Analysis of Variance

In the machine operator example, is it possible to analyze the four samples by using a *t* test for the difference in two sample means? These four samples would require ${}_4C_2 = 6$ individual *t* tests to accomplish the analysis of two groups at a time. Recall that if $\alpha = .05$ for a particular test, there is a 5% chance of rejecting a null hypothesis that is true (i.e., committing a Type I error). If enough tests are done, eventually one or more null hypotheses will be falsely rejected by chance. Hence, $\alpha = .05$ is valid only for one *t* test. In this problem, with six *t* tests, the error rate compounds, so when the analyst is finished with the problem there is a much greater than .05 chance of committing a Type I error. Fortunately, a technique has been developed that analyzes all the sample means at one time and thus precludes the buildup of error rate: analysis of variance (ANOVA). A completely randomized design is analyzed by a **one-way analysis of variance**.

In general, if *k* samples are being analyzed, the following hypotheses are being tested in a one-way ANOVA.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

H_a: At least one of the means is different from the others.

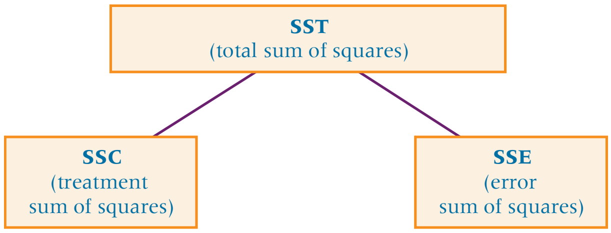
The null hypothesis states that the population means for all treatment levels are equal. Because of the way the alternative hypothesis is stated, if even one of the population means is different from the others, the null hypothesis is rejected.

Testing these hypotheses by using one-way ANOVA is accomplished by partitioning the total variance of the data into the following two variances.

1. The variance resulting from the treatment (columns)
2. The error variance, or that portion of the total variance unexplained by the treatment

FIGURE 11.3

Partitioning Total Sum of Squares of Variation



As part of this process, the total sum of squares of deviation of values around the mean can be divided into two additive and independent parts.

$$\begin{array}{ccccc} \text{SST} & = & \text{SSC} & + & \text{SSE} \\ \sum_{i=1}^{n_j}\sum_{j=1}^{C}(x_{ij}-\bar{x})^2 & = & \sum_{j=1}^{C} n_j(\bar{x}_j-\bar{x})^2 & + & \sum_{i=1}^{n_j}\sum_{j=1}^{C}(x_{ij}-\bar{x}_j)^2 \end{array}$$

where

SST = total sum of squares
SSC = sum of squares column (treatment)
SSE = sum of squares error
i = particular member of a treatment level
j = a treatment level
C = number of treatment levels
n_j = number of observations in a given treatment level
$\bar{x}$ = grand mean
$\bar{x}_j$ = mean of a treatment group or level
x_{ij} = individual value

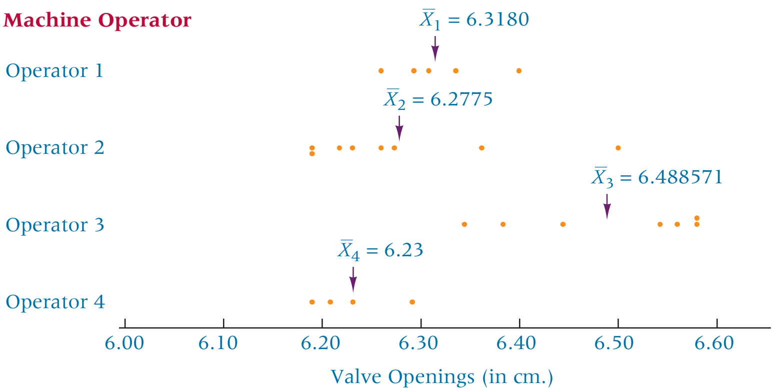
This relationship is shown in Figure 11.3. Observe that the total sum of squares of variation is partitioned into the sum of squares of treatment (columns) and the sum of squares of error.

The formulas used to accomplish one-way analysis of variance are developed from this relationship. The double summation sign indicates that the values are summed within a treatment level and across treatment levels. Basically, ANOVA compares the relative sizes of the *treatment* variation and the *error* variation (within-group variation). The error variation is unaccounted-for variation and can be viewed at this point as variation due to individual differences within treatment groups. If a significant difference in treatments is present, the treatment variation should be large relative to the error variation.

Figure 11.4 displays the data from the machine operator example in terms of treatment level. Note the variation of values (x) *within* each treatment level. Now examine the variation between levels 1 through 4 (the difference in the machine operators). In particular, note

FIGURE 11.4

Location of Mean Value Openings by Operator



that values for treatment level 3 seem to be located differently from those of levels 2 and 4. This difference also is underscored by the mean values for each treatment level:

$$\bar{x}_1 = 6.3180 \quad \bar{x}_2 = 6.2775 \quad \bar{x}_3 = 6.488571 \quad \bar{x}_4 = 6.23$$

Analysis of variance is used to determine statistically whether the variance between the treatment level means is greater than the variances within levels (error variance). Several important assumptions underlie analysis of variance:

1. Observations are drawn from normally distributed populations.
2. Observations represent random samples from the populations.
3. Variances of the populations are equal.

These assumptions are similar to those for using the *t* test for independent samples in Chapter 10. It is assumed that the populations are normally distributed and that the population variances are equal. These techniques should be used only with random samples.

An ANOVA is computed with the three sums of squares: total, treatment (columns), and error. Shown here are the formulas to compute a one-way analysis of variance. The term SS represents sum of squares, and the term MS represents mean square. SSC is the sum of squares columns, which yields the sum of squares between treatments. It measures the variation between columns or between treatments since the independent variable treatment levels are presented as columns. SSE is the sum of squares of error, which yields the variation within treatments (or columns). Some say that it is a measure of the individual differences unaccounted for by the treatments. SST is the total sum of squares and is a measure of all variation in the dependent variable. As shown previously, SST contains both SSC and SSE and can be partitioned into SSC and SSE. MSC, MSE, and MST are the mean squares of column, error, and total, respectively. Mean square is an average and is computed by dividing the sum of squares by the degrees of freedom. Finally, the *F* value is determined by dividing the treatment variance (MSC) by the error variance (MSE). As discussed in Chapter 10, the *F* is a ratio of two variances. In the ANOVA situation, the ***F* value** is *a ratio of the treatment variance to the error variance.*

FORMULAS FOR COMPUTING A ONE-WAY ANOVA

$$\text{SSC} = \sum_{j=1}^{C} n_j(\bar{x}_j - \bar{x})^2$$

$$\text{SSE} = \sum_{i=1}^{n_j} \sum_{j=1}^{C} (x_{ij} - \bar{x}_j)^2$$

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

$$\text{df}_C = C - 1$$

$$\text{df}_E = N - C$$

$$\text{df}_T = N - 1$$

$$\text{MSC} = \frac{\text{SSC}}{\text{df}_C}$$

$$\text{MSE} = \frac{\text{SSE}}{\text{df}_E}$$

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

where

i = a particular member of a treatment level
j = a treatment level
C = number of treatment levels
n_j = number of observations in a given treatment level
$\bar{x}$ = grand mean
$\bar{x}_j$ = column mean
x_{ij} = individual value

Performing these calculations for the machine operator example yields the following.

Machine Operator

1	2	3	4
6.33	6.26	6.44	6.29
6.26	6.36	6.38	6.23
6.31	6.23	6.58	6.19
6.29	6.27	6.54	6.21
6.40	6.19	6.56	
	6.50	6.34	
	6.19	6.58	
	6.22		

T_j:	$T_1 = 31.59$	$T_2 = 50.22$	$T_3 = 45.42$	$T_4 = 24.92$	$T = 152.15$
n_j:	$n_1 = 5$	$n_2 = 8$	$n_3 = 7$	$n_4 = 4$	$N = 24$
$\bar{x}_j$:	$\bar{x}_1 = 6.318$	$\bar{x}_2 = 6.2775$	$\bar{x}_3 = 6.488571$	$\bar{x}_4 = 6.230$	$\bar{x} = 6.339583$

$$\begin{aligned}\text{SSC} = \sum_{j=1}^{C} n_j(\bar{x}_j - \bar{x})^2 &= [5(6.318 - 6.339583)^2 + 8(6.2775 - 6.339583)^2 \\ &\quad + 7(6.488571 - 6.339583)^2 + 4(6.230 - 6.339583)^2] \\ &= 0.00233 + 0.03083 + 0.15538 + 0.04803 \\ &= 0.23658\end{aligned}$$

$$\begin{aligned}\text{SSE} = \sum_{i=1}^{n_j}\sum_{j=1}^{C}(x_{ij} - \bar{x}_j)^2 &= [(6.33 - 6.318)^2 + (6.26 - 6.318)^2 + (6.31 - 6.318)^2 \\ &\quad + (6.29 - 6.318)^2 + (6.40 - 6.318)^2 + (6.26 - 6.2775)^2 \\ &\quad + (6.36 - 6.2775)^2 + \ldots + (6.19 - 6.230)^2 + (6.21 - 6.230)^2 \\ &= 0.15492\end{aligned}$$

$$\begin{aligned}\text{SST} = \sum_{i=1}^{n_j}\sum_{j=1}^{C}(x_{ij} - \bar{x})^2 &= [(6.33 - 6.339583)^2 + (6.26 - 6.339583)^2 \\ &\quad + (6.31 - 6.339583)^2 + \ldots + (6.19 - 6.339583)^2 \\ &\quad + (6.21 - 6.339583)^2 \\ &= 0.39150\end{aligned}$$

$$\text{df}_C = C - 1 = 4 - 1 = 3$$

$$\text{df}_E = N - C = 24 - 4 = 20$$

$$\text{df}_T = N - 1 = 24 - 1 = 23$$

$$\text{MSC} = \frac{\text{SSC}}{\text{df}_C} = \frac{.23658}{3} = .078860$$

$$\text{MSE} = \frac{\text{SSE}}{\text{df}_E} = \frac{.15492}{20} = .007746$$

$$F = \frac{.078860}{.007746} = 10.18$$

From these computations, an analysis of variance chart can be constructed, as shown in Table 11.3. The observed F value is 10.18. It is compared to a critical value from the F table to determine whether there is a significant difference in treatment or classification.

TABLE 11.3
Analysis of Variance for the Machine Operator Example

Source of Variance	df	SS	MS	F
Between	3	0.23658	0.078860	10.18
Error	20	0.15492	0.007746	
Total	23	0.39150		

TABLE 11.4

An Abbreviated F Table for $\alpha = .05$

Denominator Degrees of Freedom	NUMERATOR DEGREES OF FREEDOM 1	2	3	4	5	6	7	8	9
.									
.									
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37

Reading the *F* Distribution Table

The ***F* distribution** table is in Table A.7. Associated with every F value in the table are two unique df values: degrees of freedom in the numerator (df_C) and degrees of freedom in the denominator (df_E). To look up a value in the F distribution table, the researcher must know both degrees of freedom. Because each F distribution is determined by a unique pair of degrees of freedom, many F distributions are possible. Space constraints limit Table A.7 to F values for only $\alpha = .005, .01, .025, .05$, and $.10$. However, statistical computer software packages for computing ANOVAs usually give a probability for the F value, which allows a hypothesis-testing decision for any alpha based on the p-value method.

In the one-way ANOVA, the df_C values are the treatment (column) degrees of freedom, $C - 1$. The df_E values are the error degrees of freedom, $N - C$. Table 11.4 contains an abbreviated F distribution table for $\alpha = .05$. For the machine operator example, $df_C = 3$ and $df_E = 20$, $F_{.05,3,20}$ from Table 11.4 is 3.10. This value is the critical value of the F test. Analysis of variance tests are always one-tailed tests with the rejection region in the upper tail. The decision rule is to reject the null hypothesis if the observed F value is greater than the critical F value ($F_{.05,3,20} = 3.10$). For the machine operator problem, the observed F value of 10.18 is larger than the table F value of 3.10. The null hypothesis is rejected. Not all means are equal, so there is a significant difference in the mean valve openings by machine operator. Figure 11.5 is a Minitab graph of an F distribution showing the critical F value for this example and the rejection region. Note that the F distribution begins at zero and contains no negative values because the F value is the ratio of two variances, and variances are always positive.

Using the Computer for One-Way ANOVA

Many researchers use the computer to analyze data with a one-way ANOVA. Figure 11.6 shows the Minitab and Excel output of the ANOVA computed for the machine operator example. The output includes the analysis of variance table presented in Table 11.3. Both Minitab and Excel ANOVA tables display the observed F value, mean squares, sum of squares, degrees of freedom, and a value of p. The value of p is the probability of an F value of 10.18 occurring by chance in an ANOVA with this structure (same degrees of freedom)

FIGURE 11.5

Minitab Graph of F Values for the Machine Operator Example

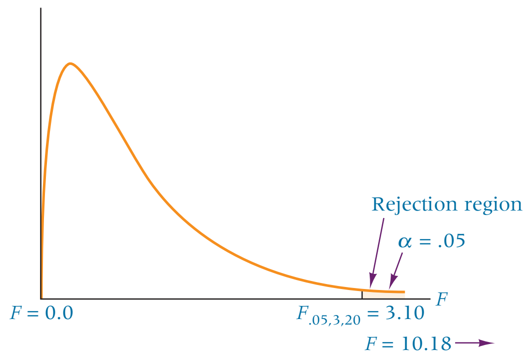


FIGURE 11.6

Minitab and Excel Analysis of the Machine Operator Problem

Minitab Output

One-way ANOVA: Operator 1, Operator 2, Operator 3, Operator 4

```
Source  DF       SS       MS      F      P
Factor   3  0.23658  0.07886  10.18  0.000
Error   20  0.15492  0.00775
Total   23  0.39150

S = 0.08801  R-Sq = 60.43%  R-Sq(adj) = 54.49%

                                Individual 95% CIs For Mean Based on
                                Pooled StDev
Level       N    Mean   StDev  ---------+---------+---------+---------+---
Operator 1  5  6.3180  0.0526          (-----*-----)
Operator 2  8  6.2775  0.1053        (----*-----)
Operator 3  7  6.4886  0.1006                          (--- * ---)
Operator 4  4  6.2300  0.0432   (------*-------)
                               ---------+---------+---------+---------+---
Pooled StDev = 0.0880                6.24      6.36      6.48      6.60
```

Excel Output

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Operator 1	5	31.59	6.318000	0.002770
Operator 2	8	50.22	6.277500	0.011079
Operator 3	7	45.42	6.488571	0.010114
Operator 4	4	24.92	6.230000	0.001867

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.236580	3	0.078860	10.18	0.000279	3.10
Within Groups	0.154916	20	0.007746			
Total	0.391496	23				

even if there is no difference between means of the treatment levels. Using the p-value method of testing hypotheses presented in Chapter 9, we can easily see that because this p-value is only .000279, the null hypothesis would be rejected using $\alpha = .05$. Most computer output yields the value of p, so there is no need to look up a table value of F against which to compare the observed F value. The Excel output also includes the critical table F value for this problem, $F_{.05,3,20} = 3.10$.

The second part of the Minitab output in Figure 11.6 contains the size of samples and sample means for each of the treatment levels. Displayed graphically are the 95% confidence levels for the population means of each treatment level group. These levels are computed by using a pooled standard deviation from all the treatment level groups. The researcher can visually observe the confidence intervals and make a subjective determination about the relative difference in the population means. More rigorous statistical techniques for testing the differences in pairs of groups are given in Section 11.3.

Comparison of *F* and *t* Values

Analysis of variance can be used to test hypotheses about the difference in two means. Analysis of data from two samples by both a t test and an ANOVA shows that the observed F value equals the observed t value squared.

$$F = t^2 \qquad \text{for df}_C = 1$$

The t test of independent samples actually is a special case of one-way ANOVA when there are only two treatment levels ($df_C = 1$). The t test is computationally simpler than ANOVA for two groups. However, some statistical computer software packages do not contain a t test. In these cases, the researcher can perform a one-way ANOVA and then either take the square root of the F value to obtain the value of t or use the generated probability with the p-value method to reach conclusions.

DEMONSTRATION PROBLEM 11.1

A company has three manufacturing plants, and company officials want to determine whether there is a difference in the average age of workers at the three locations. The following data are the ages of five randomly selected workers at each plant. Perform a one-way ANOVA to determine whether there is a significant difference in the mean ages of the workers at the three plants. Use $\alpha = .01$ and note that the sample sizes are equal.

Solution

HYPOTHESIZE:

STEP 1. The hypotheses follow.

H_0: $\mu_1 = \mu_2 = \mu_3$

H_a: At least one of the means is different from the others.

TEST:

STEP 2. The appropriate test statistic is the F test calculated from ANOVA.

STEP 3. The value of α is .01.

STEP 4. The degrees of freedom for this problem are $3 - 1 = 2$ for the numerator and $15 - 3 = 12$ for the denominator. The critical F value is $F_{.01,2,12} = 6.93$.

Because ANOVAs are always one tailed with the rejection region in the upper tail, the decision rule is to reject the null hypothesis if the observed value of F is greater than 6.93.

STEP 5.

Plant (Employee Ages)

1	2	3
29	32	25
27	33	24
30	31	24
27	34	25
28	30	26

STEP 6.

T_j:	$T_1 = 141$	$T_2 = 160$	$T_3 = 124$	$T = 425$
n_j:	$n_1 = 5$	$n_2 = 5$	$n_3 = 5$	$N = 15$
$\bar{x}_j$:	$\bar{x}_1 = 28.2$	$\bar{x}_2 = 32.0$	$\bar{x}_3 = 24.8$	$\bar{x} = 28.33$

$$SSC = 5(28.2 - 28.33)^2 + 5(32.0 - 28.33)^2 + 5(24.8 - 28.33)^2 = 129.73$$

$$SSE = (29 - 28.2)^2 + (27 - 28.2)^2 + \ldots + (25 - 24.8)^2 + (26 - 24.8)^2 = 19.60$$

$$SST = (29 - 28.33)^2 + (27 - 28.33)^2 + \ldots + (25 - 28.33)^2 + (26 - 28.33)^2 = 149.33$$

$$df_C = 3 - 1 = 2$$

$$df_E = 15 - 3 = 12$$

$$df_T = 15 - 1 = 14$$

Source of Variance	SS	df	MS	*F*
Between	129.73	2	64.87	39.80
Error	19.60	12	1.63	
Total	149.33	14		

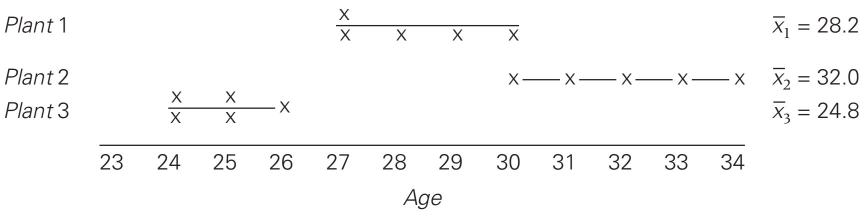
ACTION:

STEP 7. The decision is to reject the null hypothesis because the observed *F* value of 39.80 is greater than the critical table *F* value of 6.93.

