CHAPTER 12

Experimental Design and Analysis of Variance

Learning Objectives

After mastering the material in this chapter, you will be able to:

- **LO12-1** Explain the basic terminology and concepts of experimental design.
- **LO12-2** Compare several different population means by using a one-way analysis of variance.
- **LO12-3** Compare treatment effects and block effects by using a randomized block design.
- **LO12-4** Assess the effects of two factors on a response variable by using a two-way analysis of variance.
- **LO12-5** Describe what happens when two factors interact.

Chapter Outline

12.1 Basic Concepts of Experimental Design

12.2 One-Way Analysis of Variance

12.3 The Randomized Block Design

12.4 Two-Way Analysis of Variance

In Chapter 10 we learned that business improvement often involves making **comparisons.** In that chapter we presented several confidence intervals and several hypothesis testing procedures for comparing two population means. However, business improvement often requires that we compare more than two population means. For instance, we might compare the mean sales obtained by using three different advertising campaigns in order to improve a company's marketing process. Or, we might compare the mean production output obtained by using four different manufacturing process designs to improve productivity.

In this chapter we extend the methods presented in Chapter 10 by considering statistical procedures for **comparing two or more population means.** Each of the methods we discuss is called an **analysis of variance (ANOVA)** procedure. We also present some basic concepts of **experimental design,** which involves deciding how to collect data in a way that allows us to most effectively compare population means.

We explain the methods of this chapter in the context of three cases:

C

The Oil Company Case: An oil company wishes to develop a reasonably priced gasoline that will deliver improved mileages. The company uses **one-way analysis of variance** to compare the effects of three types of gasoline on mileage in order to find the gasoline type that delivers the highest mean mileage.

The Cardboard Box Case: A paper company performs an experiment to investigate the effects of four production methods on the number of defective cardboard boxes produced in an hour. The company uses a **randomized block ANOVA** to determine which production method yields the smallest mean number of defective boxes.

The Supermarket Case: A commercial bakery supplies many supermarkets. In order to improve the effectiveness of its supermarket shelf displays the company wishes to compare the effects of shelf display height (bottom, middle, or top) and width (regular or wide) on monthly demand. The bakery employs **two-way analysis of variance** to find the display height and width combination that produces the highest monthly demand.

12.1 Basic Concepts of Experimental Design

LO12-1 Explain the basic terminology and concepts of experimental design.

In many statistical studies a variable of interest, called the **response variable** (or **dependent variable**), is identified. Then data are collected that tell us about how one or more **factors** (or **independent variables**) influence the variable of interest. If we cannot control the factor(s) being studied, we say that the data obtained are **observational.** For example, suppose that in order to study how the size of a home relates to the sales price of the home, a real estate agent randomly selects 50 recently sold homes and records the square footages and sales prices of these homes. Because the real estate agent cannot control the sizes of the randomly selected homes, we say that the data are observational.

If we can control the factors being studied, we say that the data are **experimental.** Furthermore, in this case the values, or **levels,** of the factor (or combination of factors) are called **treatments.** The purpose of most experiments is **to compare and estimate the effects of the different treatments on the response variable.** For example, suppose that an oil company wishes to study how three different gasoline types (A, B, and C) affect the mileage obtained by a popular compact automobile model. Here the response variable is gasoline mileage, and the company will study a single factor—gasoline type. Because the oil company can control which gasoline type is used in the compact automobile, the data that the oil company will collect are experimental. Furthermore, the treatments—the levels of the factor gasoline type—are gasoline types A, B, and C.

In order to collect data in an experiment, the different treatments are assigned to objects (people, cars, animals, or the like) that are called **experimental units.** For example, in the gasoline mileage situation, gasoline types A, B, and C will be compared by conducting mileage tests using a compact automobile. The automobiles used in the tests are the experimental units.

In general, when a treatment is applied to more than one experimental unit, it is said to be **replicated.** Furthermore, when the analyst controls the treatments employed and how they are applied to the experimental units, a **designed experiment** is being carried out. A commonly used, simple experimental design is called the **completely randomized experimental design.**

In a **completely randomized experimental design,** independent random samples of experimental units are assigned to the treatments.

As illustrated in the following examples, we can sometimes assign *independent* random samples of experimental units to the treatments by assigning *different* random samples of experimental units to different treatments.

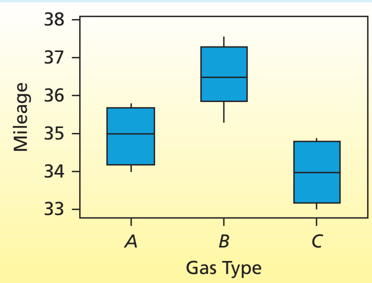
EXAMPLE 12.1 The Oil Company Case: Comparing Gasoline Types

North American Oil Company is attempting to develop a reasonably priced gasoline that will deliver improved gasoline mileages. As part of its development process, the company would like to compare the effects of three types of gasoline (A, B, and C) on gasoline mileage. For testing purposes, North American Oil will compare the effects of gasoline types A, B, and C on the gasoline mileage obtained by a popular compact model called the Lance. Suppose the company has access to 1,000 Lances that are representative of the population of all Lances, and suppose the company will utilize a completely randomized experimental design that employs samples of size five. In order to accomplish this, five Lances will be randomly selected from the 1,000 available Lances. These autos will be assigned to gasoline type A. Next, five *different* Lances will be randomly selected from the remaining 995 available Lances. These autos will be assigned to gasoline type B. Finally, five *different* Lances will be randomly selected from the remaining 990 available Lances. These autos will be assigned to gasoline type C.

Each randomly selected Lance is test driven using the appropriate gasoline type (treatment) under normal conditions for a specified distance, and the gasoline mileage for each test drive is measured. We let x_{ij} denote the j^{th} mileage obtained when using gasoline type i. The mileage data obtained are given in Table 12.1. Here we assume that the set of gasoline mileage observations obtained by using a particular gasoline type is a sample randomly selected from the infinite population of all Lance mileages that could be obtained using that gasoline type. Examining the box plots shown next to the mileage data, we see some evidence that gasoline type B yields the highest gasoline mileages.