BUSINESS IMPLICATIONS:

STEP 8. There is a significant difference in the mean ages of workers at the three plants. This difference can have hiring implications. Company leaders should understand that because motivation, discipline, and experience may differ with age, the differences in ages may call for different managerial approaches in each plant.

The chart on the next page displays the dispersion of the ages of workers from the three samples, along with the mean age for each plant sample. Note the difference in group means. The significant *F* value says that the difference between the mean ages is relatively greater than the differences of ages within each group.



Following are the Minitab and Excel output for this problem.

Minitab Output

```
One-way ANOVA: Plant 1, Plant 2, Plant 3
Source   DF      SS     MS      F      P
Factor    2  129.73  64.87  39.71  0.000
Error    12   19.60   1.63
Total    14  149.33
S = 1.278 R-Sq = 86.88% R-Sq(adj) = 84.69%
                               Individual 95% CIs For Mean
                               Based on Pooled StDev
Level     N    Mean   StDev   --+---------+---------+---------+----
Plant 1   5  28.200   1.304              (---*---)
Plant 2   5  32.000   1.581                          (---*---)
Plant 3   5  24.800   0.837    (---*---)
                              --+---------+---------+---------+----
Pooled StDev = 1.278          25.0      27.5      30.0      32.5
```

Excel Output

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Plant 1	5	141	28.2	1.7
Plant 2	5	160	32	2.5
Plant 3	5	124	24.8	0.7

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	129.73333	2	64.8667	39.71	0.0000051	3.89
Within Groups	19.6	12	1.6333			
Total	149.33333	14				

11.2 PROBLEMS

11.5 Compute a one-way ANOVA on the following data.

1	2	3
2	5	3
1	3	4
3	6	5
3	4	5
2	5	3
1		5

Determine the observed F value. Compare the observed F value with the critical table F value and decide whether to reject the null hypothesis. Use $\alpha = .05$.

11.6 Compute a one-way ANOVA on the following data.

1	2	3	4	5
14	10	11	16	14
13	9	12	17	12
10	12	13	14	13
	9	12	16	13
	10		17	12
				14

Determine the observed F value. Compare the observed F value with the critical table F value and decide whether to reject the null hypothesis. Use $\alpha = .01$.

11.7 Develop a one-way ANOVA on the following data.

1	2	3	4
113	120	132	122
121	127	130	118
117	125	129	125
110	129	135	125

Determine the observed F value. Compare it to the critical F value and decide whether to reject the null hypothesis. Use a 1% level of significance.

11.8 Compute a one-way ANOVA on the following data.

1	2
27	22
31	27
31	25
29	23
30	26
27	27
28	23

Determine the observed F value. Compare it to the critical table F value and decide whether to reject the null hypothesis. Perform a t test for independent measures on the data. Compare the t and F values. Are the results different? Use $\alpha = .05$.

11.9 Suppose you are using a completely randomized design to study some phenomenon. There are five treatment levels and a total of 55 people in the study. Each treatment level has the same sample size. Complete the following ANOVA.

Source of Variance	SS	df	MS	F
Treatment	583.39			
Error	972.18			
Total	1555.57			

11.10 Suppose you are using a completely randomized design to study some phenomenon. There are three treatment levels and a total of 17 people in the study. Complete the following ANOVA table. Use $\alpha = .05$ to find the table F value and use the data to test the null hypothesis.

Source of Variance	SS	df	MS	F
Treatment	29.64			
Error	68.42			
Total				

11.11 A milk company has four machines that fill gallon jugs with milk. The quality control manager is interested in determining whether the average fill for these machines is the same. The following data represent random samples of fill measures (in quarts) for 19 jugs of milk filled by the different machines. Use $\alpha = .01$ to test the hypotheses. Discuss the business implications of your findings.

Machine 1	Machine 2	Machine 3	Machine 4
4.05	3.99	3.97	4.00
4.01	4.02	3.98	4.02
4.02	4.01	3.97	3.99
4.04	3.99	3.95	4.01
	4.00	4.00	
	4.00		

11.12 That the starting salaries of new accounting graduates would differ according to geographic regions of the United States seems logical. A random selection of accounting firms is taken from three geographic regions, and each is asked to state the starting salary for a new accounting graduate who is going to work in auditing. The data obtained follow. Use a one-way ANOVA to analyze these data. Note that the data can be restated to make the computations more reasonable (example: $42,500 = 4.25). Use a 1% level of significance. Discuss the business implications of your findings.

South	Northeast	West
$40,500	$51,000	$45,500
41,500	49,500	43,500
40,000	49,000	45,000
41,000	48,000	46,500
41,500	49,500	46,000

11.13 A management consulting company presents a three-day seminar on project management to various clients. The seminar is basically the same each time it is given. However, sometimes it is presented to high-level managers, sometimes to midlevel managers, and sometimes to low-level managers. The seminar facilitators believe evaluations of the seminar may vary with the audience. Suppose the following data are some randomly selected evaluation scores from different levels of managers who attended the seminar. The ratings are on a scale from 1 to 10, with 10 being the highest. Use a one-way ANOVA to determine whether there is a significant difference in the evaluations according to manager level. Assume $\alpha = .05$. Discuss the business implications of your findings.

High Level	Midlevel	Low Level
7	8	5
7	9	6
8	8	5
7	10	7
9	9	4
	10	8
	8	

11.14 Family transportation costs are usually higher than most people believe because those costs include car payments, insurance, fuel costs, repairs, parking, and public transportation. Twenty randomly selected families in four major cities are asked to use their records to estimate a monthly figure for transportation cost. Use the data obtained and ANOVA to test whether there is a significant difference in monthly transportation costs for families living in these cities. Assume that $\alpha = .05$. Discuss the business implications of your findings.

Atlanta	New York	Los Angeles	Chicago
$650	$250	$850	$540
480	525	700	450
550	300	950	675
600	175	780	550
675	500	600	600

11.15 Shown here is the Minitab output for a one-way ANOVA. Analyze the results. Include the number of treatment levels, the sample sizes, the *F* value, the overall statistical significance of the test, and the values of the means.

One-Way Analysis of Variance
Analysis of Variance

Source	df	SS	MS	*F*	*p*
Factor	3	1701	567	2.95	0.040
Error	61	11728	192		
Total	64	13429			

Individual 95% CIs For Mean Based on Pooled StDev

Level	*N*	Mean	StDev
C1	18	226.73	13.59
C2	15	238.79	9.41
C3	21	232.58	12.16
C4	11	239.82	20.96

Pooled StDev = 13.87

11.16 Business is very good for a chemical company. In fact, it is so good that workers are averaging more than 40 hours per week at each of the chemical company's five plants. However, management is not certain whether there is a difference between the five plants in the average number of hours worked per week per worker. Random samples of data are taken at each of the five plants. The data are analyzed using Excel. The results follow. Explain the design of the study and determine whether there is an overall significant difference between the means at $\alpha = .05$? Why or why not? What are the values of the means? What are the business implications of this study to the chemical company?

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Plant 1	11	636.5577	57.87	63.5949
Plant 2	12	601.7648	50.15	62.4813
Plant 3	8	491.7352	61.47	47.4772
Plant 4	5	246.0172	49.20	65.6072
Plant 5	7	398.6368	56.95	140.3540

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	900.0863	4	225.0216	3.10	0.026595	2.62
Within Groups	2760.136	38	72.63516			
Total	3660.223	42				

11.3 MULTIPLE COMPARISON TESTS

Analysis of variance techniques are particularly useful in testing hypotheses about the differences of means in multiple groups because ANOVA utilizes only one single overall test. The advantage of this approach is that the probability of committing a Type I error, α, is controlled. As noted in Section 11.2, if four groups are tested two at a time, it takes six t tests ($_4C_2$) to analyze hypotheses between all possible pairs. In general, if k groups are tested two at a time, $_kC_2 = k(k-1)/2$ paired comparisons are possible.

Suppose alpha for an experiment is .05. If two different pairs of comparisons are made in the experiment using alpha of .05 in each, there is a .95 probability of not making a Type I error in each comparison. This approach results in a .9025 probability of not making a Type I error in either comparison ($.95 \times .95$), and a .0975 probability of committing a Type I error in at least one comparison ($1 - .9025$). Thus, the probability of committing a Type I error for this experiment is not .05 but .0975. In an experiment where the means of four groups are being tested two at a time, six different tests are conducted. If each is analyzed using $\alpha = .05$, the probability that no Type I error will be committed in any of the six tests is $.95 \times .95 \times .95 \times .95 \times .95 \times .95 = .735$ and the probability of committing at least one Type I error in the six tests is $1 - .735 = .265$. If an ANOVA is computed on all groups simultaneously using $\alpha = .05$, the value of alpha is maintained in the experiment.

Sometimes the researcher is satisfied with conducting an overall test of differences in groups such as the one ANOVA provides. However, when it is determined that there is an overall difference in population means, it is often desirable to go back to the groups and determine from the data which pairs of means are significantly different. Such pairwise analyses can lead to the buildup of the Type I experimental error rate, as mentioned. Fortunately, several techniques, referred to as **multiple comparisons**, have been developed to handle this problem.

Multiple comparisons are to be used only when an overall significant difference between groups has been obtained by using the F value of the analysis of variance. Some of these techniques protect more for Type I errors and others protect more for Type II errors. Some multiple comparison techniques require equal sample sizes. There seems to be some difference of opinion in the literature about which techniques are most appropriate. Here we will consider only a posteriori or post hoc pairwise comparisons.

A posteriori or **post hoc** pairwise comparisons are made *after the experiment when the researcher decides to test for any significant differences in the samples based on a significant overall F value.* In contrast, **a priori** comparisons are made when the researcher *determines before the experiment which comparisons are to be made.* The error rates for these two types of comparisons are different, as are the recommended techniques. In this text, we only consider pairwise (two-at-a-time) multiple comparisons. Other types of comparisons are possible but belong in a more advanced presentation. The two multiple comparison tests discussed here are Tukey's HSD test for designs with equal sample sizes and the Tukey-Kramer procedure for situations in which sample sizes are unequal. Minitab yields computer output for each of these tests.

Tukey's Honestly Significant Difference (HSD) Test: The Case of Equal Sample Sizes

Tukey's honestly significant difference (HSD) test, sometimes known as Tukey's T method, is a popular test for pairwise a posteriori multiple comparisons. This test, developed by John W. Tukey and presented in 1953, is somewhat limited by the fact that it requires equal sample sizes.

Tukey's HSD test takes into consideration the number of treatment levels, the value of mean square error, and the sample size. Using these values and a table value, q, the HSD determines the critical difference necessary between the means of any two treatment levels for the means to be significantly different. Once the HSD is computed, the researcher can examine the absolute value of any or all differences between pairs of means from treatment levels to determine whether there is a significant difference. The formula to compute a Tukey's HSD test follows.

TUKEY'S HSD TEST

$$\text{HSD} = q_{\alpha,C,N-C}\sqrt{\frac{\text{MSE}}{n}}$$

where:

MSE = mean square error
n = sample size
$q_{\alpha,C,N-C}$ = critical value of the studentized range distribution from Table A.10

In Demonstration Problem 11.1, an ANOVA test was used to determine that there was an overall significant difference in the mean ages of workers at the three different plants, as evidenced by the F value of 39.8. The sample data for this problem follow.

	PLANT		
	1	2	3
	29	32	25
	27	33	24
	30	31	24
	27	34	25
	28	30	26
Group Means	28.2	32.0	24.8
n_j	5	5	5

Because the sample sizes are equal in this problem, Tukey's HSD test can be used to compute multiple comparison tests between groups 1 and 2, 2 and 3, and 1 and 3. To compute the HSD, the values of MSE, n, and q must be determined. From the solution presented in Demonstration Problem 11.1, the value of MSE is 1.63. The sample size, n_j, is 5. The value of q is obtained from Table A.10 by using

$$\text{Number of Populations} = \text{Number of Treatment Means} = C$$

along with $\text{df}_E = N - C$.

In this problem, the values used to look up q are

$$C = 3$$
$$\text{df}_E = N - C = 12$$

Table A.10 has a q table for $\alpha = .05$ and one for $\alpha = .01$. In this problem, $\alpha = .01$. Shown in Table 11.5 is a portion of Table A.10 for $\alpha = .01$.

For this problem, $q_{.01,3,12} = 5.04$. HSD is computed as

$$\text{HSD} = q\sqrt{\frac{\text{MSE}}{n}} = 5.04\sqrt{\frac{1.63}{5}} = 2.88$$

TABLE 11.5
q Values for $\alpha = .01$

	Number of Populations				
Degrees of Freedom	2	3	4	5	. . .
1	90	135	164	186	
2	14	19	22.3	24.7	
3	8.26	10.6	12.2	13.3	
4	6.51	8.12	9.17	9.96	
.					
.					
.					
11	4.39	5.14	5.62	5.97	
12	4.32	5.04	5.50	5.84	

TABLE 11.6
Minitab Output for Tukey's HSD

```
Tukey 99% Simultaneous Confidence Intervals
All Pairwise Comparisons

Individual confidence level = 99.62%

Plant 1 subtracted from:

         Lower Center  Upper     -----+--------+-------+--------+-
Plant 2  0.914  3.800  6.686            (---*---) (---*---)
Plant 3 -6.286 -3.400 -0.514     -----+--------+-------+--------+-
                                     -6.0      0.0     6.0     12.0
Plant 2 subtracted from:

          Lower Center  Upper    -----+--------+-------+--------+-
Plant 3 -10.086 -7.200 -4.314    (---*---)
                                 -----+--------+-------+--------+-
                                     -6.0      0.0     6.0     12.0
```

Using this value of HSD, the business researcher can examine the differences between the means from any two groups of plants. Any of the pairs of means that differ by more than 2.88 are significantly different at $\alpha = .01$. Here are the differences for all three possible pairwise comparisons.

$$|\bar{x}_1 - \bar{x}_2| = |28.2 - 32.0| = 3.8$$
$$|\bar{x}_1 - \bar{x}_3| = |28.2 - 24.8| = 3.4$$
$$|\bar{x}_2 - \bar{x}_3| = |32.0 - 24.8| = 7.2$$

All three comparisons are greater than the value of HSD, which is 2.88. Thus, the mean ages between any and all pairs of plants are significantly different.

Using the Computer to Do Multiple Comparisons

Table 11.6 shows the Minitab output for computing a Tukey's HSD test. The computer output contains the confidence intervals for the differences in pairwise means for pairs of treatment levels. If the confidence interval includes zero, there is no significant difference in the pair of means. (If the interval contains zero, there is a possibility of no difference in the means.) Note in Table 11.6 that all three pairs of confidence intervals contain the same sign throughout the interval. For example, the confidence interval for estimating the difference in means from 1 and 2 is $0.914 \le \mu_1 - \mu_2 \le 6.686$. This interval does not contain zero, so we are confident that there is more than a zero difference in the two means. The same holds true for levels 1 and 3 and levels 2 and 3.

DEMONSTRATION PROBLEM 11.2

A metal-manufacturing firm wants to test the tensile strength of a given metal under varying conditions of temperature. Suppose that in the design phase, the metal is processed under five different temperature conditions and that random samples of size five are taken under each temperature condition. The data follow.

Tensile Strength of Metal Produced Under Five Different Temperature Settings

1	2	3	4	5
2.46	2.38	2.51	2.49	2.56
2.41	2.34	2.48	2.47	2.57
2.43	2.31	2.46	2.48	2.53
2.47	2.40	2.49	2.46	2.55
2.46	2.32	2.50	2.44	2.55

A one-way ANOVA is performed on these data by using Minitab, with the resulting analysis shown here.

```
One-way ANOVA: Tensile Strength versus Temp. Setting

Source          DF        SS        MS      F      P
Temp. Setting    4  0.108024  0.027006  43.70  0.000
Error           20  0.012360  0.000618
Total           24  0.120384

S = 0.02486  R-Sq = 89.73%  R-Sq(adj) = 87.68%
```

Note from the ANOVA table that the F value of 43.70 is statistically significant at $\alpha = .01$. There is an overall difference in the population means of metal produced under the five temperature settings. Use the data to compute a Tukey's HSD to determine which of the five groups are significantly different from the others.

Solution

From the ANOVA table, the value of MSE is .000618. The sample size, n_j, is 5. The number of treatment means, C, is 5 and the df_E are 20. With these values and $\alpha = .01$, the value of q can be obtained from Table A.10.

$$q_{.01,5,20} = 5.29$$

HSD can be computed as

$$\text{HSD} = q\sqrt{\frac{\text{MSE}}{n}} = 5.29\sqrt{\frac{.000618}{5}} = .0588$$

The treatment group means for this problem follow.

Group 1 = 2.446
Group 2 = 2.350
Group 3 = 2.488
Group 4 = 2.468
Group 5 = 2.552

Computing all pairwise differences between these means (in absolute values) produces the following data.

	Group				
	1	**2**	**3**	**4**	**5**
1	—	.096	.042	.022	.106
2	.096	—	.138	.118	.202
3	.042	.138	—	.020	.064
4	.022	.118	.020	—	.084
5	.106	.202	.064	.084	—

Comparing these differences to the value of HSD = .0588, we can determine that the differences between groups 1 and 2 (.096), 1 and 5 (.106), 2 and 3 (.138), 2 and 4 (.118), 2 and 5 (.202), 3 and 5 (.064), and 4 and 5 (.084) are significant at $\alpha = .01$.

Not only is there an overall significant difference in the treatment levels as shown by the ANOVA results, but there is a significant difference in the tensile strength of metal between seven pairs of levels. By studying the magnitudes of the individual treatment levels' means, the steel-manufacturing firm can determine which temperatures result in the greatest tensile strength. The Minitab output for this Tukey's HSD is shown on the next page. Note that the computer analysis shows significant differences between pairs 1 and 2, 1 and 5, 2 and 3, 2 and 4, 2 and 5, 3 and 5, and 4 and 5 because these confidence intervals do not contain zero. These results are consistent with the manual calculations.

```
Tukey 99% Simultaneous Confidence Intervals
All Pairwise Comparisons among Levels of Temp. Setting

Individual confidence level = 99.87%

Temp. Setting = 1 subtracted from:
Temp.
Setting     Lower     Center      Upper  -----+-------+------+-------+-
2        -0.15481   -0.09600   -0.03719      (--*--)
3        -0.01681    0.04200    0.10081               (--*--)
4        -0.03681    0.02200    0.08081             (--*--)
5         0.04719    0.10600    0.16481                  (--*--)
                                         -----+-------+------+-------+-
                                           -0.15     0.00   0.15   0.30

Temp. Setting = 2 subtracted from:
Temp.
Setting     Lower     Center      Upper  -----+-------+------+-------+-
3         0.07919    0.13800    0.19681                  (--*--)
4         0.05919    0.11800    0.17681                (--*--)
5         0.14319    0.20200    0.26081                     (--*--)
                                         -----+-------+------+-------+-
                                           -0.15     0.00   0.15   0.30

Temp. Setting = 3 subtracted from:
Temp.
Setting     Lower     Center      Upper  -----+-------+------+-------+-
4        -0.07881   -0.02000    0.03881          (--*--)
5         0.00519    0.06400    0.12281               (--*--)
                                         -----+-------+------+-------+-
                                           -0.15     0.00   0.15   0.30

Temp. Setting = 4 subtracted from:
Temp.
Setting     Lower     Center      Upper  -----+-------+------+-------+-
5         0.02519    0.08400    0.14281                (--*--)
                                         -----+-------+------+-------+-
                                           -0.15     0.00   0.15   0.30
```

Tukey-Kramer Procedure: The Case of Unequal Sample Sizes

Tukey's HSD was modified by C. Y. Kramer in the mid-1950s to handle situations in which the sample sizes are unequal. The modified version of HSD is sometimes referred to as the **Tukey-Kramer procedure.** The formula for computing the significant differences with this procedure is similar to that for the equal sample sizes, with the exception that the mean square error is divided in half and weighted by the sum of the inverses of the sample sizes under the root sign.

TUKEY-KRAMER FORMULA

$$q_{\alpha,C,N-C}\sqrt{\frac{\text{MSE}}{2}\left(\frac{1}{n_r}+\frac{1}{n_s}\right)}$$

where

MSE = mean square error
n_r = sample size for rth sample
n_s = sample size for sth sample
$q_{\alpha,\,C,\,N-C}$ = critical value of the studentized range distribution from Table A.10

TABLE 11.7

Means and Sample Sizes for the Valves Produced by Four Operators

Operator	Sample Size	Mean
1	5	6.3180
2	8	6.2775
3	7	6.4886
4	4	6.2300

As an example of the application of the Tukey-Kramer procedure, consider the machine operator example in Section 11.2. A one-way ANOVA was used to test for any difference in the mean valve openings produced by four different machine operators. An overall F of 10.18 was computed, which was significant at $\alpha = .05$. Because the ANOVA hypothesis test is significant and the null hypothesis is rejected, this problem is a candidate for multiple comparisons. Because the sample sizes are not equal, Tukey's HSD cannot be used to determine which pairs are significantly different. However, the Tukey-Kramer procedure can be applied. Shown in Table 11.7 are the means and sample sizes for the valve openings for valves produced by the four different operators.

The mean square error for this problem, MSE, is shown in Table 11.3 as .007746. The four operators in the problem represent the four levels of the independent variable, machine operator. Thus, $C = 4$, $N = 24$, and $N - C = 20$. The value of alpha in the problem is .05. With this information, the value of q is obtained from Table A.10 as

$$q_{.05,4,20} = 3.96$$

The distance necessary for the difference in the means of two samples to be statistically significant must be computed by using the Tukey-Kramer procedure for each pair because the sample sizes differ. In this problem with $C = 4$, there are $C(C - 1)/2$ or six possible pairwise comparisons. The computations follow.

For operators 1 and 2,

$$3.96\sqrt{\frac{.007746}{2}\left(\frac{1}{5} + \frac{1}{8}\right)} = .1405$$

The difference between the means of operator 1 and operator 2 is

$$6.3180 - 6.2775 = .0405.$$

Because this result is less than the critical difference of .1405, there is no significant difference between the average valve openings of valves produced by machine operators 1 and 2.

Table 11.8 reports the critical differences for each of the six pairwise comparisons as computed by using the Tukey-Kramer procedure, along with the absolute value of the

TABLE 11.8

Results of Pairwise Comparisons for the Machine Operators Example Using the Tukey-Kramer Procedure

Pair	Critical Difference	Actual Difference
1 and 2	.1405	.0405
1 and 3	.1443	.1706*
1 and 4	.1653	.0880
2 and 3	.1275	.2111*
2 and 4	.1509	.0475
3 and 4	.1545	.2586*

*Significant at $\alpha = .05$.

TABLE 11.9

Minitab Multiple Comparisons in the Machine Operator Example Using the Tukey-Kramer Procedure

```
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons

Individual confidence level = 98.89%

Operator 1 subtracted from:

                Lower      Center     Upper
Operator 2   -0.18099    -0.04050   0.09999
Operator 3    0.02627     0.17057   0.31487
Operator 4   -0.25332    -0.08800   0.07732

             -------+---------+--------+---------+-

Operator 2              (----*-----)
Operator 3                     (----*-----)
Operator 4          (-----*------)
             -------+---------+--------+---------+-
                 -0.25       0.00      0.25      0.50

Operator 2 subtracted from:
                Lower    Center     Upper
Operator 3    0.08353   0.21107   0.33862
Operator 4   -0.19841  -0.04750   0.10341

             -------+---------+--------+---------+-
Operator 3                       (----*----)
Operator 4             (----*----)
             -------+---------+--------+---------+-
                 -0.25       0.00      0.25      0.50

Operator 3 subtracted from:

                Lower     Center     Upper
Operator 4   -0.41304   -0.25857  -0.10411

             -------+---------+--------+---------+-
Operator 4   (-----*----)
             -------+---------+--------+---------+-
                 -0.25       0.00      0.25      0.50
```

actual distances between the means. Any actual distance between means that is greater than the critical distance is significant. As shown in the table, the means of three pairs of samples, operators 1 and 3, operators 2 and 3, and operators 3 and 4 are significantly different.

Table 11.9 shows the Minitab output for this problem. Minitab uses the Tukey-Kramer procedure for unequal values of n. As before with the HSD test, Minitab produces a confidence interval for the differences in means for pairs of treatment levels. If the confidence interval includes zero, there is no significant difference in the pairs of means. If the signs over the interval are the same (zero is not in the interval), there is a significant difference in the means. Note that the signs over the intervals for pairs (1, 3), (2, 3) and (3, 4) are the same, indicating a significant difference in the means of those two pairs. This conclusion agrees with the results determined through the calculations reported in Table 11.8.

11.3 PROBLEMS

11.17 Suppose an ANOVA has been performed on a completely randomized design containing six treatment levels. The mean for group 3 is 15.85, and the sample size for group 3 is eight. The mean for group 6 is 17.21, and the sample size for group 6 is seven. MSE is .3352. The total number of observations is 46. Compute the significant difference for the means of these two groups by using the Tukey-Kramer procedure. Let $\alpha = .05$.

11.18 A completely randomized design has been analyzed by using a one-way ANOVA. There are four treatment groups in the design, and each sample size is six. MSE is equal to 2.389. Using $\alpha = .05$, compute Tukey's HSD for this ANOVA.

11.19 Using the results of problem 11.5, compute a critical value by using the Tukey-Kramer procedure for groups 1 and 2. Use $\alpha = .05$. Determine whether there is a significant difference between these two groups.

11.20 Use the Tukey-Kramer procedure to determine whether there is a significant difference between the means of groups 2 and 5 in problem 11.6. Let $\alpha = .01$.

11.21 Using the results from problem 11.7, compute a Tukey's HSD to determine whether there are any significant differences between group means. Let $\alpha = .01$.

11.22 Using problem 11.8, compute Tukey's HSD and determine whether there is a significant difference in means by using this methodology. Let $\alpha = .05$.

11.23 Use the Tukey-Kramer procedure to do multiple comparisons for problem 11.11. Let $\alpha = .01$. State which pairs of machines, if any, produce significantly different mean fills.

11.24 Use Tukey's HSD test to compute multiple comparisons for the data in problem 11.12. Let $\alpha = .01$. State which regions, if any, are significantly different from other regions in mean starting salary figures.

11.25 Using $\alpha = .05$, compute critical values using the Tukey-Kramer procedure for the pairwise groups in problem 11.13. Determine which pairs of groups are significantly different, if any.

11.26 Do multiple comparisons on the data in problem 11.14 using Tukey's HSD test and $\alpha = .05$. State which pairs of cities, if any, have significantly different mean costs.

11.27 Problem 11.16 analyzed the number of weekly hours worked per person at five different plants. An F value of 3.10 was obtained with a probability of .0266. Because the probability is less than .05, the null hypothesis is rejected at $\alpha = .05$. There is an overall difference in the mean weekly hours worked by plant. Which pairs of plants have significant differences in the means, if any? To answer this question, a Minitab computer analysis was done. The data follow. Study the output in light of problem 11.16 and discuss the results.

```
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons

Individual confidence level = 99.32%

Plant 1 subtracted from:

          Lower   Center    Upper  -----+--------+-------+--------+-
Plant 2  -17.910  -7.722    2.466      (----*-----)
Plant 3   -7.743   3.598   14.939            (----*-----)
Plant 4  -21.830  -8.665    4.499    (------*-------)
Plant 5  -12.721  -0.921   10.880         (-----*-----)
                                   -----+--------+-------+--------+-
Plant 2 subtracted from:              -15        0      15       30

          Lower   Center    Upper  -----+--------+-------+--------+-
Plant 3    0.180  11.320   22.460                (------*----)
Plant 4  -13.935  -0.944   12.048          (------*-------)
Plant 5   -4.807   6.801   18.409              (-----*------)
                                   -----+--------+-------+--------+-
Plant 3 subtracted from:              -15        0      15       30

          Lower   Center    Upper  -----+--------+-------+--------+-
Plant 4  -26.178 -12.263    1.651  (-------*--------)
Plant 5  -17.151  -4.519    8.113      (------*------)
                                   -----+--------+-------+--------+-
Plant 4 subtracted from:              -15        0      15       30
                                   -----+--------+-------+--------+-
          Lower   Center    Upper               (------*-------)
Plant 5   -6.547   7.745   22.036  -----+--------+-------+--------+-
                                      -15        0      15       30
```

STATISTICS IN BUSINESS TODAY

Does National Ideology Affect a Firm's Definition of Success?

One researcher, G. C. Lodge, proposed that companies pursue different performance goals based on the ideology of their home country. L. Thurow went further by suggesting that such national ideologies drive U.S. firms to be short-term profit maximizers, Japanese firms to be growth maximizers, and European firms to be a mix of the two.

Three other researchers, J. Katz, S. Werner, and L. Brouthers, decided to test these suggestions by studying 114 international banks from the United States, the European Union (EU), and Japan listed in the Global 1000. Specifically, there were 34 banks from the United States, 45 banks from the European Union, and 35 banks from Japan in the study. Financial and market data were gathered and averaged on each bank over a five-year period to limit the effect of single-year variations. All statistics were converted by Morgan Stanley Capital International to U.S. dollar denominations on the same day of each year to ensure consistency of measurement.

The banks were compared on general measures of success such as profitability, capitalization, growth, size, risk, and earnings distribution by specifically examining 11 measures. Eleven one-way analyses of variance designs were computed, one for each dependent variable. These included return on equity, return on assets, yield, capitalization, assets, market value, growth, Tobin's Q, price-to-earnings ratio, payout ratio, and risk. The independent variable in each ANOVA was country, with three levels: U.S., EU, and Japan.

In all 11 ANOVAs, there was a significant difference between banks in the three countries ($\alpha = .01$) supporting the theme of different financial success goals for different national cultures. Because of the overall significant difference attained in the ANOVAs, each analysis of variance was followed by a Duncan's multiple range test (multiple comparison) to determine which, if any, of the pairs were significantly different. These comparisons revealed that U.S. and EU banks maintained significantly higher levels than Japanese banks on return on equity, return on assets, and yield. This result underscores the notion that U.S. and EU banks have more of a short-term profit orientation than do Japanese banks. There was a significant difference in banks from each of the three countries on amount of capitalization. U.S. banks had the highest level of capitalization followed by EU banks and then Japanese banks. This result may reflect the cultural attitude about how much capital is needed to ensure a sound economy, with U.S. banks maintaining higher levels of capital.

The study found that Japanese banks had significantly higher levels on growth, Tobin's Q, and price-to-earnings ratio than did the other two national entities. This result confirms the hypothesis that Japanese firms are more interested in growth. In addition, Japanese banks had a significantly higher asset size and market value of equity than did U.S. banks. The researchers had hypothesized that EU banks would have a greater portfolio risk than that of U.S. or Japanese banks. They found that EU banks did have significantly higher risk and paid out significantly higher dividends than did either Japanese or U.S. banks.

Source: Adapted from Jeffrey P. Katz, Steve Werner, and Lance Brouthers, "Does Winning Mean the Same Thing Around the World? National Ideology and the Performance of Global Competitors," *Journal of Business Research*, vol. 44, no. 2 (February 1999), pp. 117–126.

11.4 THE RANDOMIZED BLOCK DESIGN

Video

A second research design is the **randomized block design**. The randomized block design is similar to the completely randomized design in that it focuses on one independent variable (treatment variable) of interest. However, the randomized block design also includes a second variable, referred to as a blocking variable, that can be used to control for confounding or concomitant variables.

Confounding variables, or **concomitant variables**, are *variables that are not being controlled by the researcher in the experiment but can have an effect on the outcome of the treatment being studied.* For example, Demonstration Problem 11.2 showed how a completely randomized design could be used to analyze the effects of temperature on the tensile strengths of metal. However, other variables not being controlled by the researcher in this experiment may affect the tensile strength of metal, such as humidity, raw materials, machine, and shift. One way to control for these variables is to include them in the experimental design. The randomized block design has the capability of adding one of these variables into the analysis as a blocking variable. A **blocking variable** is *a variable that the researcher wants to control but is not the treatment variable of interest.*

One of the first people to use the randomized block design was Sir Ronald A. Fisher. He applied the design to the field of agriculture, where he was interested in studying the growth patterns of varieties of seeds for a given type of plant. The seed variety was his independent variable. However, he realized that as he experimented on different plots of ground, the "block" of ground might make some difference in the experiment. Fisher designated several different plots of ground as blocks, which he controlled as a second variable. Each of the seed varieties was planted on each of the blocks. The main thrust of his study was to compare the seed varieties (independent variable). He merely wanted to control for the difference in plots of ground (blocking variable).

In Demonstration Problem 11.2, examples of blocking variables might be machine number (if several machines are used to make the metal), worker, shift, or day of the week. The researcher probably already knows that different workers or different machines will produce at least slightly different metal tensile strengths because of individual differences. However, designating the variable (machine or worker) as the blocking variable and computing a randomized block design affords the potential for a more powerful analysis. In other experiments, some other possible variables that might be used as blocking variables include sex of subject, age of subject, intelligence of subject, economic level of subject, brand, supplier, or vehicle.

A special case of the randomized block design is the repeated measures design. The **repeated measures design** is a randomized block design in which each block level is an individual item or person, and that person or item is measured across all treatments. Thus, where a block level in a randomized block design is night shift and items produced under different treatment levels on the night shift are measured, in a repeated measures design, a block level might be an individual machine or person; items produced by that person or machine are then randomly chosen across all treatments. Thus, a repeated measure of the person or machine is made across all treatments. This repeated measures design is an extension of the t test for dependent samples presented in Section 10.3.

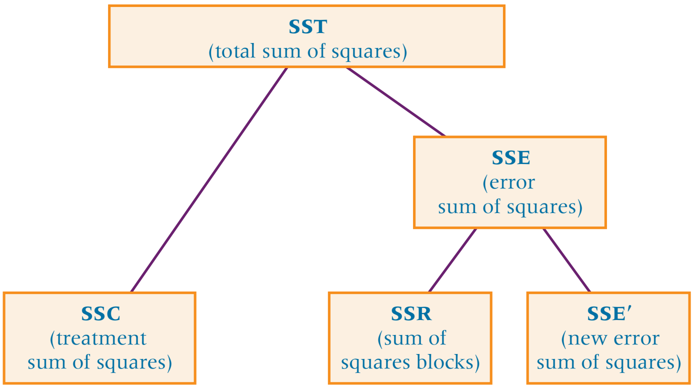
The sum of squares in a completely randomized design is

$$\text{SST} = \text{SSC} + \text{SSE}$$

In a randomized block design, the sum of squares is

$$\text{SST} = \text{SSC} + \text{SSR} + \text{SSE}$$

FIGURE 11.7
Partitioning the Total Sum of Squares in a Randomized Block Design



where

SST = sum of squares total
SSC = sum of squares columns (treatment)
SSR = sum of squares rows (blocking)
SSE = sum of squares error

SST and SSC are the same for a given analysis whether a completely randomized design or a randomized block design is used. For this reason, the SSR (blocking effects) comes out of the SSE; that is, some of the error variation in the completely randomized design is accounted for in the blocking effects of the randomized block design, as shown in Figure 11.7. By reducing the error term, it is possible that the value of F for treatment will increase (the denominator of the F value is decreased). However, if there is not sufficient difference between levels of the blocking variable, the use of a randomized block design can lead to a less powerful result than would a completely randomized design computed on the same problem. Thus, the researcher should seek out blocking variables that he or she believes are significant contributors to variation among measurements of the dependent variable. Figure 11.8 shows the layout of a randomized block design.

In each of the intersections of independent variable and blocking variable in Figure 11.8, one measurement is taken. In the randomized block design, one measurement is given for each treatment level under each blocking level.

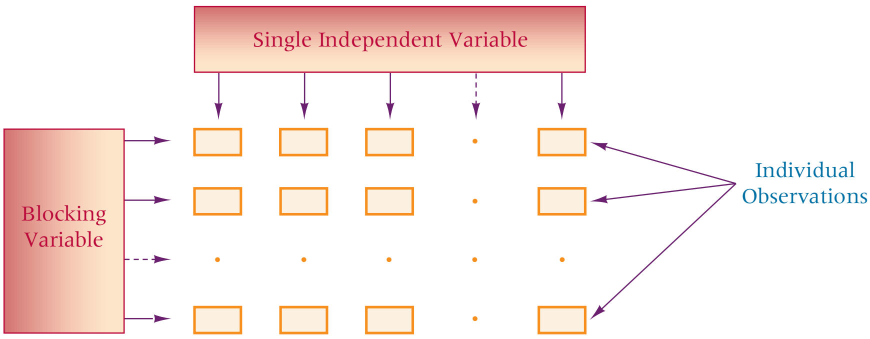
The null and alternate hypotheses for the treatment effects in the randomized block design are

H_0: $\mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3} = \ldots = \mu_{\cdot C}$
H_a: At least one of the treatment means is different from the others.

For the blocking effects, they are

H_0: $\mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \ldots = \mu_{R\cdot}$
H_a: At least one of the blocking means is different from the others.

FIGURE 11.8
A Randomized Block Design



Essentially, we are testing the null hypothesis that the population means of the treatment groups are equal. If the null hypothesis is rejected, at least one of the population means does not equal the others.

The formulas for computing a randomized block design follow.