TABLE 12.1 The Gasoline Mileage Data DS GasMile2

Gasoline Type *A*	Gasoline Type *B*	Gasoline Type *C*
$x_{A1} = 34.0$	$x_{B1} = 35.3$	$x_{C1} = 33.3$
$x_{A2} = 35.0$	$x_{B2} = 36.5$	$x_{C2} = 34.0$
$x_{A3} = 34.3$	$x_{B3} = 36.4$	$x_{C3} = 34.7$
$x_{A4} = 35.5$	$x_{B4} = 37.0$	$x_{C4} = 33.0$
$x_{A5} = 35.8$	$x_{B5} = 37.6$	$x_{C5} = 34.9$

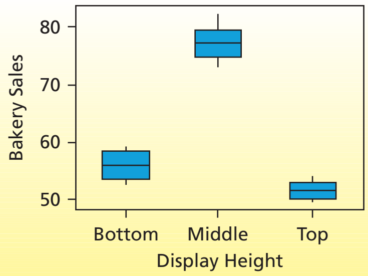


EXAMPLE 12.2 The Supermarket Case: Studying the Effect of Display Height

The Tastee Bakery Company supplies a bakery product to many supermarkets in a metropolitan area. The company wishes to study the effect of the shelf display height employed by the supermarkets on monthly sales (measured in cases of 10 units each) for this product. Shelf display height, the factor to be studied, has three levels—bottom (B), middle (M), and top (T)—which are the treatments. To compare these treatments, the bakery uses a completely randomized experimental design. For each shelf height, six supermarkets (the experimental units) of equal sales potential are randomly selected, and each supermarket displays the product using its assigned shelf height for a month. At the end of the month, sales of the bakery product (the response variable) at the 18 participating stores are recorded, giving the data in Table 12.2. Here we assume that the set of sales amounts for each display height is a sample randomly selected from the population of all sales amounts that could be obtained (at supermarkets of the given sales potential) at that display height. Examining the box plots that are shown next to the sales data, we seem to have evidence that a middle display height gives the highest bakery product sales.

TABLE 12.2 The Bakery Product Sales Data DS BakeSale

	Shelf Display Height	
Bottom (*B*)	Middle (*M*)	Top (*T*)
58.2	73.0	52.4
53.7	78.1	49.7
55.8	75.4	50.9
55.7	76.2	54.0
52.5	78.4	52.1
58.9	82.1	49.9



12.2 One-Way Analysis of Variance

LO12-2 Compare several different population means by using a one-way analysis of variance.

Suppose we wish to study the effects of p **treatments** (treatments 1, 2, . . . , p) on a **response variable.** For any particular treatment, say treatment i, we define μ_i and σ_i to be the mean and standard deviation of the population of all possible values of the response variable that could potentially be observed when using treatment i. Here we refer to μ_i as **treatment mean *i*.** The goal of **one-way analysis of variance** (often called **one-way ANOVA**) is to estimate and compare the effects of the different treatments on the response variable. We do this by **estimating and comparing the treatment means** $\mu_1, \mu_2, \ldots, \mu_p$. Here we assume that a sample has been randomly selected for each of the p treatments by employing a completely randomized experimental design. We let n_i denote the size of the sample that has been randomly selected for treatment i, and we let x_{ij} denote the j^{th} value of the response variable that is observed when using treatment i. It then follows that the point estimate of μ_i is $\bar{x}_i$, the average of the sample of n_i values of the response variable observed when using treatment i. It further follows that the point estimate of σ_i is s_i, the standard deviation of the sample of n_i values of the response variable observed when using treatment i.

For example, consider the gasoline mileage situation. We let μ_A, μ_B, and μ_C denote the means and σ_A, σ_B, and σ_C denote the standard deviations of the populations of all possible gasoline mileages using gasoline types A, B, and C. To estimate these means and standard deviations, North American Oil has employed a completely randomized experimental design and has obtained the samples of mileages in Table 12.1. The means of these samples—$\bar{x}_A = 34.92$, $\bar{x}_B = 36.56$, and $\bar{x}_C = 33.98$—are the point estimates of μ_A, μ_B, and μ_C. The standard deviations of these samples—$s_A = .7662$, $s_B = .8503$, and $s_C = .8349$—are the point estimates of σ_A, σ_B, and σ_C. Using these point estimates, we will (later in this section) test to see whether there are any statistically significant differences between the treatment means μ_A, μ_B, and μ_C. If such differences exist, we will estimate the magnitudes of these differences. This will allow North American Oil to judge whether these differences have practical importance.

The one-way ANOVA formulas allow us to test for significant differences between treatment means and allow us to estimate differences between treatment means. The validity of these formulas requires that the following assumptions hold:

Assumptions for One-Way Analysis of Variance

1 **Constant variance**—the p populations of values of the response variable associated with the treatments have equal variances.

2 **Normality**—the p populations of values of the response variable associated with the treatments all have normal distributions.

3 **Independence**—the samples of experimental units associated with the treatments are randomly selected, independent samples.

The one-way ANOVA results are not very sensitive to violations of the equal variances assumption. Studies have shown that this is particularly true when the sample sizes employed are equal (or nearly equal). Therefore, a good way to make sure that unequal variances will not be a problem is to take samples that are the same size. In addition, it is useful to compare the sample standard deviations $s_1, s_2, \ldots, s_p$ to see if they are reasonably equal. As a general rule, *the one-way ANOVA results will be approximately correct if the largest sample standard deviation is no more than twice the smallest sample standard deviation.* The variations of the samples can also be compared by constructing a box plot for each sample (as we have done for the gasoline mileage data in Table 12.1). Several statistical tests also employ the sample variances to test the equality of the population variances [see Bowerman and O'Connell (1990) for two of these tests]. However, these tests have some drawbacks—in particular, their results are very sensitive to violations of the normality assumption. Because of this, there is controversy as to whether these tests should be performed.