FORMULAS FOR COMPUTING A RANDOMIZED BLOCK DESIGN

$$SSC = n\sum_{j=1}^{C}(\bar{x}_j - \bar{x})^2$$

$$SSR = C\sum_{i=1}^{n}(\bar{x}_i - \bar{x})^2$$

$$SSE = \sum_{i=1}^{n}\sum_{j=1}^{C}(x_{ij} - \bar{x}_j - \bar{x}_i + \bar{x})^2$$

$$SST = \sum_{i=1}^{n}\sum_{j=1}^{C}(x_{ij} - \bar{x})^2$$

where

i = block group (row)
j = treatment level (column)
C = number of treatment levels (columns)
n = number of observations in each treatment level (number of blocks or rows)
x_{ij} = individual observation
$\bar{x}_j$ = treatment (column) mean
$\bar{x}_i$ = block (row) mean
$\bar{x}$ = grand mean
N = total number of observations

$$df_C = C - 1$$

$$df_R = n - 1$$

$$df_E = (C - 1)(n - 1) = N - n - C + 1$$

$$MSC = \frac{SSC}{C - 1}$$

$$MSR = \frac{SSR}{n - 1}$$

$$MSE = \frac{SSE}{N - n - C + 1}$$

$$F_{\text{treatments}} = \frac{MSC}{MSE}$$

$$F_{\text{blocks}} = \frac{MSR}{MSE}$$

The observed F value for treatments computed using the randomized block design formula is tested by comparing it to a table F value, which is ascertained from Appendix A.7 by using α, df_C (treatment), and df_E (error). If the observed F value is greater than the table value, the null hypothesis is rejected for that alpha value. Such a result would indicate that not all population treatment means are equal. At this point, the business researcher has the option of computing multiple comparisons if the null hypothesis has been rejected.

Some researchers also compute an F value for blocks even though the main emphasis in the experiment is on the treatments. The observed F value for blocks is compared to a critical table F value determined from Appendix A.7 by using α, df_R (blocks), and df_E (error). If the F value for blocks is greater than the critical F value, the null hypothesis that all block population means are equal is rejected. This result tells the business researcher that including the blocking in the design was probably worthwhile and that a significant amount of variance was drawn off from the error term, thus increasing the power of the treatment test. In this text, we have omitted F_{blocks} from the normal presentation and problem solving. We leave the use of this F value to the discretion of the reader.

As an example of the application of the randomized block design, consider a tire company that developed a new tire. The company conducted tread-wear tests on the tire to determine whether there is a significant difference in tread wear if the average speed with which the automobile is driven varies. The company set up an experiment in which the independent variable was speed of automobile. There were three treatment levels: slow speed (car is driven 20 miles per hour), medium speed (car is driven 40 miles per hour), and high speed (car is driven 60 miles per hour). Company researchers realized that several possible variables could confound the study. One of these variables was supplier. The company uses five suppliers to provide a major component of the rubber from which the tires are made. To control for this variable experimentally, the researchers used supplier as a blocking variable. Fifteen tires were randomly selected for the study, three from each supplier. Each of the three was assigned to be tested under a different speed condition. The data are given here, along with treatment and block totals. These figures represent tire wear in units of 10,000 miles.

	Speed			Block Means
Supplier	**Slow**	**Medium**	**Fast**	$\bar{x}_i$
1	3.7	4.5	3.1	3.77
2	3.4	3.9	2.8	3.37
3	3.5	4.1	3.0	3.53
4	3.2	3.5	2.6	3.10
5	3.9	4.8	3.4	4.03
Treatment Means $\bar{x}_j$	3.54	4.16	2.98	$\bar{x} = 3.56$

To analyze this randomized block design using $\alpha = .01$, the computations are as follows.

$$C = 3$$
$$n = 5$$
$$N = 15$$

$$\begin{aligned} SSC &= n\sum_{j=1}^{C}(\bar{x}_j - \bar{x})^2 \\ &= 5[(3.54 - 3.56)^2 + (4.16 - 3.56)^2 + (2.98 - 3.56)^2] \\ &= 3.484 \end{aligned}$$

$$\begin{aligned} SSR &= C\sum_{i=1}^{n}(\bar{x}_i - \bar{x})^2 \\ &= 3[(3.77 - 3.56)^2 + (3.37 - 3.56)^2 + (3.53 - 3.56)^2 + (3.10 - 3.56)^2 \\ &\quad + (4.03 - 3.56)^2] \\ &= 1.549 \end{aligned}$$

$$\begin{aligned} SSE &= \sum_{i=1}^{n}\sum_{j=1}^{C}(x_{ij} - \bar{x}_j - \bar{x}_i + \bar{x})^2 \\ &= (3.7 - 3.54 - 3.77 + 3.56)^2 + (3.4 - 3.54 - 3.37 + 3.56)^2 \\ &\quad + \cdots + (2.6 - 2.98 - 3.10 + 3.56)^2 + (3.4 - 2.98 - 4.03 + 3.56)^2 \\ &= .143 \end{aligned}$$

$$\text{SST} = \sum_{i=1}^{n}\sum_{j=1}^{C}(x_{ij} - \bar{x})^2$$
$$= (3.7 - 3.56)^2 + (3.4 - 3.56)^2 + \cdots + (2.6 - 3.56)^2 + (3.4 - 3.56)^2$$
$$= 5.176$$

$$\text{MSC} = \frac{\text{SSC}}{C - 1} = \frac{3.484}{2} = 1.742$$
$$\text{MSR} = \frac{\text{SSR}}{n - 1} = \frac{1.549}{4} = .38725$$
$$\text{MSE} = \frac{\text{SSE}}{N - n - C + 1} = \frac{.143}{8} = .017875$$
$$F = \frac{\text{MSC}}{\text{MSE}} = \frac{1.742}{.017875} = 97.45$$

Source of Variation	SS	df	MS	F
Treatment	3.484	2	1.742	97.45
Block	1.549	4	.38725	
Error	.143	8	.017875	
Total	5.176	14		

For alpha of .01, the critical F value is

$$F_{.01,2,8} = 8.65$$

Because the observed value of F for treatment (97.45) is greater than this critical F value, the null hypothesis is rejected. At least one of the population means of the treatment levels is not the same as the others; that is, there is a significant difference in tread wear for cars driven at different speeds. If this problem had been set up as a completely randomized design, the SSR would have been a part of the SSE. The degrees of freedom for the blocking effects would have been combined with degrees of freedom of error. Thus, the value of SSE would have been $1.549 + .143 = 1.692$, and df_E would have been $4 + 8 = 12$. These would then have been used to recompute MSE = 1.692/12 = .141. The value of F for treatments would have been

$$F = \frac{\text{MSC}}{\text{MSE}} = \frac{1.742}{0.141} = 12.35$$

Thus, the F value for treatment with the blocking was 97.45 and *without* the blocking was 12.35. By using the random block design, a much larger observed F value was obtained.

Using the Computer to Analyze Randomized Block Designs

Both Minitab and Excel have the capability of analyzing a randomized block design. The computer output from each of these software packages for the tire tread wear example is displayed in Table 11.10. The randomized block design analysis is done on Minitab by using the same process as the two-way ANOVA, which will be discussed in Section 11.5.

The Minitab output includes F values and their associated p-values for both the treatment and the blocking effects. As with most standard ANOVA tables, the sum of squares, mean squares, and degrees of freedom for each source of variation are included.

Excel treats a randomized block design like a two-way ANOVA (Section 11.5) that has only one observation per cell. The Excel output includes sums, averages, and variances for each row and column. The Excel ANOVA table displays the observed F values for the treatment (columns) and the blocks (rows). An important inclusion in the Excel output is the p-value for each F, along with the critical (table) F values.

TABLE 11.10

Minitab and Excel Output for the Tread Wear Example

Minitab Output

Two-way ANOVA: Mileage versus Supplier, Speed

```
Source     DF       SS        MS      F      P
Supplier    4  1.54933  0.38733  21.72  0.000
Speed       2  3.48400  1.74200  97.68  0.000
Error       8  0.14267  0.01783
Total      14  5.17600
S = 0.1335  R-Sq = 97.24%  R-Sq(adj) = 95.18%
```

Excel Output

Anova: Two-Factor Without Replication

SUMMARY	*Count*	*Sum*	*Average*	*Variance*
Supplier 1	3	11.3	3.767	0.4933
Supplier 2	3	10.1	3.367	0.3033
Supplier 3	3	10.6	3.533	0.3033
Supplier 4	3	9.3	3.100	0.2100
Supplier 5	3	12.1	4.033	0.5033
Slow	5	17.7	3.54	0.073
Medium	5	20.8	4.16	0.258
Fast	5	14.9	2.98	0.092

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	1.549333	4	0.38733333	21.72	0.0002357	7.01
Columns	3.484000	2	1.742000	97.68	0.0000024	8.65
Error	0.142667	8	0.01783333			
Total	5.176000	14				

DEMONSTRATION PROBLEM 11.3

Suppose a national travel association studied the cost of premium unleaded gasoline in the United States during the summer of 2010. From experience, association directors believed there was a significant difference in the average cost of a gallon of premium gasoline among urban areas in different parts of the country. To test this belief, they placed random calls to gasoline stations in five different cities. In addition, the researchers realized that the brand of gasoline might make a difference. They were mostly interested in the differences between cities, so they made city their treatment variable. To control for the fact that pricing varies with brand, the researchers included brand as a blocking variable and selected six different brands to participate. The researchers randomly telephoned one gasoline station for each brand in each city, resulting in 30 measurements (five cities and six brands). Each station operator was asked to report the current cost of a gallon of premium unleaded gasoline at that station. The data are shown here. Test these data by using a randomized block design analysis to determine whether there is a significant difference in the average cost of premium unleaded gasoline by city. Let $\alpha = .01$.

Geographic Region

Brand	**Miami**	**Philadelphia**	**Minneapolis**	**San Antonio**	**Oakland**	$\bar{x}_i$
A	3.47	3.40	3.38	3.32	3.50	3.414
B	3.43	3.41	3.42	3.35	3.44	3.410
C	3.44	3.41	3.43	3.36	3.45	3.418
D	3.46	3.45	3.40	3.30	3.45	3.412
E	3.46	3.40	3.39	3.39	3.48	3.424
F	3.44	3.43	3.42	3.39	3.49	3.434
$\bar{x}_j$	3.450	3.4167	3.4067	3.3517	3.4683	$\bar{x} = 3.4187$

Solution

HYPOTHESIZE:

STEP 1. The hypotheses follow.

For treatments,

H_0: $\mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3} = \mu_{\cdot 4} = \mu_{\cdot 5}$

H_a: At least one of the treatment means is different from the others.

For blocks,

H_0: $\mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot} = \mu_{5\cdot} = \mu_{6\cdot}$

H_a: At least one of the blocking means is different from the others.

TEST:

STEP 2. The appropriate statistical test is the F test in the ANOVA for randomized block designs.

STEP 3. Let $\alpha = .01$.

STEP 4. There are four degrees of freedom for the treatment ($C - 1 = 5 - 1 = 4$), five degrees of freedom for the blocks ($n - 1 = 6 - 1 = 5$), and 20 degrees of freedom for error [$(C - 1)(n - 1) = (4)(5) = 20$]. Using these, $\alpha = .01$, and Table A.7, we find the critical F values.

$$F_{.01,4,20} = 4.43 \text{ for treatments}$$
$$F_{.01,5,20} = 4.10 \text{ for blocks}$$

The decision rule is to reject the null hypothesis for treatments if the observed F value for treatments is greater than 4.43 and to reject the null hypothesis for blocking effects if the observed F value for blocks is greater than 4.10.

STEP 5. The sample data including row and column means and the grand mean are given in the preceding table.

STEP 6.

$$\begin{aligned}\text{SSC} &= n\sum_{j=1}^{C}(\bar{x}_j - \bar{x})^2 \\ &= 6[(3.450 - 3.4187)^2 + (3.4167 - 3.4187)^2 + (3.4067 - 3.4187)^2 \\ &\quad + (3.3517 - 3.4187)^2 + (3.4683 - 3.4187)^2] \\ &= .04846\end{aligned}$$

$$\begin{aligned}\text{SSR} &= C\sum_{i=1}^{n}(\bar{x}_i - \bar{x})^2 \\ &= 5[(3.414 - 3.4187)^2 + (3.410 - 3.4187)^2 + (3.418 - 3.4187)^2 \\ &\quad + (3.412 - 3.4187)^2 + (3.424 - 3.4187)^2 + (3.434 - 3.4187)^2] \\ &= .00203\end{aligned}$$

$$\begin{aligned}\text{SSE} &= \sum_{i=1}^{n}\sum_{j=1}^{C}(x_{ij} - \bar{x}_j - \bar{x}_i + \bar{x})^2 \\ &= (3.47 - 3.450 - 3.414 + 3.4187)^2 + (3.43 - 3.450 - 3.410 + 3.4187)^2 + \ldots \\ &\quad + (3.48 - 3.4683 - 3.424 + 3.4187)^2 + (3.49 - 3.4683 - 3.434 + 3.4187)^2 = .01281\end{aligned}$$

$$\begin{aligned}\text{SST} &= \sum_{i=1}^{n}\sum_{j=1}^{C}(x_{ij} - \bar{x})^2 \\ &= (3.47 - 3.4187)^2 + (3.43 - 3.4187)^2 + \ldots + (3.48 - 3.4187)^2 + (3.49 - 3.4187)^2 \\ &= .06330\end{aligned}$$

$$\text{MSC} = \frac{\text{SSC}}{C - 1} = \frac{.04846}{4} = .01213$$

$$\text{MSR} = \frac{\text{SSR}}{n - 1} = \frac{.00203}{5} = .00041$$

$$\text{MSE} = \frac{\text{SSE}}{(C - 1)(n - 1)} = \frac{.01281}{20} = .00064$$

$$F = \frac{\text{MSC}}{\text{MSE}} = \frac{.01213}{.00064} = 18.95$$

Source of Variance	SS	df	MS	*F*
Treatment	.04846	4	.01213	18.95
Block	.00203	5	.00041	
Error	.01281	20	.00064	
Total	.06330	29		

ACTION:

STEP 7. Because $F_{\text{treat}} = 18.95 > F_{.01,4,20} = 4.43$, the null hypothesis is rejected for the treatment effects. There is a significant difference in the average price of a gallon of premium unleaded gasoline in various cities.

A glance at the MSR reveals that there appears to be relatively little blocking variance. The result of determining an *F* value for the blocking effects is

$$F = \frac{\text{MSR}}{\text{MSE}} = \frac{.00041}{.00064} = 0.64$$

The value of *F* for blocks is not significant at $\alpha = .01$ ($F_{.01,5,20} = 4.10$). This result indicates that the blocking portion of the experimental design did not contribute significantly to the analysis. If the blocking effects (SSR) are added back into SSE and the df_R are included with df_E, the MSE becomes .00059 instead of .00064. Using the value .00059 in the denominator for the treatment *F* increases the observed treatment *F* value to 20.56. Thus, including nonsignificant blocking effects in the original analysis caused a loss of power.

Shown here are the Minitab and Excel ANOVA table outputs for this problem.

Minitab Output

```
Two-way ANOVA: Gas Prices versus Brand, City

Source  DF         SS         MS      F      P
Brand    5  0.0020267  0.0004053   0.63  0.677
City     4  0.0485133  0.0121283  18.94  0.000
Error   20  0.0128067  0.0006403
Total   29  0.0633467
```

Excel Output

Anova: Two-Factor Without Replication

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	0.002027	5	0.000405	0.63	0.6768877	4.10
Columns	0.048513	4	0.012128	18.94	0.0000014	4.43
Error	0.012807	20	0.000640			
Total	0.063347	29				

BUSINESS IMPLICATIONS:

STEP 8. The fact that there is a significant difference in the price of gasoline in different parts of the country can be useful information to decision makers. For example, companies in the ground transportation business are greatly affected by increases in the cost of fuel. Knowledge of price differences in fuel can help these companies plan strategies and routes. Fuel price differences can sometimes be indications of cost-of-living differences or distribution problems, which can affect a company's relocation decision or cost-of-living increases given to employees who transfer to the higher-priced locations. Knowing that the price of gasoline varies around the country can generate interest among market researchers who might want to study why the differences are there and what drives them. This information can sometimes result in a better understanding of the marketplace.

11.4 PROBLEMS

11.28 Use ANOVA to analyze the data from the randomized block design given here. Let $\alpha = .05$. State the null and alternative hypotheses and determine whether the null hypothesis is rejected.

		Treatment Level			
		1	*2*	*3*	*4*
	1	23	26	24	24
	2	31	35	32	33
Block	3	27	29	26	27
	4	21	28	27	22
	5	18	25	27	20

11.29 The following data were gathered from a randomized block design. Use $\alpha = .01$ to test for a significant difference in the treatment levels. Establish the hypotheses and reach a conclusion about the null hypothesis.

		Treatment Level		
		1	*2*	*3*
	1	1.28	1.29	1.29
Block	2	1.40	1.36	1.35
	3	1.15	1.13	1.19
	4	1.22	1.18	1.24

11.30 A randomized block design has a treatment variable with six levels and a blocking variable with 10 blocks. Using this information and $\alpha = .05$, complete the following table and reach a conclusion about the null hypothesis.

Source of Variance	SS	df	MS	*F*
Treatment	2,477.53			
Blocks	3,180.48			
Error	11,661.38			
Total				

11.31 A randomized block design has a treatment variable with four levels and a blocking variable with seven blocks. Using this information and $\alpha = .01$, complete the following table and reach a conclusion about the null hypothesis.

Source of Variance	SS	df	MS	*F*
Treatment	199.48			
Blocks	265.24			
Error	306.59			
Total				

11.32 Safety in motels and hotels is a growing concern among travelers. Suppose a survey was conducted by the National Motel and Hotel Association to determine U.S. travelers' perception of safety in various motel chains. The association chose four different national chains from the economy lodging sector and randomly selected 10 people who had stayed overnight in a motel in each of the four chains in the past two years. Each selected traveler was asked to rate each motel chain on a scale from 0 to 100 to indicate how safe he or she felt at that motel. A score of 0 indicates completely unsafe and a score of 100 indicates perfectly safe. The scores follow. Test this randomized block design to determine whether there is a significant difference in the safety ratings of the four motels. Use $\alpha = .05$

Traveler	Motel 1	Motel 2	Motel 3	Motel 4
1	40	30	55	45
2	65	50	80	70
3	60	55	60	60
4	20	40	55	50
5	50	35	65	60
6	30	30	50	50
7	55	30	60	55
8	70	70	70	70
9	65	60	80	75
10	45	25	45	50

11.33 In recent years, the debate over the U.S. economy has been constant. The electorate seems somewhat divided as to whether the economy is in a recovery or not. Suppose a survey was undertaken to ascertain whether the perception of economic recovery differs according to political affiliation. People were selected for the survey from the Democratic Party, the Republican Party, and those classifying themselves as independents. A 25-point scale was developed in which respondents gave a score of 25 if they felt the economy was definitely in complete recovery, a 0 if the economy was definitely not in a recovery, and some value in between for more uncertain responses. To control for differences in socioeconomic class, a blocking variable was maintained using five different socioeconomic categories. The data are given here in the form of a randomized block design. Use $\alpha = .01$ to determine whether there is a significant difference in mean responses according to political affiliation.