The normality assumption says that each of the p populations is normally distributed. This assumption is not crucial. It has been shown that the one-way ANOVA results are approximately valid for mound-shaped distributions. It is useful to construct a box plot and/or a stem-and-leaf display for each sample. If the distributions are reasonably symmetric, and if there are no outliers, the ANOVA results can be trusted for sample sizes as small as 4 or 5. As an example, consider the gasoline mileage study of Example 12.1. The box plots of Table 12.1 suggest that the variability of the mileages in each of the three samples is roughly the same. Furthermore, the sample standard deviations $s_A = .7662$, $s_B = .8503$, and $s_C = .8349$ are reasonably equal (the largest is not even close to twice the smallest). Therefore, it is reasonable to believe that the constant variance assumption is satisfied. Moreover, because the sample sizes are the same, unequal variances would probably not be a serious problem anyway. Many small, independent factors influence gasoline mileage, so the distributions of mileages for gasoline types A, B, and C are probably mound-shaped. In addition, the box plots of Table 12.1 indicate that each distribution is roughly symmetric with no outliers. Thus, the normality assumption probably approximately holds. Finally, because North American Oil has employed a completely randomized design, the independence assumption probably holds. This is because the gasoline mileages in the different samples were obtained for *different* Lances.

Testing for significant differences between treatment means As a preliminary step in one-way ANOVA, we wish to determine whether there are any statistically significant differences between the treatment means $\mu_1, \mu_2, \ldots, \mu_p$. To do this, we test the null hypothesis

$$H_0\text{: } \mu_1 = \mu_2 = \cdots = \mu_p$$

This hypothesis says that all the treatments have the same effect on the mean response. We test H_0 versus the alternative hypothesis

$$H_a\text{: At least two of } \mu_1, \mu_2, \ldots, \mu_p \text{ differ}$$

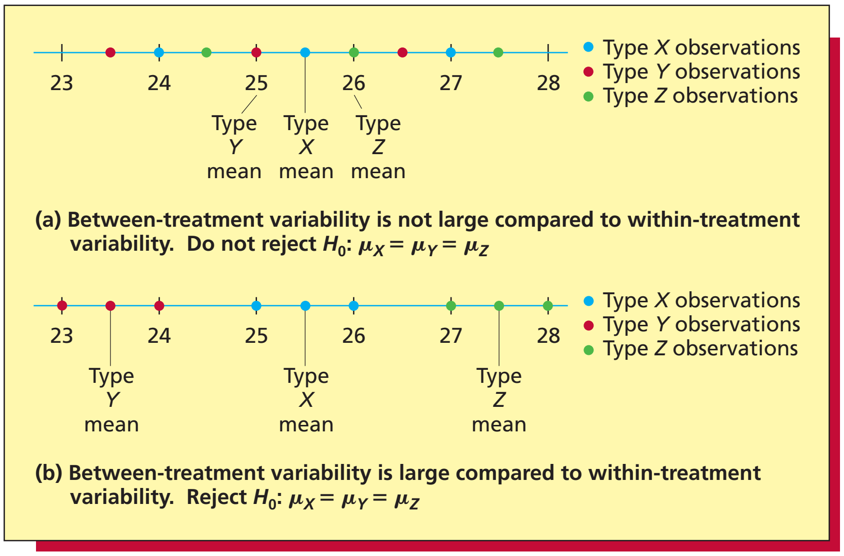
This alternative says that at least two treatments have different effects on the mean response.

To carry out such a test, we compare what we call the **between-treatment variability** to the **within-treatment variability.** For instance, suppose we wish to study the effects of three gasoline types (X, Y, and Z) on mean gasoline mileage, and consider Figure 12.1(a). This figure depicts three independent random samples of gasoline mileages obtained using gasoline types X, Y, and Z. Observations obtained using gasoline type X are plotted as blue dots (●), observations obtained using gasoline type Y are plotted as red dots (●), and observations obtained using gasoline type Z are plotted as green dots (●). Furthermore, the sample treatment means are labeled as "type X mean," "type Y mean," and "type Z mean." We see that the variability of the sample treatment means—that is, the **between-treatment variability**—is not large compared to the variability within each sample (the **within-treatment variability**). In this case, the differences between the sample treatment means could quite easily be the result of sampling variation. Thus we would not have sufficient evidence to reject

$$H_0\text{: } \mu_X = \mu_Y = \mu_Z$$

Next look at Figure 12.1(b), which depicts a different set of three independent random samples of gasoline mileages. Here the variability of the sample treatment means (the between-treatment variability) is large compared to the variability within each sample. This would probably provide

FIGURE 12.1 Comparing Between-Treatment Variability and Within-Treatment Variability



enough evidence to tell us to reject H_0: $\mu_X = \mu_Y = \mu_Z$ in favor of H_a: At least two of μ_X, μ_Y, and μ_Z differ. We would conclude that at least two of gasoline types X, Y, and Z have different effects on mean mileage.

In order to numerically compare the between-treatment and within-treatment variability, we can define several **sums of squares** and **mean squares.** To begin, we define n to be the total number of experimental units employed in the one-way ANOVA, and we define $\bar{x}$ to be the overall mean of all observed values of the response variable. Then we define the following:

The **treatment sum of squares** is

$$SST = \sum_{i=1}^{p} n_i(\bar{x}_i - \bar{x})^2$$

In order to compute SST, we calculate the difference between each sample treatment mean $\bar{x}_i$ and the overall mean $\bar{x}$, we square each of these differences, we multiply each squared difference by the number of observations for that treatment, and we sum over all treatments. The SST measures the variability of the sample treatment means. For instance, if all the sample treatment means ($\bar{x}_i$ values) were equal, then the treatment sum of squares would be equal to 0. The more the $\bar{x}_i$ values vary, the larger will be SST. In other words, the **treatment sum of squares** measures the amount of **between-treatment variability.**

As an example, consider the gasoline mileage data in Table 12.1. In this experiment we employ a total of

$$n = n_A + n_B + n_C = 5 + 5 + 5 = 15$$

experimental units. Furthermore, the overall mean of the 15 observed gasoline mileages is

$$\bar{x} = \frac{34.0 + 35.0 + \cdots + 34.9}{15} = \frac{527.3}{15} = 35.153$$

Then

$$\begin{aligned} SST &= \sum_{i=A,B,C} n_i(\bar{x}_i - \bar{x})^2 \\ &= n_A(\bar{x}_A - \bar{x})^2 + n_B(\bar{x}_B - \bar{x})^2 + n_C(\bar{x}_C - \bar{x})^2 \\ &= 5(34.92 - 35.153)^2 + 5(36.56 - 35.153)^2 + 5(33.98 - 35.153)^2 \\ &= 17.0493 \end{aligned}$$

In order to measure the within-treatment variability, we define the following quantity:

The **error sum of squares** is

$$SSE = \sum_{j=1}^{n_1}(x_{1j} - \bar{x}_1)^2 + \sum_{j=1}^{n_2}(x_{2j} - \bar{x}_2)^2 + \cdots + \sum_{j=1}^{n_p}(x_{pj} - \bar{x}_p)^2$$

Here x_{1j} is the j^{th} observed value of the response in the first sample, x_{2j} is the j^{th} observed value of the response in the second sample, and so forth. The formula above says that we compute SSE by calculating the squared difference between each observed value of the response and its corresponding sample treatment mean and by summing these squared differences over all the observations in the experiment.

The SSE measures the variability of the observed values of the response variable around their respective sample treatment means. For example, if there were no variability within each sample, the error sum of squares would be equal to 0. The more the values within the samples vary, the larger will be SSE.

As an example, in the gasoline mileage study, the sample treatment means are $\bar{x}_A = 34.92$, $\bar{x}_B = 36.56$, and $\bar{x}_C = 33.98$. It follows that

$$\begin{aligned} SSE &= \sum_{j=1}^{n_A}(x_{Aj} - \bar{x}_A)^2 + \sum_{j=1}^{n_B}(x_{Bj} - \bar{x}_B)^2 + \sum_{j=1}^{n_C}(x_{Cj} - \bar{x}_C)^2 \\ &= [(34.0 - 34.92)^2 + (35.0 - 34.92)^2 + (34.3 - 34.92)^2 + (35.5 - 34.92)^2 + (35.8 - 34.92)^2] \\ &\quad + [(35.3 - 36.56)^2 + (36.5 - 36.56)^2 + (36.4 - 36.56)^2 + (37.0 - 36.56)^2 + (37.6 - 36.56)^2] \\ &\quad + [(33.3 - 33.98)^2 + (34.0 - 33.98)^2 + (34.7 - 33.98)^2 + (33.0 - 33.98)^2 + (34.9 - 33.98)^2] \\ &= 8.028 \end{aligned}$$

Finally, we define a sum of squares that measures the total amount of variability in the observed values of the response:

The **total sum of squares** is

$$SSTO = SST + SSE$$

The variability in the observed values of the response must come from one of two sources—the between-treatment variability or the within-treatment variability. It follows that the total sum of squares equals the sum of the treatment sum of squares and the error sum of squares. Therefore, the ***SST* and *SSE* are said to partition the total sum of squares.** For the gasoline mileage study

$$SSTO = SST + SSE = 17.0493 + 8.028 = 25.0773$$

In order to decide whether there are any statistically significant differences between the treatment means, it makes sense to compare the amount of between-treatment variability to the amount of within-treatment variability. This comparison suggests the following F-test:

An *F*-Test for Differences between Treatment Means

Suppose that we wish to compare p treatment means $\mu_1, \mu_2, \ldots, \mu_p$ and consider testing

H_0: $\mu_1 = \mu_2 = \cdots = \mu_p$ (all treatment means are equal) versus H_a: At least two of $\mu_1, \mu_2, \ldots, \mu_p$ differ (at least two treatment means differ)

To perform the hypothesis test, define the **treatment mean square** to be $MST = SST/(p - 1)$ and define the **error mean square** to be $MSE = SSE/(n - p)$. Also, define the F statistic

$$F = \frac{MST}{MSE} = \frac{SST/(p-1)}{SSE/(n-p)}$$

and its p-value to be the area under the F curve with $p - 1$ and $n - p$ degrees of freedom to the right of F. We can reject H_0 in favor of H_a at level of significance α if either of the following equivalent conditions holds:

1 $F > F_\alpha$ **2** p-value $< \alpha$

Here the F_α point is based on $p - 1$ numerator and $n - p$ denominator degrees of freedom.

A large value of F results when SST, which measures the between-treatment variability, is large compared to SSE, which measures the within-treatment variability. If F is large enough, this implies that H_0 should be rejected. The rejection point F_α tells us when F is large enough to allow us to reject H_0 at level of significance α. When F is large, the associated p-value is small. If this p-value is less than α, we can reject H_0 at level of significance α.

EXAMPLE 12.3 The Oil Company Case: Comparing Gasoline Types

C

Consider the North American Oil Company data in Table 12.1. The company wishes to determine whether any of gasoline types A, B, and C have different effects on mean Lance gasoline mileage. That is, we wish to see whether there are any statistically significant differences between μ_A, μ_B, and μ_C. To do this, we test the null hypothesis H_0: $\mu_A = \mu_B = \mu_C$, which says that gasoline types A, B, and C have the same effects on mean gasoline mileage. We test H_0 versus the alternative H_a: At least two of μ_A, μ_B, and μ_C differ, which says that at least two of gasoline types A, B, and C have different effects on mean gasoline mileage.

Because we have previously computed SST to be 17.0493 and SSE to be 8.028, and because we are comparing $p = 3$ treatment means, we have

$$MST = \frac{SST}{p - 1} = \frac{17.0493}{3 - 1} = 8.525$$

and

$$MSE = \frac{SSE}{n - p} = \frac{8.028}{15 - 3} = 0.669$$

It follows that

$$F = \frac{MST}{MSE} = \frac{8.525}{0.669} = 12.74$$

In order to test H_0 at the .05 level of significance, we use $F_{.05}$ with $p - 1 = 3 - 1 = 2$ numerator and $n - p = 15 - 3 = 12$ denominator degrees of freedom. Table A.7 (page 796) tells us that this F point equals 3.89, so we have

$$F = 12.74 > F_{.05} = 3.89$$

Therefore, we reject H_0 at the .05 level of significance. This says we have strong evidence that at least two of the treatment means μ_A, μ_B, and μ_C differ. In other words, we conclude that at least two of gasoline types A, B, and C have different effects on mean gasoline mileage.

The results of an analysis of variance are often summarized in what is called an **analysis of variance table.** This table gives the sums of squares (SST, SSE, $SSTO$), the mean squares (MST and MSE), and the F statistic and its related p-value for the ANOVA. The table also gives the degrees of freedom associated with each source of variation—treatments, error, and total. Table 12.3 gives the ANOVA table for the gasoline mileage problem. Notice that in the column labeled "Sums of Squares," the values of SST and SSE sum to $SSTO$.