	Political Affiliation		
Socioeconomic Class	Democrat	Republican	Independent
Upper	11	5	8
Upper middle	15	9	8
Middle	19	14	15
Lower middle	16	12	10
Lower	9	8	7

11.34 As part of a manufacturing process, a plastic container is supposed to be filled with 46 ounces of saltwater solution. The plant has three machines that fill the containers. Managers are concerned that the machines might not be filling the containers with the same amount of saltwater solution, so they set up a randomized block design to test this concern. A pool of five machine operators operates each of the three machines at different times. Company technicians randomly select five containers filled by each machine (one container for each of the five operators). The measurements are gathered and analyzed. The Minitab output from this analysis follows. What was the structure of the design? How many blocks were there? How many treatment classifications? Is there a statistical difference in the treatment means? Are the blocking effects significant? Discuss the implications of the output.

```
Two-way ANOVA: Measurement versus Machine, Operator
Source          DF        SS        MS        F        P
Machine          2     78.30     39.15     6.72     .019
Operator         4      5.09      1.27     0.22     .807
Error            8     46.66      5.83
Total           14    130.06
```

11.35 The comptroller of a company is interested in determining whether the average length of long-distance calls by managers varies according to type of telephone. A randomized block design experiment is set up in which a long-distance call by each of five managers is sampled for four different types of telephones: cellular, computer, regular, and cordless. The treatment is type of telephone and the blocks are the managers. The results of analysis by Excel are shown here. Discuss the results and any implications they might have for the company.

Anova: Two-Factor Without Replication

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Managers	11.3346	4	2.8336	12.74	0.00028	3.26
Phone Type	10.6043	3	3.5348	15.89	0.00018	3.49
Error	2.6696	12	0.2225			
Total	24.6085	19				

11.5 A FACTORIAL DESIGN (TWO-WAY ANOVA)

Video

Some experiments are designed so that *two or more treatments* (independent variables) *are explored simultaneously.* Such experimental designs are referred to as **factorial designs**. In factorial designs, *every level of each treatment is studied under the conditions of every level of all other treatments.* Factorial designs can be arranged such that three, four, or *n* treatments or independent variables are studied simultaneously in the same experiment. As an example, consider the valve opening data in Table 11.1. The mean valve opening for the 24 measurements is 6.34 centimeters. However, every valve but one in the sample measures something other than the mean. Why? Company management realizes that valves at this firm are made on different machines, by different operators, on different shifts, on different days, with raw materials from different suppliers. Business researchers who are interested in finding the sources of variation might decide to set up a factorial design that incorporates all five of these independent variables in one study. In this text, we explore the factorial designs with two treatments only.

Advantages of the Factorial Design

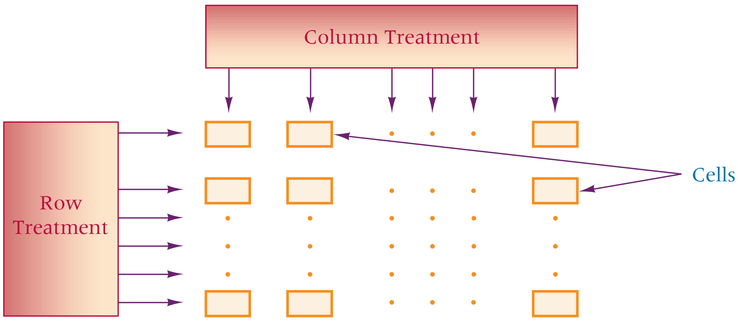
If two independent variables are analyzed by using a completely randomized design, the effects of each variable are explored separately (one per design). Thus, it takes two completely randomized designs to analyze the effects of the two independent variables. By using a factorial design, the business researcher can analyze both variables at the same time in one design, saving the time and effort of doing two different analyses and minimizing the experiment-wise error rate.

Some business researchers use the factorial design as a way to control confounding or concomitant variables in a study. By building variables into the design, the researcher attempts to control for the effects of multiple variables *in* the experiment. With the completely randomized design, the variables are studied in isolation. With the factorial design, there is potential for increased power over the completely randomized design because the additional effects of the second variable are removed from the error sum of squares.

The researcher can explore the possibility of interaction between the two treatment variables in a two-factor factorial design if multiple measurements are taken under every combination of levels of the two treatments. Interaction will be discussed later.

Factorial designs with two treatments are similar to randomized block designs. However, whereas randomized block designs focus on one treatment variable and control for a blocking effect, a two-treatment factorial design focuses on the effects of both variables. Because the randomized block design contains only one measure for each (treatment-block) combination, interaction cannot be analyzed in randomized block designs.

FIGURE 11.9
Two-Way Factorial Design



Factorial Designs with Two Treatments

The structure of a two-treatment factorial design is featured in Figure 11.9. Note that there are two independent variables (two treatments) and that there is an intersection of each level of each treatment. These intersections are referred to as *cells*. One treatment is arbitrarily designated as *row* treatment (forming the rows of the design) and the other treatment is designated as *column* treatment (forming the columns of the design). Although it is possible to analyze factorial designs with unequal numbers of items in the cells, the analysis of unequal cell designs is beyond the scope of this text. All factorial designs discussed here have cells of equal size.

Treatments (independent variables) of factorial designs must have at least two levels each. The simplest factorial design is a 2×2 factorial design, where each treatment has two levels. If such a factorial design were diagrammed in the manner of Figure 11.9, it would include two rows and two columns, forming four cells.

In this section, we study only factorial designs with $n > 1$ measurements for each combination of treatment levels (cells). This approach allows us to attempt to measure the interaction of the treatment variables. As with the completely randomized design and the randomized block design, a factorial design contains only one dependent variable.

Applications

Many applications of the factorial design are possible in business research. For example, the natural gas industry can design an experiment to study usage rates and how they are affected by temperature and precipitation. Theorizing that the outside temperature and type of precipitation make a difference in natural gas usage, industry researchers can gather usage measurements for a given community over a variety of temperature and precipitation conditions. At the same time, they can make an effort to determine whether certain types of precipitation, combined with certain temperature levels, affect usage rates differently than other combinations of temperature and precipitation (interaction effects).

Stock market analysts can select a company from an industry such as the construction industry and observe the behavior of its stock under different conditions. A factorial design can be set up by using volume of the stock market and prime interest rate as two independent variables. For volume of the market, business researchers can select some days when the volume is up from the day before, some days when the volume is down from the day before, and some other days when the volume is essentially the same as on the preceding day. These groups of days would constitute three levels of the independent variable, market volume. Business researchers can do the same thing with prime rate. Levels can be selected such that the prime rate is (1) up, (2) down, and (3) essentially the same. For the dependent variable, the researchers would measure how much the company's stock rises or falls on those randomly selected days (stock change). Using the factorial design, the business researcher can determine whether stock changes are different under various levels of market volume, whether stock changes are different under various levels of the prime interest rate, and whether stock changes react differently under various combinations of volume and prime rate (interaction effects).

Statistically Testing the Factorial Design

Analysis of variance is used to analyze data gathered from factorial designs. For factorial designs with two factors (independent variables), a **two-way analysis of variance (two-way ANOVA)** is used to test hypotheses statistically. The following hypotheses are tested by a two-way ANOVA.

Row effects:	H_0: Row means all are equal.
	H_a: At least one row mean is different from the others.
Column effects:	H_0: Column means are all equal.
	H_a: At least one column mean is different from the others.
Interaction effects:	H_0: The interaction effects are zero.
	H_a: An interaction effect is present.

Formulas for computing a two-way ANOVA are given in the following box. These formulas are computed in a manner similar to computations for the completely randomized design and the randomized block design. *F* values are determined for three effects:

1. Row effects
2. Column effects
3. Interaction effects

The row effects and the column effects are sometimes referred to as the main effects. Although *F* values are determined for these main effects, an *F* value is also computed for interaction effects. Using these observed *F* values, the researcher can make a decision about the null hypotheses for each effect.

Each of these observed *F* values is compared to a table *F* value. The table *F* value is determined by α, df_{num}, and df_{denom}.The degrees of freedom for the numerator (df_{num}) are determined by the effect being studied. If the observed *F* value is for columns, the degrees of freedom for the numerator are $C - 1$. If the observed *F* value is for rows, the degrees of freedom for the numerator are $R - 1$. If the observed *F* value is for interaction, the degrees of freedom for the numerator are $(R - 1) \cdot (C - 1)$. The number of degrees of freedom for the denominator of the table value for each of the three effects is the same, the error degrees of freedom, $RC(n - 1)$. The table *F* values (critical *F*) for a two-way ANOVA follow.

TABLE *F* VALUES FOR A TWO-WAY ANOVA

Row effects:	$F_{\alpha,R-1,RC(n-1)}$
Column effects:	$F_{\alpha,C-1,RC(n-1)}$
Interaction effects:	$F_{\alpha,(R-1)(C-1),RC(n-1)}$

FORMULAS FOR COMPUTING A TWO-WAY ANOVA

$$\text{SSR} = nC\sum_{i=1}^{R}(\bar{x}_i - \bar{x})^2$$

$$\text{SSC} = nR\sum_{j=1}^{C}(\bar{x}_j - \bar{x})^2$$

$$\text{SSI} = n\sum_{i=1}^{R}\sum_{j=1}^{C}(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$$

$$\text{SSE} = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij})^2$$

$$\text{SST} = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{n}(x_{ijk} - \bar{x})^2$$

$$\text{df}_R = R - 1$$
$$\text{df}_C = C - 1$$
$$\text{df}_I = (R - 1)(C - 1)$$
$$\text{df}_E = RC(n - 1)$$
$$\text{df}_T = N - 1$$

$$\text{MSR} = \frac{\text{SSR}}{R - 1}$$

$$\text{MSC} = \frac{\text{SSC}}{C - 1}$$

$$\text{MSI} = \frac{\text{SSI}}{(R - 1)(C - 1)}$$

$$\text{MSE} = \frac{\text{SSE}}{RC(n - 1)}$$

$$F_R = \frac{\text{MSR}}{\text{MSE}}$$

$$F_C = \frac{\text{MSC}}{\text{MSE}}$$

$$F_I = \frac{\text{MSI}}{\text{MSE}}$$

where

n = number of observations per cell
C = number of column treatments
R = number of row treatments
i = row treatment level
j = column treatment level
k = cell member
x_{ijk} = individual observation
$\bar{x}_{ij}$ = cell mean
$\bar{x}_i$ = row mean
$\bar{x}_j$ = column mean
$\bar{x}$ = grand mean

Interaction

As noted before, along with testing the effects of the two treatments in a factorial design, it is possible to test for the interaction effects of the two treatments whenever multiple measures are taken in each cell of the design. **Interaction** occurs *when the effects of one treatment vary according to the levels of treatment of the other effect.* For example, in a study examining the impact of temperature and humidity on a manufacturing process, it is possible that temperature and humidity will interact in such a way that the effect of temperature on the process varies with the humidity. Low temperatures might not be a significant manufacturing factor when humidity is low but might be a factor when humidity is high. Similarly, high temperatures might be a factor with low humidity but not with high humidity.

As another example, suppose a business researcher is studying the amount of red meat consumed by families per month and is examining economic class and religion as two independent variables. Class and religion might interact in such a way that with certain religions, economic class does not matter in the consumption of red meat, but with other religions, class does make a difference.

In terms of the factorial design, interaction occurs when the pattern of cell means in one row (going across columns) varies from the pattern of cell means in other rows. This

FIGURE 11.10

A 2 × 3 Factorial Design with Interaction

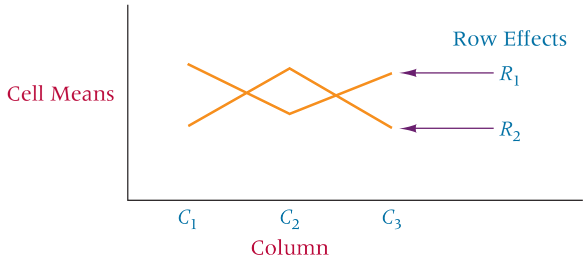
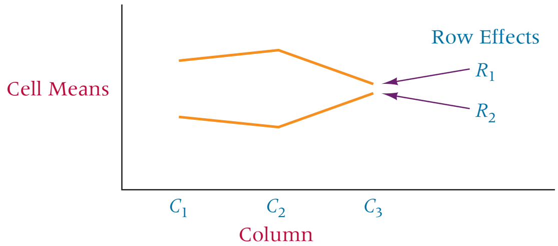


FIGURE 11.11

A 2 × 3 Factorial Design with Some Interaction



variation indicates that the differences in column effects depend on which row is being examined. Hence, an interaction of the rows and columns occurs. The same thing can happen when the pattern of cell means within a column is different from the pattern of cell means in other columns.

Interaction can be depicted graphically by plotting the cell means within each row (and can also be done by plotting the cell means within each column). The means within each row (or column) are then connected by a line. If the broken lines for the rows (or columns) are parallel, no interaction is indicated.

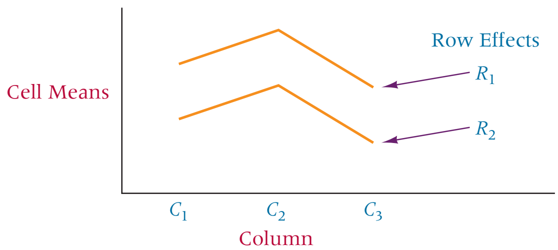
Figure 11.10 is a graph of the means for each cell in each row in a 2 × 3 (2 rows, 3 columns) factorial design with interaction. Note that the lines connecting the means in each row cross each other. In Figure 11.11 the lines converge, indicating the likely presence of some interaction. Figure 11.12 depicts a 2 × 3 factorial design with no interaction.

When the interaction effects are significant, the main effects (row and column) are confounded and should not be analyzed in the usual manner. In this case, it is not possible to state unequivocally that the row effects or the column effects are significantly different because the difference in means of one main effect varies according to the level of the other main effect (interaction is present). Some specific procedures are recommended for examining main effects when significant interaction is present. However, these techniques are beyond the scope of material presented here. Hence, in this text, whenever interaction effects are present (F_{inter} is significant), the researcher should *not* attempt to interpret the main effects (F_{row} and F_{col}).

As an example of a factorial design, consider the fact that at the end of a financially successful fiscal year, CEOs often must decide whether to award a dividend to stockholders

FIGURE 11.12

A 2 × 3 Factorial Design with No Interaction



or to make a company investment. One factor in this decision would seem to be whether attractive investment opportunities are available.* To determine whether this factor is important, business researchers randomly select 24 CEOs and ask them to rate how important "availability of profitable investment opportunities" is in deciding whether to pay dividends or invest. The CEOs are requested to respond to this item on a scale from 0 to 4, where 0 = no importance, 1 = slight importance, 2 = moderate importance, 3 = great importance, and 4 = maximum importance. The 0–4 response is the dependent variable in the experimental design.

The business researchers are concerned that where the company's stock is traded (New York Stock Exchange, American Stock Exchange, and over-the-counter) might make a difference in the CEOs' response to the question. In addition, the business researchers believe that how stockholders are informed of dividends (annual reports versus presentations) might affect the outcome of the experiment. Thus, a two-way ANOVA is set up with "where the company's stock is traded" and "how stockholders are informed of dividends" as the two independent variables. The variable "how stockholders are informed of dividends" has two treatment levels, or classifications.

1. Annual/quarterly reports
2. Presentations to analysts

The variable "where company stock is traded" has three treatment levels, or classifications.

1. New York Stock Exchange
2. American Stock Exchange
3. Over-the-counter

This factorial design is a 2 × 3 design (2 rows, 3 columns) with four measurements (ratings) per cell, as shown in the following table.

How Stockholders Are Informed of Dividends		*Where Company Stock Is Traded*			
		New York Stock Exchange	*American Stock Exchange*	*Over the Counter*	$\bar{X}_i =$
	Annual Quarterly Reports	2 1 2 1 $\bar{X}_{11} = 1.5$	2 3 3 2 $\bar{X}_{12} = 2.5$	4 3 4 3 $\bar{X}_{13} = 3.5$	2.5
	Presentations to Analysts	2 3 1 2 $\bar{X}_{21} = 2.0$	3 3 2 4 $\bar{X}_{22} = 3.0$	4 4 3 4 $\bar{X}_{23} = 3.75$	2.9167
	$\bar{X}_j =$	1.75	2.75	3.625	$\bar{X} = 2.7083$

These data are analyzed by using a two-way analysis of variance and $\alpha = .05$.

$$\text{SSR} = nC\sum_{i=1}^{R}(\bar{x}_i - \bar{x})^2$$

$$= 4(3)[(2.5 - 2.7083)^2 + (2.9167 - 2.7083)^2] = 1.0418$$

*Adapted from H. Kent Baker, "Why Companies Pay No Dividends," *Akron Business and Economic Review*, vol. 20 (Summer 1989), pp. 48–61.

$$\text{SSC} = nR\sum_{j=1}^{C}(\bar{x}_j - \bar{x})^2$$
$$= 4(2)[(1.75 - 2.7083)^2 + (2.75 - 2.7083)^2 + (3.625 - 2.7083)^2] = 14.0833$$

$$\text{SSI} = n\sum_{i=1}^{R}\sum_{j=1}^{C}(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$$
$$= 4[(1.5 - 2.5 - 1.75 + 2.7083)^2 + (2.5 - 2.5 - 2.75 + 2.7083)^2$$
$$+ (3.5 - 2.5 - 3.625 + 2.7083)^2 + (2.0 - 2.9167 - 1.75 + 2.7083)^2$$
$$+ (3.0 - 2.9167 - 2.75 + 2.7083)^2 + (3.75 - 2.9167 - 3.625 + 2.7083)^2] = .0833$$

$$\text{SSE} = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij})^2$$
$$= (2 - 1.5)^2 + (1 - 1.5)^2 + \cdots + (3 - 3.75)^2 + (4 - 3.75)^2 = 7.7500$$

$$\text{SST} = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{n}(x_{ijk} - \bar{x})^2$$
$$= (2 - 2.7083)^2 + (1 - 2.7083)^2 + \cdots + (3 - 2.7083)^2 + (4 - 2.7083)^2 = 22.9583$$

$$\text{MSR} = \frac{\text{SSR}}{R - 1} = \frac{1.0418}{1} = 1.0418$$

$$\text{MSC} = \frac{\text{SSC}}{C - 1} = \frac{14.0833}{2} = 7.0417$$

$$\text{MSI} = \frac{\text{SSI}}{(R - 1)(C - 1)} = \frac{.0833}{2} = .0417$$

$$\text{MSE} = \frac{\text{SSE}}{RC(n - 1)} = \frac{7.7500}{18} = .4306$$

$$F_R = \frac{\text{MSR}}{\text{MSE}} = \frac{1.0418}{.4306} = 2.42$$

$$F_C = \frac{\text{MSC}}{\text{MSE}} = \frac{7.0417}{.4306} = 16.35$$

$$F_I = \frac{\text{MSI}}{\text{MSE}} = \frac{.0417}{.4306} = 0.10$$

Source of Variation	SS	df	MS	F
Row	1.0418	1	1.0418	2.42
Column	14.0833	2	7.0417	16.35*
Interaction	.0833	2	.0417	0.10
Error	7.7500	18	.4306	
Total	22.9583	23		

*Denotes significance at $\alpha = .05$.