Figure 12.2 gives the MINITAB and Excel output of an analysis of variance of the gasoline mileage data. Note that the upper portion of the MINITAB output and the lower portion of the Excel output give the ANOVA table of Table 12.3. Also, note that each output gives the value $F = 12.74$ and the related p-value, which equals .001(rounded). Because this p-value is less than .05, we reject H_0 at the .05 level of significance.

Pairwise comparisons If the one-way ANOVA F test says that at least two treatment means differ, then we investigate which treatment means differ and we estimate how large the differences are. We do this by making what we call **pairwise comparisons** (that is, we compare treatment means *two at a time*). One way to make these comparisons is to compute point estimates of and confidence intervals for **pairwise differences.** For example, in the oil company case we might estimate the pairwise differences $\mu_A - \mu_B$, $\mu_A - \mu_C$, and $\mu_B - \mu_C$. Here, for instance, the pairwise difference $\mu_A - \mu_B$ can be interpreted as the change in mean mileage achieved by changing from using gasoline type B to using gasoline type A.

TABLE 12.3 Analysis of Variance (ANOVA) Table for Testing H_0: $\mu_A = \mu_B = \mu_C$ in the Oil Company Case ($p = 3$ Gasoline Types, $n = 15$ Observations)

Source	Degrees of Freedom	Sums of Squares	Mean Squares	F Statistic	p-Value
Treatments	$p - 1 = 3 - 1 = 2$	$SST = 17.0493$	$MST = \frac{SST}{p-1} = \frac{17.0493}{3-1} = 8.525$	$F = \frac{MST}{MSE} = \frac{8.525}{0.669} = 12.74$	0.001
Error	$n - p = 15 - 3 = 12$	$SSE = 8.028$	$MSE = \frac{SSE}{n-p} = \frac{8.028}{15-3} = 0.669$		
Total	$n - 1 = 15 - 1 = 14$	$SSTO = 25.0773$			

FIGURE 12.2 MINITAB and Excel Output of an Analysis of Variance of the Oil Company Gasoline Mileage Data in Table 12.1

(a) The MINITAB output

```
One-way  ANOVA: Type A,  Type B,  Type C                   Tukey 95% Simultaneous
Source    DF      SS         MS        F        P           Confidence Intervals
Gas Type   2 [1]  17.049 [4] 8.525 [7] 12.74 [9] 0.001 [10]
Error     12 [2]   8.028 [5] 0.669 [8]                     Type A subtracted from:
Total     14 [3]  25.077 [6]                                          Lower    Center    Upper
                                Individual 95%             Type B    0.2610    1.6400   3.0190
                                CIs For Mean Based on Pooled StDev   Type C   -2.3190   -0.9400   0.439
Level   N  Mean         StDev  ---+---------+---------+---------+------
Type A  5  34.920 [11]  0.766         (------*------)       Type B subtracted from:
Type B  5  36.560 [12]  0.850                   (------*-----)        Lower    Center    Upper
Type C  5  33.980 [13]  0.835 (-----*------)              Type C   -3.9590   -2.5800  -1.2010
                               ---+---------+---------+---------+------
Pooled StDev = 0.818             33.6      34.8      36.0      37.2
```

(b) The Excel output

SUMMARY

Groups	Count	Sum	Average	Variance
Type A	5	174.6	34.92 [11]	0.587
Type B	5	182.8	36.56 [12]	0.723
Type C	5	169.9	33.98 [13]	0.697

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	17.0493 [4]	2 [1]	8.5247 [7]	12.7424 [9]	0.0011 [10]	3.8853 [14]
Within Groups	8.0280 [5]	12 [2]	0.6690 [8]			
Total	25.0773 [6]	14 [3]				

[1] $p-1$ [2] $n-p$ [3] $n-1$ [4] SST [5] SSE [6] $SSTO$ [7] MST [8] MSE [9] F statistic [10] p-value related to F [11] $\bar{x}_A$ [12] $\bar{x}_B$ [13] $\bar{x}_C$ [14] $F_{.05}$

There are two approaches to calculating confidence intervals for pairwise differences. The first involves computing the usual, or **individual, confidence interval** for each pairwise difference. Here, if we are computing $100(1 - \alpha)$ percent confidence intervals, we are $100(1 - \alpha)$ percent confident that each individual pairwise difference is contained in its respective interval. That is, the confidence level associated with each (individual) comparison is $100(1 - \alpha)$ percent, and we refer to α as the **comparisonwise error rate.** However, we are less than $100(1 - \alpha)$ percent confident that all of the pairwise differences are simultaneously contained in their respective intervals. A more conservative approach is to compute **simultaneous confidence intervals**. Such intervals make us $100(1 - \alpha)$ percent confident that all of the pairwise differences are simultaneously contained in their respective intervals. That is, when we compute simultaneous intervals, the overall confidence level associated with all the comparisons being made in the experiment is $100(1 - \alpha)$ percent, and we refer to α as the **experimentwise error rate.**

Several kinds of simultaneous confidence intervals can be computed. In this book we present what is called the **Tukey formula** for simultaneous intervals. We do this because, *if we are interested in studying all pairwise differences between treatment means, the Tukey formula yields the most precise (shortest) simultaneous confidence intervals.*

Estimation in One-Way ANOVA

1. Consider the **pairwise difference** $\boldsymbol{\mu_i - \mu_h}$, which can be interpreted to be the change in the mean value of the response variable associated with changing from using treatment h to using treatment i. Then, a **point estimate of the difference** $\boldsymbol{\mu_i - \mu_h}$ **is** $\boldsymbol{\bar{x}_i - \bar{x}_h}$, where $\bar{x}_i$ and $\bar{x}_h$ are the sample treatment means associated with treatments i and h.

2. A **Tukey simultaneous 100(1 − α) percent confidence interval for** $\boldsymbol{\mu_i - \mu_h}$ is

$$\left[(\bar{x}_i - \bar{x}_h) \pm q_\alpha \sqrt{\frac{MSE}{m}}\right]$$

Here, the value q_α is obtained from Table A.10 (pages 799–800), which is a **table of percentage points of the studentized range.** In this table q_α is listed corresponding to values of p and $n - p$. Furthermore, we assume that the sample sizes n_i and n_h are equal to the same value, which we denote as m. If n_i and n_h are not equal, we replace $q_\alpha\sqrt{MSE/m}$ by $(q_\alpha/\sqrt{2})\sqrt{MSE[(1/n_i) + (1/n_h)]}$.

3. A **point estimate of the treatment mean** $\boldsymbol{\mu_i}$ **is** $\boldsymbol{\bar{x}_i}$ and an **individual 100(1 − α) percent confidence interval for** $\boldsymbol{\mu_i}$ is

$$\left[\bar{x}_i \pm t_{\alpha/2}\sqrt{\frac{MSE}{n_i}}\right]$$

Here, the $t_{\alpha/2}$ point is based on $n - p$ degrees of freedom.

EXAMPLE 12.4 The Oil Company Case: Comparing Gasoline Types

C

Part 1: Using confidence intervals In the gasoline mileage study, we are comparing $p = 3$ treatment means (μ_A, μ_B, and μ_C). Furthermore, each sample is of size $m = 5$, there are a total of $n = 15$ observed gas mileages, and the MSE found in Table 12.3 is .669. Because $q_{.05} = 3.77$ is the entry found in Table A.10 (page 799) corresponding to $p = 3$ and $n - p = 12$, a Tukey simultaneous 95 percent confidence interval for $\mu_B - \mu_A$ is

$$\left[(\bar{x}_B - \bar{x}_A) \pm q_{.05}\sqrt{\frac{MSE}{m}}\right] = \left[(36.56 - 34.92) \pm 3.77\sqrt{\frac{.669}{5}}\right]$$
$$= [1.64 \pm 1.379]$$
$$= [.261, 3.019]$$

Similarly, Tukey simultaneous 95 percent confidence intervals for $\mu_A - \mu_C$ and $\mu_B - \mu_C$ are, respectively,

$$[(\bar{x}_A - \bar{x}_C) \pm 1.379] = [(34.92 - 33.98) \pm 1.379] = [-0.439, 2.319]$$

and

$$[(\bar{x}_B - \bar{x}_C) \pm 1.379] = [(36.56 - 33.98) \pm 1.379] = [1.201, 3.959]$$

These intervals make us simultaneously 95 percent confident that (1) changing from gasoline type A to gasoline type B increases mean mileage by between .261 and 3.019 mpg, (2) changing from gasoline type C to gasoline type A might decrease mean mileage by as much as .439 mpg or might increase mean mileage by as much as 2.319 mpg, *and* (3) changing from gasoline type C to gasoline type B increases mean mileage by between 1.201 and 3.959 mpg. The first and third of these intervals make us 95 percent confident that μ_B is at least .261 mpg greater than μ_A and at least 1.201 mpg greater than μ_C. Therefore, we have strong evidence that gasoline type B yields the highest mean mileage of the gasoline types tested. Furthermore, noting that $t_{.025}$ based on

BI

$n - p = 12$ degrees of freedom is 2.179, it follows that an individual 95 percent confidence interval for μ_B is

$$\left[\bar{x}_B \pm t_{.025}\sqrt{\frac{MSE}{n_B}}\right] = \left[36.56 \pm 2.179\sqrt{\frac{.669}{5}}\right]$$
$$= [35.763, 37.357]$$

This interval says we can be 95 percent confident that the mean mileage obtained by using gasoline type B is between 35.763 and 37.357 mpg. Notice that this confidence interval is graphed on the MINITAB output of Figure 12.2. This output also shows the 95 percent confidence intervals for μ_A and μ_C and gives Tukey simultaneous 95 percent intervals for $\mu_B - \mu_A$, $\mu_C - \mu_A$, and $\mu_C - \mu_B$. Note that the last two Tukey intervals on the output are the "negatives" of the Tukey intervals that we hand calculated for $\mu_A - \mu_C$ and $\mu_B - \mu_C$.

Part 2: Using hypothesis testing (optional) We next consider testing $H_0: \mu_i - \mu_h = 0$ versus $H_a: \mu_i - \mu_h \neq 0$. The test statistic t for performing this test is calculated by dividing $\bar{x}_i - \bar{x}_h$ by $\sqrt{MSE\,[(1/n_i) + (1/n_h)]}$. For example, consider testing $H_0: \mu_B - \mu_A = 0$ versus $H_a: \mu_B - \mu_A \neq 0$. Since $\bar{x}_B - \bar{x}_A = 36.56 - 34.92 = 1.64$ and $\sqrt{MSE\,[(1/n_B) + (1/n_A)]} = \sqrt{.669[(1/5) + (1/5)]} = .5173$, the test statistic t equals $1.64/.5173 = 3.17$. This test statistic value is given in the leftmost portion of the following Excel add-in (MegaStat) output, as is the test statistic value for testing $H_0: \mu_B - \mu_C = 0$ ($t = 4.99$) and the test statistic value for testing $H_0: \mu_A - \mu_C = 0$ ($t = 1.82$):

Tukey simultaneous comparison t-values (d.f. = 12)

		Type C 33.98	Type A 34.92	Type B 36.56
Type C	33.98			
Type A	34.92	1.82		
Type B	36.56	4.99	3.17	

critical values for experimentwise error rate:

0.05	2.67
0.01	3.56

If we wish to use the **Tukey simultaneous comparison procedure** having an experimentwise error rate of α, we reject $H_0: \mu_i - \mu_h = 0$ in favor of $H_a: \mu_i - \mu_h \neq 0$ if the absolute value of t is greater than the critical value $q_\alpha/\sqrt{2}$. Table A.10 (page 799) tells us that $q_{.05}$ is 3.77 and $q_{.01}$ is 5.04. Therefore, the critical values for experimentwise error rates of .05 and .01 are, respectively, $3.77/\sqrt{2} = 2.67$ and $5.04/\sqrt{2} = 3.56$ (see the right portion of the MegaStat output). Suppose we set α equal to .05. Then, since the test statistic value for testing $H_0: \mu_B - \mu_A = 0$ ($t = 3.17$) and the test statistic value for testing $H_0: \mu_B - \mu_C = 0$ ($t = 4.99$) are greater than the critical value 2.67, we reject both null hypotheses. This, along with the fact that $\bar{x}_B = 36.56$ is greater than $\bar{x}_A = 34.92$ and $\bar{x}_C = 33.98$, leads us to conclude that gasoline type B yields the highest mean mileage.