The critical F value for the interaction effects at $\alpha = .05$ is

$$F_{.05,2,18} = 3.55.$$

The observed F value for interaction effects is 0.10. Because this value is less than the critical table value (3.55), no significant interaction effects are evident. Because no significant interaction effects are present, it is possible to examine the main effects.

The critical F value of the row effects at $\alpha = .05$ is $F_{.05,1,18} = 4.41$. The observed F value of 2.42 is less than the table value. Hence, no significant row effects are present.

The critical F value of the column effects at $\alpha = .05$ is $F_{.05,2,18} = 3.55$. This value is coincidently the same as the critical table value for interaction because in this problem the degrees of freedom are the same for interaction and column effects. The observed F value for columns (16.35) is greater than this critical value. Hence, a significant difference in row effects is evident at $\alpha = .05$.

A significant difference is noted in the CEOs' mean ratings of the item "availability of profitable investment opportunities" according to where the company's stock is traded. A cursory examination of the means for the three levels of the column effects (where stock is traded) reveals that the lowest mean was from CEOs whose company traded stock on the New York Stock Exchange. The highest mean rating was from CEOs whose company traded

FIGURE 11.13

Minitab and Excel Output for the CEO Dividend Problem

Minitab Output

```
Two-way ANOVA: Rating versus How Reported, Where Traded

Source          DF       SS        MS       F       P
How Reported     1    1.0417   1.04167    2.42   0.137
Where Traded     2   14.0833   7.04167   16.35   0.000
Interaction      2    0.0833   0.04167    0.10   0.908
Error           18    7.7500   0.43056
Total           23   22.9583

S = 0.6562    R-Sq = 66.24%    R-Sq(adj) = 56.87%

                          Individual 95% CIs For Mean Based on Pooled StDev
How  Reported   Mean      +------ + ---- + ---- + ----
1            2.50000      (---------*--------)
2            2.91667                (-------*-------)
                          +------ + ---- + ---- + ----
                        2.10    2.45    2.80    3.15

                          Individual 95% CIs For Mean Based on Pooled StDev
Where   Traded  Mean      - + ---- + ---- + ---- + ----
1              1.750      (----*----)
2              2.750                (----*----)
3              3.625                          (----*----)
                          - + ---- + ---- + ---- + ----
                           1.60    2.40    3.20    4.00
```

Excel Output

ANOVA: Two-Factor With Replication

SUMMARY	NYSE	ASE	OTC	Total
A.Q. Reports				
Count	4	4	4	12
Sum	6	10	14	30
Average	1.5	2.5	3.5	2.5
Variance	0.33333	0.33333	0.33333	1
Pres. to Analysts				
Count	4	4	4	12
Sum	8	12	15	35
Average	2	3	3.75	2.91667
Variance	0.66667	0.66667	0.25	0.99242
Total				
Count	8	8	8	
Sum	14	22	29	
Average	1.75	2.75	3.625	
Variance	0.5	0.5	0.26786	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	1.04167	1	1.04167	2.42	0.137251	4.41
Columns	14.08333	2	7.04167	16.35	0.000089	3.55
Interaction	0.08333	2	0.04167	0.10	0.90823	3.55
Within	7.75	18	0.43056			
Total	22.95833	23				

stock over-the-counter. Using multiple comparison techniques, the business researchers can statistically test for differences in the means of these three groups.

Because the sample sizes within each column are equal, Tukey's HSD test can be used to compute multiple comparisons. The value of MSE is .431 for this problem. In testing the column means with Tukey's HSD test, the value of n is the number of items in a column, which is eight. The number of treatments is $C = 3$ for columns and $N - C = 24 - 3 = 21$.

With these two values and $\alpha = .05$, a value for q can be determined from Table A.10:

$$q_{.05,3,21} = 3.58$$

From these values, the honestly significant difference can be computed:

$$\text{HSD} = q\sqrt{\frac{\text{MSE}}{n}} = 3.58\sqrt{\frac{.431}{8}} = .831$$

The mean ratings for the three columns are

$$\bar{x}_1 = 1.75,\ \bar{x}_2 = 2.75,\ \bar{x}_3 = 3.625$$

The absolute value of differences between means are as follows:

$$|\bar{x}_1 - \bar{x}_2| = |1.75 - 2.75| = 1.00$$
$$|\bar{x}_1 - \bar{x}_3| = |1.75 - 3.625| = 1.875$$
$$|\bar{x}_2 - \bar{x}_3| = |2.75 - 3.625| = .875$$

All three differences are greater than .831 and are therefore significantly different at $\alpha = .05$ by the HSD test. Where a company's stock is traded makes a difference in the way a CEO responds to the question.

Using a Computer to Do a Two-Way ANOVA

A two-way ANOVA can be computed by using either Minitab or Excel. Figure 11.13 displays the Minitab and Excel output for the CEO example. The Minitab output contains an ANOVA table with each of the three F values and their associated p-values. In addition, there are individual 95% confidence intervals for means of both row and column effects. These intervals give the researcher a visual idea of differences between means. A more formal test of multiple comparisons of the column means is done with Minitab by using Tukey's HSD test. This output is displayed in Figure 11.14. Observe that in all three comparisons the signs on each end of the particular confidence interval are the same (and thus zero is not included); hence there is a significant difference in the means in each of the three pairs.

FIGURE 11.14
Tukey's Pairwise Comparisons for Column Means

```
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons among levels of where Traded

Individual confidence level = 98.00%

Where Traded = 1 subtracted from:

Where
Traded      Lower   Center   Upper  ----+------+-------+-------+----
2          0.1818  1.0000  1.8182             (----*-----)
3          1.0568  1.8750  2.6932                   (-----*-----)
                                    ----+------+-------+-------+----
                                      -1.2    0.0     1.2     2.4

Where Traded = 2 subtracted from:

Where
Traded      Lower   Center   Upper  ----+------+-------+-------+----
3          0.0568  0.8750  1.6932             (-----*-----)
                                    ----+------+-------+-------+----
                                      -1.2    0.0     1.2     2.4
```

The Excel output for two-way ANOVA with replications on the CEO dividend example is included in Figure 11.13. The Excel output contains cell, column, and row means along with observed *F* values for rows (sample), columns, and interaction. The Excel output also contains *p*-values and critical *F* values for each of these *F*'s. Note that the output here is virtually identical to the findings obtained by the manual calculations.

DEMONSTRATION PROBLEM 11.4

Some theorists believe that training warehouse workers can reduce absenteeism.* Suppose an experimental design is structured to test this belief. Warehouses in which training sessions have been held for workers are selected for the study. The four types of warehouses are (1) general merchandise, (2) commodity, (3) bulk storage, and (4) cold storage. The training sessions are differentiated by length. Researchers identify three levels of training sessions according to the length of sessions: (1) 1–20 days, (2) 21–50 days, and (3) more than 50 days. Three warehouse workers are selected randomly for each particular combination of type of warehouse and session length. The workers are monitored for the next year to determine how many days they are absent. The resulting data are in the following 4 × 3 design (4 rows, 3 columns) structure. Using this information, calculate a two-way ANOVA to determine whether there are any significant differences in effects. Use $\alpha = .05$.

Solution

HYPOTHESIZE:

STEP 1. The following hypotheses are being tested.

For row effects:

H_0: $\mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$
H_a: At least one of the row means is different from the others.

For column effects:

H_0: $\mu_{\cdot 1} = \mu_{\cdot 2} = \mu_{\cdot 3}$
H_a: At least one of the column means is different from the others.

For interaction effects:

H_0: The interaction effects are zero.
H_a: There is an interaction effect.

TEST:

STEP 2. The two-way ANOVA with the *F* test is the appropriate statistical test.

STEP 3. $\alpha = .05$

STEP 4.

$$df_{rows} = 4 - 1 = 3$$
$$df_{columns} = 3 - 1 = 2$$
$$df_{interaction} = (3)(2) = 6$$
$$df_{error} = (4)(3)(2) = 24$$

For row effects, $F_{.05,3,24} = 3.01$; for column effects, $F_{.05,2,24} = 3.40$; and for interaction effects, $F_{.05,6,24} = 2.51$. For each of these effects, if any observed *F* value is greater than its associated critical *F* value, the respective null hypothesis will be rejected.

*Adapted from Paul R. Murphy and Richard F. Poist, "Managing the Human Side of Public Warehousing: An Overview of Modern Practices," *Transportation Journal*, vol. 31 (Spring 1992), pp. 54–63.

STEP 5.

Types of Warehouses		Length of Training Session (Days) 1–20	21–50	More than 50	$\bar{X}_r$
	General Merchandise	3 4.5 4	2 2.5 2	2.5 1 1.5	2.5556
	Commodity	5 4.5 4	1 3 2.5	0 1.5 2	2.6111
	Bulk Storage	2.5 3 3.5	1 3 1.5	3.5 3.5 4	2.8333
	Cold Storage	2 2 3	5 4.5 2.5	4 4.5 5	3.6111
	$\bar{X}_c$	3.4167	2.5417	2.75	

$\bar{X} = 2.9028$

STEP 6. The Minitab and Excel (ANOVA table only) output for this problem follows

Minitab Output

```
Two-way ANOVA: Absences versus Type of Ware, Length
   Source         DF        SS         MS       F       P
   Type of ware    3    6.4097    2.13657    3.46   0.032
   Length          2    5.0139    2.50694    4.06   0.030
   Interaction     6   33.1528    5.52546    8.94   0.000
   Error          24   14.8333    0.61806
   Total          35   59.4097
```

Excel Output

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Types of Warehouses	6.40972	3	2.136574	3.46	0.032205	3.01
Length of Training Session	5.01389	2	2.506944	4.06	0.030372	3.40
Interaction	33.15278	6	5.525463	8.94	0.000035	2.51
Within	14.83333	24	0.618056			
Total	59.40972	35				

ACTION:

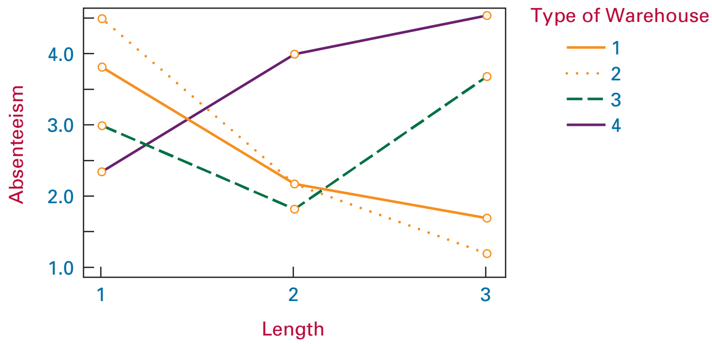
STEP 7. Looking at the source of variation table, we must first examine the interaction effects. The observed *F* value for interaction is 8.94 for both Excel and Minitab. The observed *F* value for interaction is greater than the critical *F* value. The interaction effects are statistically significant at $\alpha = .05$. The *p*-value for interaction shown in Excel is .000035. The interaction effects are significant at $\alpha = .0001$. The business researcher should not bother to examine the main effects because the significant interaction confounds the main effects.

BUSINESS IMPLICATIONS:

STEP 8. The significant interaction effects indicate that certain warehouse types in combination with certain lengths of training session result in different absenteeism rates than do other combinations of levels for these two variables. Using the cell means shown here, we can depict the interactions graphically.

		Length of Training Session (Days)		
		1–20	21–50	More than 50
Type of Warehouse	General Merchandise	3.8	2.2	1.7
	Commodity	4.5	2.2	1.2
	Bulk Storage	3.0	1.8	3.7
	Cold Storage	2.3	4.0	4.5

Minitab produces the following graph of the interaction.



Note the intersecting and crossing lines, which indicate interaction. Under the short-length training sessions, 1, cold-storage workers had the lowest rate of absenteeism and workers at commodity warehouses had the highest. However, for medium-length sessions, 2, cold-storage workers had the highest rate of absenteeism and bulk-storage had the lowest. For the longest training sessions, 3, commodity warehouse workers had the lowest rate of absenteeism, even though these workers had the highest rate of absenteeism for short-length sessions. Thus, the rate of absenteeism for workers at a particular type of warehouse depended on length of session. There was an interaction between type of warehouse and length of session. This graph could be constructed with the row levels along the bottom axis instead of column levels.

11.5 PROBLEMS

11.36 Describe the following factorial design. How many independent and dependent variables are there? How many levels are there for each treatment? If the data were known, could interaction be determined from this design? Compute all degrees of freedom. Each data value is represented by an *x*.

	Variable 1			
Variable 2	x_{111}	x_{121}	x_{131}	x_{141}
	x_{112}	x_{122}	x_{132}	x_{142}
	x_{113}	x_{123}	x_{133}	x_{143}
	x_{211}	x_{221}	x_{231}	x_{241}
	x_{212}	x_{222}	x_{232}	x_{242}
	x_{213}	x_{223}	x_{233}	x_{243}

11.37 Describe the following factorial design. How many independent and dependent variables are there? How many levels are there for each treatment? If the data were known, could interaction be determined from this design? Compute all degrees of freedom. Each data value is represented by an x.

	Variable 1		
	x_{111}	x_{121}	x_{131}
	x_{112}	x_{122}	x_{132}
	x_{211}	x_{221}	x_{231}
	x_{212}	x_{222}	x_{232}
Variable 2	x_{311}	x_{321}	x_{331}
	x_{312}	x_{322}	x_{332}
	x_{411}	x_{421}	x_{431}
	x_{412}	x_{422}	x_{432}

11.38 Complete the following two-way ANOVA table. Determine the critical table F values and reach conclusions about the hypotheses for effects. Let $\alpha = .05$.

Source of Variance	SS	df	MS	F
Row	126.98	3		
Column	37.49	4		
Interaction	380.82			
Error	733.65	60		
Total				

11.39 Complete the following two-way ANOVA table. Determine the critical table F values and reach conclusions about the hypotheses for effects. Let $\alpha = .05$.

Source of Variance	SS	df	MS	F
Row	1.047	1		
Column	3.844	3		
Interaction	0.773			
Error	____	__		
Total	12.632	23		

11.40 The data gathered from a two-way factorial design follow. Use the two-way ANOVA to analyze these data. Let $\alpha = .01$.

		Treatment 1		
		A	*B*	*C*
	A	23	21	20
		25	21	22
Treatment 2				
	B	27	24	26
		28	27	27

11.41 Suppose the following data have been gathered from a study with a two-way factorial design. Use $\alpha = .05$ and a two-way ANOVA to analyze the data. State your conclusions.

		Treatment 2															
		A				*B*				*C*				D			
Treatment 1	**A**	1.2	1.3	1.3	1.5	2.2	2.1	2.0	2.3	1.7	1.8	1.7	1.6	2.4	2.3	2.5	2.4
	B	1.9	1.6	1.7	2.0	2.7	2.5	2.8	2.8	1.9	2.2	1.9	2.0	2.8	2.6	2.4	2.8

11.42 Children are generally believed to have considerable influence over their parents in the purchase of certain items, particularly food and beverage items. To study this notion further, a study is conducted in which parents are asked to report how many food and beverage items purchased by the family per week are purchased mainly because of the influence of their children. Because the age of the child may have an effect on the study, parents are asked to focus on one particular child in the family for the week, and to report the age of the child. Four age categories are selected for the children: 4–5 years, 6–7 years, 8–9 years, and 10–12 years. Also, because the number of children in the family might make a difference, three different sizes of families are chosen for the study: families with one child, families with two children, and families with three or more children. Suppose the following data represent the reported number of child-influenced buying incidents per week. Use the data to compute a two-way ANOVA. Let $\alpha = .05$.

		Number of Children in Family		
		1	**2**	**3 or more**
Age of Child (years)	*4–5*	2	1	1
		4	2	1
	6–7	5	3	2
		4	1	1
	8–9	8	4	2
		6	5	3
	10–12	7	3	4
		8	5	3

11.43 A shoe retailer conducted a study to determine whether there is a difference in the number of pairs of shoes sold per day by stores according to the number of competitors within a 1-mile radius and the location of the store. The company researchers selected three types of stores for consideration in the study: stand-alone suburban stores, mall stores, and downtown stores. These stores vary in the numbers of competing stores within a 1-mile radius, which have been reduced to four categories: 0 competitors, 1 competitor, 2 competitors, and 3 or more competitors. Suppose the following data represent the number of pairs of shoes sold per day for each of these types of stores with the given number of competitors. Use $\alpha = .05$ and a two-way ANOVA to analyze the data.

		Number of Competitors			
		0	**1**	**2**	**3 or more**
Store Location	*Stand-Alone*	41	38	59	47
		30	31	48	40
		45	39	51	39
	Mall	25	29	44	43
		31	35	48	42
		22	30	50	53
	Downtown	18	22	29	24
		29	17	28	27
		33	25	26	32

11.44 Study the following analysis of variance table that was produced by using Minitab. Describe the design (number of treatments, sample sizes, etc.). Are there any significant effects? Discuss the output.

```
Two-way ANOVA:DV versus RowEffect, ColEffect
Source          DF         SS        MS        F        P
RowEffect        2      92.31    46.156    13.23    0.000
ColEffect        4     998.80   249.700    71.57    0.000
Interaction      8     442.13    55.267    15.84    0.000
Error           30     104.67     3.489
Total           44    1637.91
```

11.45 Consider the valve opening data displayed in Table 11.1. Suppose the data represent valves produced on four different machines on three different shifts and that the quality controllers want to know whether there is any difference in the mean measurements of valve openings by shift or by machine. The data are given here, organized by machine and shift. In addition, Excel has been used to analyze the data with a two-way ANOVA. What are the hypotheses for this problem? Study the output in terms of significant differences. Discuss the results obtained. What conclusions might the quality controllers reach from this analysis?

		Valve Openings (cm)		
		Shift 1	Shift 2	Shift 3
	1	6.56	6.38	6.29
		6.40	6.19	6.23
	2	6.54	6.26	6.19
Machine		6.34	6.23	6.33
	3	6.58	6.22	6.26
		6.44	6.27	6.31
	4	6.36	6.29	6.21
		6.50	6.19	6.58

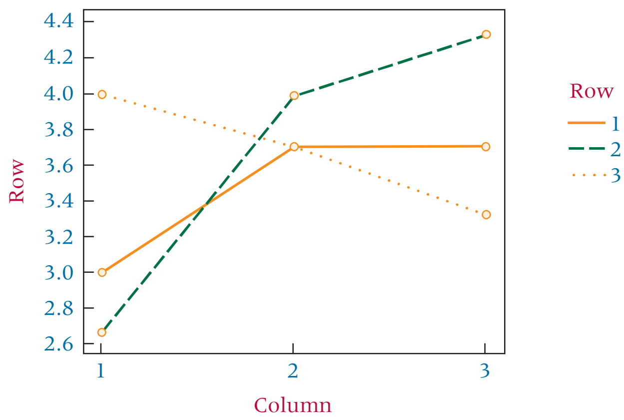
ANOVA: Two-Factor with Replication

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.00538	3	0.00179	0.14	0.9368	3.49
Columns	0.19731	2	0.09865	7.47	0.0078	3.89
Interaction	0.03036	6	0.00506	0.38	0.8760	3.00
Within	0.15845	12	0.01320			
Total	0.39150	23				

11.46 Finish the computations in the Minitab ANOVA table shown below and on the next page and determine the critical table F values. Interpret the analysis. Examine the associated Minitab graph and interpret the results. Discuss this problem, including the structure of the design, the sample sizes, and decisions about the hypotheses.

```
Two-way ANOVA: depvar versus row, column
Source          DF        SS         MS
Row              2     0.296      0.148
Column           2     1.852      0.926
Interaction      4     4.370      1.093
Error           18    14.000      0.778
Total           26    20.519
```



Decision Dilemma Solved

Is there a difference in the job satisfaction ratings of self-initiated expatriates by industry? The data presented in the Decision Dilemma to study this question represent responses on a seven-point Likert scale by 24 self-initiated expatriates from five different industries. The Likert scale score is the dependent variable. There is only one independent variable, industry, with five classification levels: IT, finance, education, healthcare, and consulting. If a series of t tests for the difference of two means from independent populations were used to analyze these data, there would be ${}_5C_2$ or 10 different t tests on this one problem. Using $\alpha = .05$ for each test, the probability of at least one of the 10 tests being significant by chance when the null hypothesis is true is $1 - (.95)^{10} = .4013$. That is, performing 10 t tests on this problem could result in an overall probability of committing a Type I error equal to .4013, not .05. In order to control the overall error, a one-way ANOVA is used on this completely randomized design to analyze these data by producing a single value of F and holding the probability of committing a Type I error at .05. Both Excel and Minitab have the capability of analyzing these data, and Minitab output for this problem is shown below.

```
One-way ANOVA: IT, Finance, Education, Healthcare, Consulting
Source      DF       SS       MS       F      P
Factor       4   43.175   10.794   15.25  0.000
Error       19   13.450    0.708
Total       23   56.625
S = 0.8414    R-Sq = 76.25%  R-Sq(adj) = 71.25%

                                  Individual 95% CIs For Mean Based on
                                  Pooled StDev
Level          N      Mean    StDev  -----+-------+------+-------+-
IT             4    5.7500   0.9574                  (----*----)
Finance        5    4.0000   0.7071           (----*----)
Education      6    2.5000   0.5477 (---*---)
Healthcare     5    3.4000   1.1402     (--- *----)
Consulting     4    6.0000   0.8165                    (----*----)
                                     -----+-------+------+-------+-
                                         3.0     4.5    6.0     7.5
```

With an F value of 15.25 and a p-value of 0.000, the results of the one-way ANOVA show that there is an overall significant difference in job satisfaction between the five industries. Examining the Minitab confidence intervals shown graphically suggests that there might be a significant difference between some pairs of industries. Because there was an overall significant difference in the industries, it is appropriate to use Tukey's HSD test to determine which of the pairs of industries are significantly different. Tukey's test controls for the overall error so that the problem mentioned previously arising from computing ten *t* tests is avoided. The Minitab output for Tukey's test is:

```
Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons

Individual confidence level = 99.27%

IT subtracted from:
                Lower      Center      Upper   -----+-------+------+-------+-
Finance       -3.4462     -1.7500    -0.0538        (----*----)
Education     -4.8821     -3.2500    -1.6179   (----*----)
Healthcare    -4.0462     -2.3500    -0.6538      (----*-----)
Consulting    -1.5379      0.2500     2.0379             (-----*-----)
                                               -----+-------+------+-------+-
                                                  -3.0     0.0    3.0     6.0

Finance subtracted from:
                Lower      Center      Upper   -----+-------+------+-------+-
Education     -3.0311     -1.5000     0.0311        (----*----)
Healthcare    -2.1991     -0.6000     0.9991           (----*----)
Consulting     0.3038      2.0000     3.6962                 (----*----)
                                               -----+-------+------+-------+-
                                                  -3.0     0.0    3.0     6.0

Education subtracted from:
                Lower      Center      Upper   -----+-------+------+-------+-
Healthcare    -0.6311      0.9000     2.4311               (----*----)
Consulting     1.8679      3.5000     5.1321                      (--- *---)
                                               -----+-------+------+-------+-
                                                  -3.0     0.0    3.0     6.0

Healthcare subtracted from:
               Lower      Center      Upper    -----+-------+------+-------+-
Consulting    0.9038      2.6000      4.2962                  (----*----)
                                               -----+-------+------+-------+-
                                                  -3.0     0.0    3.0     6.0
```

Any confidence interval in which the sign of the value for the lower end of the interval is the same as the sign of the value for the upper end indicates that zero is not in the interval and that there is a significant difference between the pair in that case. Examining the Minitab output reveals that IT and Finance, IT and Education, IT and Healthcare, Finance and Consulting, Education and Consulting, and Healthcare and Consulting all are significantly different pairs of industries.

In analyzing career satisfaction, self-initiated expatriates were sampled from three age categories and four categories of time in the host country. This experimental design is a two-way factorial design with age and time in the host country as independent variables and individual scores on the seven-point Likert scale being the dependent variable. There are three classification levels under the independent variable Age: 30–39, 40–49, and over 50, and there are four classifications under the independent variable Time in Host Country: <1 year, 1 to 2 years, 3 to 4 years, and 5 or more years. Because there is more than one score per cell, interaction can be analyzed. A two-way ANOVA with replication is run in Excel to analyze the data and the result is shown below.

Excel Output:

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Age	3.5556	2	1.7778	2.91	0.073920	3.40
Time in Host Country	20.9722	3	6.9907	11.44	0.000075	3.01
Interaction	1.1111	6	0.1852	0.30	0.929183	2.51
Within	14.6667	24	0.6111			
Total	40.3056	35				

An examination of this output reveals no significant interaction effects (p-value of 0.929183). Since there are no significant interaction effects, it is appropriate to examine the main effects. Because of a p-value of 0.000075, there is a significant difference in Time in Host Country using an alpha of 0.0001. Multiple comparison analysis could be done to determine which, if any, pairs of Time in Host Country are significantly different. The p-value for Age is 0.073920 indicating that there is a significant difference between Age classifications at alpha 0.10 but not at alpha of 0.05.

ETHICAL CONSIDERATIONS

In theory, any phenomenon that affects the dependent variable in an experiment should be either entered into the experimental design or controlled in the experiment. Researchers will sometimes report statistical findings from an experiment and fail to mention the possible concomitant variables that were neither controlled by the experimental setting nor controlled by the experimental design. The findings from such studies are highly questionable and often lead to spurious conclusions. Scientifically, the researcher needs to conduct the experiment in an environment such that as many concomitant variables are controlled as possible. To the extent that they are not controlled, the researcher has an ethical responsibility to report that fact in the findings.

Other ethical considerations enter into conducting research with experimental designs. Selection of treatment levels should be done with fairness or even randomness in cases where the treatment has several possibilities for levels. A researcher can build in skewed views of the treatment effects by erroneously selecting treatment levels to be studied. Some researchers believe that reporting significant main effects from a factorial design when there are confounding interaction effects is unethical or at least misleading.

Another ethical consideration is the leveling of sample sizes. Some designs, such as the two-way factorial design or completely randomized design with Tukey's HSD, require equal sample sizes. Sometimes unequal sample sizes arise either through the selection process or through attrition. A number of techniques for approaching this problem are not presented in this book. It remains highly unethical to make up data values or to eliminate values arbitrarily to produce equal sample sizes.

SUMMARY

Sound business research requires that the researcher plan and establish a design for the experiment before a study is undertaken. The design of the experiment should encompass the treatment variables to be studied, manipulated, and controlled. These variables are often referred to as the independent variables. It is possible to study several independent variables and several levels, or classifications, of each of those variables in one design. In addition, the researcher selects one measurement to be taken from sample items under the conditions of the experiment. This measurement is referred to as the dependent variable because if the treatment effect is significant, the measurement of the dependent variable will "depend" on the independent variable(s) selected. This chapter explored three types of experimental designs: completely randomized design, randomized block design, and the factorial experimental designs.

The completely randomized design is the simplest of the experimental designs presented in this chapter. It has only one independent, or treatment, variable. With the completely randomized design, subjects are assigned randomly to treatments. If the treatment variable has only two levels, the design becomes identical to the one used to test the difference in means of independent populations presented in Chapter 10. The data from a completely randomized design are analyzed by a one-way analysis of variance (ANOVA). A one-way ANOVA produces an F value that can be compared to table F values in Appendix A.7 to determine whether the ANOVA F value is statistically significant. If it is, the null hypothesis that all population means are equal is rejected and at least one of the means is different from the others. Analysis of variance does not tell the researcher which means, if any, are significantly different from others. Although the researcher can visually examine means to determine which ones are greater and lesser, statistical techniques called multiple comparisons must be used to determine statistically whether pairs of means are significantly different.

Two types of multiple comparison techniques are presented and used in this chapter: Tukey's HSD test and the Tukey-Kramer procedure. Tukey's HSD test requires that equal sample sizes be used. It utilizes the mean square of error from the ANOVA, the sample size, and a q value that is obtained from Table A.10 to solve for the least difference between a pair of means that would be significant (HSD). The absolute value of the difference in sample means is compared to the HSD value to determine statistical significance. The Tukey-Kramer procedure is used in the case of unequal sample sizes.

A second experimental design is the randomized block design. This design contains a treatment variable (independent variable) and a blocking variable. The independent variable is the main variable of interest in this design. The blocking variable is a variable the researcher is interested in controlling rather than studying. A special case of randomized block design is the repeated measures design, in which the blocking variable represents subjects or items for which repeated measures are taken across the full range of treatment levels.

In randomized block designs, the variation of the blocking variable is removed from the error variance. This approach can potentially make the test of treatment effects more powerful. If the blocking variable contains no significant differences, the blocking can make the treatment effects test less powerful. Usually an F is computed only for the treatment effects in a randomized block design. Sometimes an F value is computed for blocking effects to determine whether the blocking was useful in the experiment.

A third experimental design is the factorial design. A factorial design enables the researcher to test the effects of two or more independent variables simultaneously. In complete factorial designs, every treatment level of each independent variable is studied under the conditions of every other treatment level for all independent variables. This chapter focused only on factorial designs with two independent variables. Each independent variable can have two or more treatment levels. These two-way factorial designs are analyzed by two-way analysis of variance (ANOVA). This analysis produces an F value for each of the two treatment effects and for interaction. Interaction is present when the results of one treatment vary significantly according to the levels of the other treatment. At least two measurements per cell must be present in order to compute interaction. If the F value for interaction is statistically significant, the main effects of the experiment are confounded and should not be examined in the usual manner.

KEY TERMS

Flash Cards

a posteriori
a priori
analysis of variance (ANOVA)
blocking variable
classification variable
classifications
completely randomized design
concomitant variables
confounding variables
dependent variable
experimental design
F distribution
F value
factorial design
factors
independent variable
interaction
levels
multiple comparisons
one-way analysis of variance
post hoc
randomized block design
repeated measures design
treatment variable
Tukey-Kramer procedure
Tukey's HSD test
two-way analysis of variance

FORMULAS

Formulas for computing a one-way ANOVA

$$\text{SSC} = \sum_{j=1}^{C} n_j(\bar{x}_j - \bar{x})^2$$

$$\text{SSE} = \sum_{i=1}^{n_j}\sum_{j=1}^{C}(x_{ij} - \bar{x}_j)^2$$

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

$$\text{df}_C = C - 1$$

$$\text{df}_E = N - C$$

$$\text{df}_T = N - 1$$

$$\text{MSC} = \frac{\text{SSC}}{\text{df}_C}$$

$$\text{MSE} = \frac{\text{SSE}}{\text{df}_E}$$

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

Tukey's HSD test

$$\text{HSD} = q_{\alpha,C,N-C}\sqrt{\frac{\text{MSE}}{n}}$$

Tukey-Kramer formula

$$q_{\alpha,C,N-C}\sqrt{\frac{\text{MSE}}{2}\left(\frac{1}{n_r} + \frac{1}{n_s}\right)}$$

Formulas for computing a randomized block design

$$\text{SSC} = n\sum_{j=1}^{C}(\bar{x}_j - \bar{x})^2$$

$$\text{SSR} = C\sum_{i=1}^{n}(\bar{x}_i - \bar{x})^2$$

$$\text{SSE} = \sum_{i=1}^{n}\sum_{j=1}^{C}(x_{ij} - \bar{x}_j - \bar{x}_i + \bar{x})^2$$

$$\text{SST} = \sum_{i=1}^{n}\sum_{j=1}^{C}(x_{ij} - \bar{x})^2$$

$$\text{df}_C = C - 1$$

$$\text{df}_R = n - 1$$

$$\text{df}_E = (C - 1)(n - 1) = N - n - C + 1$$

$$\text{MSC} = \frac{\text{SSC}}{C - 1}$$

$$\text{MSR} = \frac{\text{SSR}}{n - 1}$$

$$\text{MSE} = \frac{\text{SSE}}{N - n - C + 1}$$

$$F_{\text{treatments}} = \frac{\text{MSC}}{\text{MSE}}$$

$$F_{\text{blocks}} = \frac{\text{MSR}}{\text{MSE}}$$

Formulas for computing a two-way ANOVA

$$SSR = nC\sum_{i=1}^{R}(\bar{x}_i - \bar{x})^2$$

$$SSC = nR\sum_{j=1}^{C}(\bar{x}_j - \bar{x})^2$$

$$SSI = n\sum_{i=1}^{R}\sum_{j=1}^{C}(\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$$

$$SSE = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij})^2$$

$$SST = \sum_{i=1}^{R}\sum_{j=1}^{C}\sum_{k=1}^{n}(x_{ijk} - \bar{x})^2$$

$$df_R = R - 1$$

$$df_C = C - 1$$

$$df_I = (R - 1)(C - 1)$$

$$df_E = RC(n - 1)$$

$$df_T = N - 1$$

$$MSR = \frac{SSR}{R - 1}$$

$$MSC = \frac{SSC}{C - 1}$$

$$MSI = \frac{SSI}{(R - 1)(C - 1)}$$

$$MSE = \frac{SSE}{RC(n - 1)}$$

$$F_R = \frac{MSR}{MSE}$$

$$F_C = \frac{MSC}{MSE}$$

$$F_I = \frac{MSI}{MSE}$$

SUPPLEMENTARY PROBLEMS

CALCULATING THE STATISTICS

11.47 Compute a one-way ANOVA on the following data. Use $\alpha = .05$. If there is a significant difference in treatment levels, use Tukey's HSD to compute multiple comparisons. Let $\alpha = .05$ for the multiple comparisons.

Treatment			
1	*2*	*3*	*4*
10	9	12	10
12	7	13	10
15	9	14	13
11	6	14	12

11.48 Complete the following ANOVA table.

Source of Variance	SS	df	MS	F
Treatment				
Error	249.61	19		
Total	317.80	25		

11.49 You are asked to analyze a completely randomized design that has six treatment levels and a total of 42 measurements. Complete the following table, which contains some information from the study.

Source of Variance	SS	df	MS	F
Treatment	210			
Error	655			
Total				

11.50 Compute a one-way ANOVA of the following data. Let $\alpha = .01$. Use the Tukey-Kramer procedure to conduct multiple comparisons for the means.

Treatment		
1	*2*	*3*
7	11	8
12	17	6
9	16	10
11	13	9
8	10	11
9	15	7
11	14	10
10	18	
7		
8		

11.51 Examine the structure of the following experimental design. Determine which of the three designs presented in the chapter would be most likely to characterize this structure. Discuss the variables and the levels of variables. Determine the degrees of freedom.

	Methodology		
Person	*Method 1*	*Method 2*	*Method 3*
1	x_{11}	x_{12}	x_{13}
2	x_{21}	x_{22}	x_{23}
3	x_{31}	x_{32}	x_{33}
4	x_{41}	x_{42}	x_{43}
5	x_{51}	x_{52}	x_{53}
6	x_{61}	x_{62}	x_{63}

11.52 Complete the following ANOVA table and determine whether there is any significance in treatment effects. Let $\alpha = .05$.

Source of Variance	SS	df	MS	F
Treatment	20,994	3		
Blocking		9		
Error	33,891			
Total	71,338			

11.53 Analyze the following data, gathered from a randomized block design using $\alpha = .05$. If there is a significant difference in the treatment effects, use Tukey's HSD test to do multiple comparisons.

		Treatment			
		A	*B*	*C*	*D*
Blocking Variable	1	17	10	9	21
	2	13	9	8	16
	3	20	17	18	22
	4	11	6	5	10
	5	16	13	14	22
	6	23	19	20	28

11.54 A two-way ANOVA has been computed on a factorial design. Treatment 1 has five levels and treatment 2 has two levels. Each cell contains four measures. Complete the following ANOVA table. Use $\alpha = .05$ to test to determine significance of the effects. Comment on your findings.

Source of Variance	SS	df	MS	F
Treatment 1	29.13			
Treatment 2	12.67			
Interaction	73.49			
Error	110.38			
Total				

11.55 Compute a two-way ANOVA on the following data ($\alpha = .01$).

		Treatment 1		
		A	*B*	*C*
Treatment 2	*A*	5	2	2
		3	4	3
		6	4	5
	B	11	9	13
		8	10	12
		12	8	10
	C	6	7	4
		4	6	6
		5	7	8
	D	9	8	8
		11	12	9
		9	9	11

TESTING YOUR UNDERSTANDING

11.56 A company conducted a consumer research project to ascertain customer service ratings from its customers. The customers were asked to rate the company on a scale from 1 to 7 on various quality characteristics. One question was the promptness of company response to a repair problem. The following data represent customer responses to this question. The customers were divided by geographic region and by age. Use analysis of variance to analyze the responses. Let $\alpha = .05$. Compute multiple comparisons where they are appropriate. Graph the cell means and observe any interaction.

		Geographic Region			
		Southeast	*West*	*Midwest*	*Northeast*
Age	21–35	3	2	3	2
		2	4	3	3
		3	3	2	2
	36–50	5	4	5	6
		5	4	6	4
		4	6	5	5
	Over 50	3	2	3	3
		1	2	2	2
		2	3	3	1

11.57 A major automobile manufacturer wants to know whether there is any difference in the average mileage of four different brands of tires (A, B, C, and D), because the manufacturer is trying to select the best supplier in terms of tire durability. The manufacturer selects comparable levels of tires from each company and tests some on comparable cars. The mileage results follow.

A	B	C	D
31,000	24,000	30,500	24,500
25,000	25,500	28,000	27,000
28,500	27,000	32,500	26,000
29,000	26,500	28,000	21,000
32,000	25,000	31,000	25,500
27,500	28,000		26,000
	27,500		

Use $\alpha = .05$ to test whether there is a significant difference in the mean mileage of these four brands. Assume tire mileage is normally distributed.