In general, when we use a completely randomized experimental design, it is important to compare the treatments by using experimental units that are essentially the same with respect to the characteristic under study. For example, in the oil company case we have used cars of the same type (Lances) to compare the different gasoline types, and in the supermarket case we have used grocery stores of the same sales potential for the bakery product to compare the shelf display heights (the reader will analyze the data for this case in the exercises). Sometimes, however, it is not possible to use experimental units that are essentially the same with respect to the characteristic under study. One approach to dealing with this situation is to employ **a randomized block design.** This experimental design is discussed in Section 12.3.

To conclude this section, we note that if we fear that the normality and/or equal variances assumptions for one-way analysis of variance do not hold, we can use a nonparametric approach to compare several populations. See Section 18.4 of Chapter 18.

Exercises for Section 12.2

CONCEPTS

connect

12.1 Define the meaning of the terms *response variable, factor, treatments,* and *experimental units.*

12.2 Explain the assumptions that must be satisfied in order to validly use the one-way ANOVA formulas.

FIGURE 12.3 MINITAB Output of a One-Way ANOVA of the Bakery Sales Data in Table 12.2

```
One-way ANOVA: Bakery Sales versus Display Height
                                                                          Tukey 95% Simultaneous
Source          DF       SS       MS       F      P                        Confidence Intervals
Display Height   2  2273.88  1136.94  184.57  0.000
Error           15    92.40     6.16                                      Bottom subtracted from:
Total           17  2366.28                                                         Lower   Center   Upper
                                      Individual 95%                      Middle   17.681   21.400  25.119
                                      CIs For Mean Based on Pooled StDev  Top      -8.019   -4.300  -0.581
Level     N    Mean  StDev  --------+---------+---------+---------+-
Bottom    6  55.800  2.477       (--*-)                                   Middle subtracted from:
Middle    6  77.200  3.103                                  (--*-)                 Lower   Center   Upper
Top       6  51.500  1.648  (-*--)                                        Top    -29.419  -25.700 -21.981
                            --------+---------+---------+---------+-
 Pooled StDev = 2.482            56.0      64.0      72.0      80.0
```

FIGURE 12.4 MINITAB Output of a One-Way ANOVA of the Display Panel Study Data in Table 12.4

```
One-way ANOVA: Time versus Display                                        Tukey 95% Simultaneous
                                                                           Confidence Intervals
Source   DF      SS      MS      F      P
Display   2  500.17  250.08  30.11  0.000                                 A subtracted from:
Error     9   74.75    8.31                                                    Lower  Center   Upper
Total    11  574.92                                                       B   -9.692  -4.000   1.692
                                 Individual 95%                           C    5.558  11.250  16.942
                                 CIs For Mean Based on Pooled StDev
Level   N    Mean  StDev  -+---------+---------+---------+--------
A       4  24.500  2.646        (-----*----)                              B subtracted from:
B       4  20.500  2.646  (----*-----)                                         Lower  Center   Upper
C       4  35.750  3.304                               (-----*----)       C    9.558  15.250  20.942
                          -+---------+---------+---------+--------
Pooled StDev = 2.882     18.0      24.0      30.0      36.0
```

12.3 Explain the difference between the between-treatment variability and the within-treatment variability when performing a one-way ANOVA.

12.4 Explain why we conduct pairwise comparisons of treatment means.

METHODS AND APPLICATIONS

12.5 THE SUPERMARKET CASE DS BakeSale

Consider Example 12.2, and let μ_B, μ_M, and μ_T represent the mean monthly sales when using the bottom, middle, and top shelf display heights, respectively. Figure 12.3 gives the MINITAB output of a one-way ANOVA of the bakery sales study data in Table 12.2 (page 429).

a Test the null hypothesis that μ_B, μ_M, and μ_T are equal by setting $\alpha = .05$. On the basis of this test, can we conclude that the bottom, middle, and top shelf display heights have different effects on mean monthly sales?

b Consider the pairwise differences $\mu_M - \mu_B$, $\mu_T - \mu_B$, and $\mu_T - \mu_M$. Find a point estimate of and a Tukey simultaneous 95 percent confidence interval for each pairwise difference. Interpret the meaning of each interval in practical terms. Which display height maximizes mean sales?

c Find 95 percent confidence intervals for μ_B, μ_M, and μ_T. Interpret each interval.

12.6 A study compared three different display panels for use by air traffic controllers. Each display panel was tested in a simulated emergency condition; 12 highly trained air traffic controllers took part in the study. Four controllers were randomly assigned to each display panel. The time (in seconds) needed to stabilize the emergency condition was recorded. The results of the study are given in Table 12.4. Let μ_A, μ_B, and μ_C represent the mean times to stabilize the emergency condition when using display panels A, B, and C, respectively. Figure 12.4 gives the MINITAB output of a one-way ANOVA of the display panel data DS Display

a Test the null hypothesis that μ_A, μ_B, and μ_C are equal by setting $\alpha = .05$. On the basis of this test, can we conclude that display panels A, B, and C have different effects on the mean time to stabilize the emergency condition?

b Consider the pairwise differences $\mu_B - \mu_A$, $\mu_C - \mu_A$, and $\mu_C - \mu_B$. Find a point estimate of and a Tukey simultaneous 95 percent confidence interval for each pairwise difference. Interpret the

TABLE 12.4 Display Panel Study Data DS Display

Display Panel		
A	*B*	*C*
21	24	40
27	21	36
24	18	35
26	19	32

TABLE 12.5 Bottle Design Study Data DS BottleDes

Bottle Design		
A	*B*	*C*
16	33	23
18	31	27
19	37	21
17	29	28
13	34	25

FIGURE 12.5 Excel Output of a One-Way ANOVA of the Bottle Design Study Data

SUMMARY

Groups	Count	Sum	Average	Variance
DESIGN A	5	83	16.6	5.3
DESIGN B	5	164	32.8	9.2
DESIGN C	5	124	24.8	8.2

ANOVA

Source of Variation	SS	df	MS	F	P-Value	F crit
Between Groups	656.1333	2	328.0667	43.35683	3.23E-06	3.88529
Within Groups	90.8	12	7.566667			
Total	746.9333	14				