11.58 Agricultural researchers are studying three different ways of planting peanuts to determine whether significantly different levels of production yield will result. The researchers have access to a large peanut farm on which to conduct their tests. They identify six blocks of land. In each block of land, peanuts are planted in each of the three different ways. At the end of the growing season, the peanuts are harvested and the average number of pounds per acre is determined for peanuts planted under each method in each block. Using the following data and $\alpha = .01$, test to determine whether there is a significant difference in yields among the planting methods.

Block	Method 1	Method 2	Method 3
1	1310	1080	850
2	1275	1100	1020
3	1280	1050	780
4	1225	1020	870
5	1190	990	805
6	1300	1030	910

11.59 The Construction Labor Research Council lists a number of construction labor jobs that seem to pay approximately the same wages per hour. Some of these are bricklaying, iron working, and crane operation. Suppose a labor researcher takes a random sample of workers from each of these types of construction jobs and from across the country and asks what are their hourly wages. If this survey yields the following data, is there a significant difference in mean hourly wages for these three jobs? If there is a significant difference, use the Tukey-Kramer procedure to determine which pairs, if any, are also significantly different. Let $\alpha = .05$.

Job Type		
Bricklaying	*Iron Working*	*Crane Operation*
19.25	26.45	16.20
17.80	21.10	23.30
20.50	16.40	22.90
24.33	22.86	19.50
19.81	25.55	27.00
22.29	18.50	22.95
21.20		25.52
		21.20

11.60 Why are mergers attractive to CEOs? One of the reasons might be a potential increase in market share that can come with the pooling of company markets. Suppose a random survey of CEOs is taken, and they are asked to respond on a scale from 1 to 5 (5 representing strongly agree) whether increase in market share is a good reason for considering a merger of their company with another. Suppose also that the data are as given here and that CEOs have been categorized by size of company and years they have been with their company. Use a two-way ANOVA to determine whether there are any significant differences in the responses to this question. Let $\alpha = .05$.

		Company Size ($ million per year in sales)			
		0–5	*6–20*	*21–100*	*>100*
		2	2	3	3
		3	1	4	4
	0–2	2	2	4	4
		2	3	5	3
Years		2	2	3	3
with the		1	3	2	3
Company	3–5	2	2	4	3
		3	3	4	4
		2	2	3	2
		1	3	2	3
	Over 5	1	1	3	2
		2	2	3	3

11.61 Are some unskilled office jobs viewed as having more status than others? Suppose a study is conducted in which eight unskilled, unemployed people are interviewed. The people are asked to rate each of five positions on a scale from 1 to 10 to indicate the status of the position, with 10 denoting most status and 1 denoting least status. The resulting data are given below. Use $\alpha = .05$ to analyze the repeated measures randomized block design data.

		Job				
		Mail Clerk	*Typist*	*Receptionist*	*Secretary*	*Telephone Operator*
	1	4	5	3	7	6
	2	2	4	4	5	4
	3	3	3	2	6	7
Respondent	4	4	4	4	5	4
	5	3	5	1	3	5
	6	3	4	2	7	7
	7	2	2	2	4	4
	8	3	4	3	6	6

INTERPRETING THE OUTPUT

11.62 Analyze the following Minitab output. Describe the design of the experiment. Using $\alpha = .05$ determine whether there are any significant effects; if so, explain why. Discuss any other ramifications of the output.

```
One-way ANOVA: Dependent Variable versus Factor
Analysis of Variance
Source          DF      SS       MS      F      p
Factor          3     876.6    292.2   3.01   0.045
Error           32   3107.5     97.1
Total           35   3984.1
                               Individual 95% CIs for
                               Mean  Based  on  Pooled
                               StDev
Level   N    Mean     StDev   -+----+----+----+--
C1      8    307.73   5.98    (----*----)
C2      7    313.20   9.71         (----*----)
C3      11   308.60   9.78      (----*----)
C4      10   319.74   12.18             (---*---)
                              -+----+----+----+--
Pooled StDev = 9.85        301.0 308.0 315.0 322.0
```

11.63 Following is Excel output for an ANOVA problem. Describe the experimental design. The given value of alpha was .05. Discuss the output in terms of significant findings.

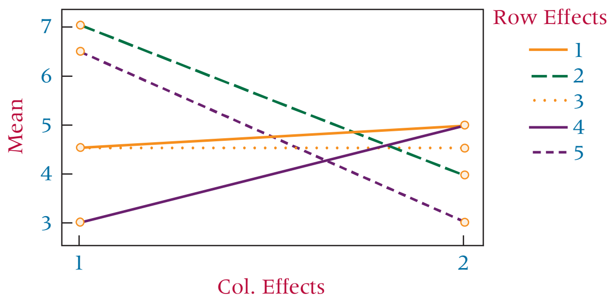
ANOVA: Two-Factor Without Replication

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Rows	48.278	5	9.656	3.16	0.057	3.33
Columns	10.111	2	5.056	1.65	0.23	4.10
Error	30.556	10	3.056			
Total	88.944	17				

11.64 Study the following Minitab output and graph. Discuss the meaning of the output.

```
Two-Way ANOVA: Dependent Variable Versus
Row Effects, Column Effects
Source        DF     SS     MS     F      p
Row Eff        4   4.70   1.17  0.98  0.461
Col. Eff       1   3.20   3.20  2.67  0.134
Interaction    4  22.30   5.57  4.65  0.022
Error         10  12.00   1.20
Total         19  42.20
```



11.65 Interpret the following Excel output. Discuss the structure of the experimental design and any significant effects. Alpha is .05.

ANOVA: Two-Factor With Replication

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	2913.889	3	971.296	4.30	0.0146	3.01
Columns	240.389	2	120.194	0.53	0.5940	3.40
Interaction	1342.944	6	223.824	0.99	0.4533	2.51
Within	5419.333	24	225.806			
Total	9916.556	35				

11.66 Study the following Minitab output. Determine whether there are any significant effects and discuss the results. What kind of design was used and what was the size of it?

```
Two-Way Analysis of Variance
Source       df       SS       MS
Blocking      4    41.44    10.36
Treatment     4   143.93    35.98
Error        16   117.82     7.36
Total        24   303.19
```

11.67 Discuss the following Minitab output.

```
One-Way Analysis of Variance
Source     df    SS     MS     F     P
Treatment   3  138.0  46.0  3.51  0.034
Error      20  262.2  13.1
Total      23  400.3
Individual 95% CIs For Mean Based on Pooled
StDev
Level  N   Mean    StDev ----+----+----+----+
1      6  53.778  5.470   (----*----)
2      6  54.665  1.840    (----*----)
3      6  59.911  3.845           (----*----)
4      6  57.293  2.088       (----*----)
                         ----+----+----+----+
Pooled StDev = 3.621      52.5 56.0 59.5 63.0

Tukey's pairwise comparisons
Family error rate = 0.0500
Individual error rate = 0.0111
Critical value = 3.96
Intervals for (column level mean) − (row level
mean)
        1         2         3
2   −6.741
     4.967
3  −11.987   −11.100
    −0.279     0.608
4   −9.369    −8.482   −3.236
     2.339     3.225    8.472
```

ANALYZING THE DATABASES

see www.wiley.com/college/black

1. Do various financial indicators differ significantly according to type of company? Use a one-way ANOVA and the financial database to answer this question. Let Type of Company be the independent variable with seven levels (Apparel, Chemical, Electric Power, Grocery, Healthcare Products, Insurance, and Petroleum). Compute three one-way ANOVAs, one for each of the following dependent variables: Earnings Per Share, Dividends Per Share, and Average P/E Ratio. On each ANOVA, if there is a significant overall difference between Type of Company, compute multiple comparisons to determine which pairs of types of companies, if any, are significantly different.

2. In the Manufacturing database, the Value of Industrial Shipments has been recoded into four classifications (1–4) according to magnitude of value. Let this value be the independent variable with four levels of classifications. Compute a one-way ANOVA to determine whether there is any significant difference in classification of the Value of Industrial Shipments on the Number of Production Workers (dependent variable). Perform the same analysis using End-of-Year Inventories as the dependent variable. Now change the independent variable to Industry Group, of which there are 20, and perform first a one-way ANOVA using Number of Production Workers as the dependent variable and then a one-way ANOVA using End-of-Year Inventory as the dependent variable.
3. The hospital database contains data on hospitals from seven different geographic regions. Let this variable be the independent variable. Determine whether there is a significant difference in Admissions for these geographic regions using a one-way ANOVA. Perform the same analysis using Births as the dependent variable. Control is a variable with four levels of classification denoting the type of control the hospital is under (such as federal government or for-profit). Use this variable as the independent variable and test to determine whether there is a significant difference in the Admissions of a hospital by Control. Perform the same test using Births as the dependent variable.
4. The Consumer Food database contains data on Annual Food Spending, Annual Household Income, and Non-Mortgage Household Debt broken down by Region and Location. Using Region as an independent variable with four classification levels (four regions of the U.S.), perform three different one-way ANOVA's—one for each of the three dependent variables (Annual Food Spending, Annual Household Income, Non-Mortgage Household Debt). Did you find any significant differences by region? If so, conduct multiple comparisons to determine which regions, if any, are significantly different.

CASE

THE CLARKSON COMPANY: A DIVISION OF TYCO INTERNATIONAL

In 1950, J. R. Clarkson founded a family-owned industrial valve design and manufacturing company in Sparks, Nevada. For almost a half century, the company, known as the Clarkson Company, worked on advancing metal and mineral processing. The Clarkson Company became known for its knife-gate and control valves, introduced in the 1970s, that are able to halt and isolate sections of slurry flow. By the late 1990s, the company had become a key supplier of knife-gate valves, helping to control the flow in many of the piping systems around the world in different industries, including mining, energy, and wastewater treatment.

The knife-gate valve uses a steel gate like a blade that lowers into a slurry flow to create a bubble-tight seal. While conventional metal gates fill with hardened slurry and fail easily thereby requiring high maintenance, Clarkson's design introduced an easily replaceable snap-in elastomer sleeve that is durable, versatile, and handles both high pressure and temperature variation. Pipeline operators value Clarkson's elastomer sleeve because traditional seals have cost between $75 and $500 to replace, and considerable revenue is lost when a slurry system is stopped for maintenance repairs. Clarkson's product lasts longer and is easier to replace.

In the late 1990s, the Clarkson Company was acquired by Tyco Valves & Controls, a division of Tyco International, Ltd. Tyco Valves & Controls, located in Reno, Nevada, and having ISO 9000 certification, continues to produce, market, and distribute products under the Clarkson brand name, including the popular knife-gate valve.

Discussion

1. The successful Clarkson knife-gate valve contains a wafer that is thin and light. Yet, the wafer is so strong it can operate with up to 150 pounds-per-square-inch (psi) of pressure on it, making it much stronger than those of competing brands. Suppose Tyco engineers have developed a new wafer that is even stronger. They want to set up an experimental design to test the strength of the wafer but they want to conduct the tests under three different temperature conditions, 70°, 110°, and 150°. In addition, suppose Tyco uses two different suppliers (company A and company B) of the synthetic materials that are used to manufacture the wafers. Some wafers are made primarily of raw materials supplied by company A, and some are made primarily of raw materials from company B. Thus, the engineers have set up a 2×3 factorial design with temperature and supplier as the independent variables and pressure (measured in psi) as the dependent variable. Data are gathered and are shown here. Analyze the data and discuss the business implications of the findings. If you were conducting the study, what would you report to the engineers?

	Temperature		
	70°	*110°*	*150°*
Supplier A	163	157	146
	159	162	137
	161	155	140
Supplier B	158	159	150
	154	157	142
	164	160	155

2. Pipeline operators estimate that it costs between $75 and $500 in U.S. currency to replace each seal, thus making the Clarkson longer-lasting valves more attractive. Tyco

does business with pipeline companies around the world. Suppose in an attempt to develop marketing materials, Tyco marketers are interested in determining whether there is a significant difference in the cost of replacing pipeline seals in different countries. Four countries—Canada, Colombia, Taiwan, and the United States—are chosen for the study. Pipeline operators from equivalent operations are selected from companies in each country. The operators keep a cost log of seal replacements. A random sample of the data follows. Use these data to help Tyco determine whether there is a difference in the cost of seal replacements in the various countries. Explain your answer and tell how Tyco might use the information in their marketing materials.

Canada	Colombia	Taiwan	United States
\$215	\$355	\$170	\$230
205	280	190	190
245	300	235	225
270	330	195	220
290	360	205	215
260	340	180	245
225	300	190	230

3. In the late 1980s, the Clarkson Company installed a manufacturing resource planning system. Using this and other quality improvement approaches, the company was able to reduce lead-time from six to eight weeks to less than two weeks. Suppose that Tyco now uses a similar system and wants to test to determine whether lead-times differ significantly according to the type of valve it is manufacturing. As a control of the experiment, they are including in the study, as a blocking variable, the day of the week the valve was ordered. One lead-time was selected per valve per day of the week. The data are given here in weeks. Analyze the data and discuss your findings.

	Type of Valve					
	Safety	***Butterfly***	***Clack***	***Slide***	***Poppet***	***Needle***
Monday	1.6	2.2	1.3	1.8	2.5	0.8
Tuesday	1.8	2.0	1.4	1.5	2.4	1.0
Wednesday	1.0	1.8	1.0	1.6	2.0	0.8
Thursday	1.8	2.2	1.4	1.6	1.8	0.6
Friday	2.0	2.4	1.5	1.8	2.2	1.2

Source: Adapted from "J. R. Clarkson Co., From Afterthought to Forethought," Real-World Lessons for America's Small Businesses: Insights from the Blue Chip Enterprise Initiative. Published by *Nation's Business* magazine on behalf of Connecticut Mutual Life Insurance Company and the U.S. Chamber of Commerce in association with the Blue Chip Enterprise Initiative, 1992; the Clarkson Co., Company Profile, Thomas Register Industry Answer Results, available at http://www.thomasregister.com; "The Clarkson Company Saves Time and Money Improving Piping Valves," ALGOR, pp. 1–4, available at http://www.algor.com; "Controlling the Flow," *Mechanical Engineering*, December 1998, pp. 1–5, available at http://www.memagazine.org. Tyco Valves & Controls Web site at www.tycovalves.com, and http://www.knifegate.com/, 2009.

USING THE COMPUTER

EXCEL

- Excel has the capability of performing a completely randomized design (one-way ANOVA), a randomized block design, and a two-way factorial design (two-way ANOVA).
- Each of the tests presented here in Excel is accessed through the **Data Analysis** feature.
- To conduct a one-way ANOVA, begin by selecting the **Data** tab on the Excel worksheet. From the **Analysis** panel at the right top of the **Data** tab worksheet, click on **Data Analysis**. If your Excel worksheet does not show the **Data Analysis** option, then you can load it as an add-in following directions given in Chapter 2. From the **Data Analysis** pulldown menu, select **Anova: Single Factor**. Click and drag over the data and enter in **Input Range**. Check **Labels in the First Row** if you included labels in the data. Insert the value of alpha in **Alpha**.
- To conduct a randomized block design, load the treatment observations into columns. Data may be loaded either with or without labels. Select the **Data** tab on the Excel worksheet. From the **Analysis** panel at the right top of the **Data** tab worksheet, click on **Data Analysis**. If your Excel worksheet does not show the **Data Analysis** option, then you can load it as an add-in following directions given in Chapter 2. From the **Data Analysis** pulldown menu, select **Anova: Two-Factor Without Replication**. Click and drag over the data under **Input Range**. Check **Labels in the First Row** if you have included labels in the data. Insert the value of alpha in **Alpha**.
- To conduct a two-way ANOVA, load the treatment observations into columns. Excel is quite particular about how the data are entered for a two-way ANOVA. Data must be loaded in rows and columns as with most two-way designs. However, two-way ANOVA in Excel requires labels for both rows and columns; and if labels are not supplied, Excel will incorrectly use some of the data for labels. Since cells will have multiple values, there need only be a label for each new row (cell). Select the **Data** tab on the Excel worksheet. From the **Analysis** panel at the right top of the **Data** tab worksheet, click on **Data Analysis**. If your Excel worksheet does not show the **Data Analysis** option, then you can load it as an add-in following directions given in Chapter 2. From the **Data Analysis** pulldown menu, select **Anova: Two-Factor With Replication**. Click and drag over the data under **Input Range**. Enter the number of values per cell in **Rows per sample**. Insert the value of alpha in **Alpha**.

MINITAB

- Minitab also has the capability to perform a completely randomized design (one-way ANOVA), a randomized block design, and a two-way factorial design along with multiple comparisons.
- There are two ways to compute a one-way ANOVA in Minitab, stacking all observations in one column with group identifiers in another column, or entering the observations unstacked in separate columns.
- To begin a one-way ANOVA with **Unstacked Data**, select **Stat** from the menu bar. Select **ANOVA** from the pulldown menu. Select **One-Way (Unstacked)**. In the slot, **Responses (in separate columns)**, list the columns containing the data. For multiple comparisons, select **Comparisons** and make your selection from the dialog box that appears. The multiple comparison options are Tukey's, Fisher's, Dunnett's, or Hsu's MCB tests. In the multiple comparison dialog box, you can insert the family error rate in the box on the right as a whole number.
- To begin a one-way ANOVA with **Stacked Data**, select **Stat** from the menu bar. Select **ANOVA** from the pulldown menu. Select **One-Way**. In the slot **Response**, list the column containing the observations. In the slot **Factor**, list the column containing the group identifiers. For multiple comparisons, select **Comparisons** and make your selection from the dialog box that appears. The multiple comparison options are Tukey's, Fisher's, Dunnett's, or Hsu's MCB tests. In the multiple comparison dialog box, you can insert the family error rate in the box on the right as a whole number.
- There are two ways to compute a randomized block design or a two-way ANOVA in Minitab, using the **Two-Way** procedure or using the **Balanced ANOVA** procedure. Both the randomized block design and the two-way ANOVA are analyzed in the same way and are presented together here.
- To begin using the **Two-Way** procedure, select **Stat** from the menu bar. Select **ANOVA** from the pulldown menu. Select **Two-Way**. The observations should be "stacked," that is, listed in one column. Place the location of these observations in the **Response** box. The group identifiers for the row effects should be listed in another column. Place the location of the row identifiers in **Row factor**. The group identifiers for the column effects should be listed in another column. Place the location of the column identifiers in **Column factor**.
- To begin using the **Balanced ANOVA** procedure, select **Stat** from the menu bar. Select **ANOVA** from the pulldown menu. Select **Balanced ANOVA** from the pulldown menu. Place the location of the observations in the **Responses** box. Place the columns containing the group identifiers to be analyzed in the **Model** box. To compute a two-way ANOVA with interaction, place the location of the column containing the row effects identifiers and the location of the column containing column effects identifiers in the **Model** box. To test for interaction also place a third term in the **Model** box that is the product of the row and column effects. For example, if the row effects are X and the column effects are Y, place a X*Y in the **Model** box along with X and Y.