FIGURE 12.6 Excel Output of a One-Way ANOVA of the Golf Ball Durability Data

SUMMARY

Groups	Count	Sum	Average	Variance
Alpha	5	1268	253.6	609.3
Best	5	1532	306.4	740.3
Century	5	1209	241.8	469.7
Divot	5	1683	336.6	605.3

ANOVA

Source of Variation	SS	df	MS	F	P-Value	F crit
Between Groups	29860.4	3	9953.4667	16.420798	3.853E-05	3.2388715
Within Groups	9698.4	16	606.15			
Total	39558.8	19				

Tukey simultaneous comparison t-values (d.f. = 16)

		Century 241.8	Alpha 253.6	Best 306.4	Divot 336.6
Century	241.8				
Alpha	253.6	0.76			
Best	306.4	4.15	3.39		
Divot	336.6	6.09	5.33	1.94	

Critical values for experimentwise error rate:

0.05	2.86
0.01	3.67

results by describing the effects of changing from using each display panel to using each of the other panels. Which display panel minimizes the time required to stabilize the emergency condition?

12.7 A consumer preference study compares the effects of three different bottle designs (A, B, and C) on sales of a popular fabric softener. A completely randomized design is employed. Specifically, 15 supermarkets of equal sales potential are selected, and 5 of these supermarkets are randomly assigned to each bottle design. The number of bottles sold in 24 hours at each supermarket is recorded. The data obtained are displayed in Table 12.5. Let μ_A, μ_B, and μ_C represent mean daily sales using bottle designs A, B, and C, respectively. Figure 12.5 gives the Excel output of a one-way ANOVA of the bottle design study data. DS BottleDes

a Test the null hypothesis that μ_A, μ_B, and μ_C are equal by setting $\alpha = .05$. That is, test for statistically significant differences between these treatment means at the .05 level of significance. Based on this test, can we conclude that bottle designs A, B, and C have different effects on mean daily sales?

b Consider the pairwise differences $\mu_B - \mu_A$, $\mu_C - \mu_A$, and $\mu_C - \mu_B$. Find a point estimate of and a Tukey simultaneous 95 percent confidence interval for each pairwise difference. Interpret the results in practical terms. Which bottle design maximizes mean daily sales?

c Find and interpret a 95 percent confidence interval for each of the treatment means μ_A, μ_B, and μ_C.

TABLE 12.6 Golf Ball Durability Test Results DS GolfBall

Brand	
Alpha	**Best**
281	270
220	334
274	307
242	290
251	331
Century	**Divot**
218	364
244	302
225	325
273	337
249	355

12.8 In order to compare the durability of four different brands of golf balls (ALPHA, BEST, CENTURY, and DIVOT), the National Golf Association randomly selects five balls of each brand and places each ball into a machine that exerts the force produced by a 250-yard drive. The number of simulated drives needed to crack or chip each ball is recorded. The results are given in Table 12.6. The Excel output of a one-way ANOVA of these data is shown in Figure 12.6. Test for statistically significant differences between the treatment means μ_{ALPHA}, μ_{BEST}, $\mu_{CENTURY}$, and μ_{DIVOT}. Set $\alpha = .05$. DS GolfBall

12.9 Perform pairwise comparisons of the treatment means in Exercise 12.8 by **(1)** Using Tukey simultaneous 95 percent confidence intervals **(2)** Optionally using t statistics and critical values (see the right side of Figure 12.6 and page 436). Which brands are most durable? Find and interpret a 95 percent confidence interval for each of the treatment means.

12.3 The Randomized Block Design

LO12-3 Compare treatment effects and block effects by using a randomized block design.

Not all experiments employ a completely randomized design. For instance, suppose that when we employ a completely randomized design, we fail to reject the null hypothesis of equality of treatment means because the within-treatment variability (which is measured by the *SSE*) is large. This could happen because differences between the experimental units are concealing true differences between the treatments. We can often remedy this by using what is called a **randomized block design.**

EXAMPLE 12.5 The Cardboard Box Case: Comparing Production Methods

C

The Universal Paper Company manufactures cardboard boxes. The company wishes to investigate the effects of four production methods (methods 1, 2, 3, and 4) on the number of defective boxes produced in an hour. To compare the methods, the company could utilize a completely randomized design. For each of the four production methods, the company would select several (say, as an example, three) machine operators, train each operator to use the production method to which he or she has been assigned, have each operator produce boxes for one hour, and record the number of defective boxes produced. The three operators using any one production method would be *different* from those using any other production method. That is, the completely randomized design would utilize a total of 12 machine operators. However, the abilities of the machine operators could differ substantially. These differences might tend to conceal any real differences between the production methods. To overcome this disadvantage, the company will employ a **randomized block experimental design.** This involves randomly selecting three machine operators and training each operator thoroughly to use all four production methods. Then each operator will produce boxes for one hour using each of the four production methods. The order in which each operator uses the four methods should be random. We record the number of defective boxes produced by each operator using each method. The advantage of the randomized block design is that the defective rates obtained by using the four methods result from employing the *same* three operators. Thus any true differences in the effectiveness of the methods would not be concealed by differences in the operators' abilities.

When Universal Paper employs the randomized block design, it obtains the 12 defective box counts in Table 12.7. We let x_{ij} denote the number of defective boxes produced by machine operator j using production method i. For example, $x_{32} = 5$ says that 5 defective boxes were produced by machine operator 2 using production method 3 (see Table 12.7). In addition to the 12 defective box counts, Table 12.7 gives the sample mean of these 12 observations, which is $\bar{x} = 7.5833$, and also gives **sample treatment means** and **sample block means.** The sample treatment means are the average defective box counts obtained when using production methods 1, 2, 3, and 4. Denoting these sample treatment means as $\bar{x}_{1\bullet}$, $\bar{x}_{2\bullet}$, $\bar{x}_{3\bullet}$, and $\bar{x}_{4\bullet}$, we see from Table 12.7 that $\bar{x}_{1\bullet} = 10.3333$, $\bar{x}_{2\bullet} = 10.3333$, $\bar{x}_{3\bullet} = 5.0$, and $\bar{x}_{4\bullet} = 4.6667$. Because $\bar{x}_{3\bullet}$ and $\bar{x}_{4\bullet}$ are less than $\bar{x}_{1\bullet}$ and $\bar{x}_{2\bullet}$, we estimate that the mean number of defective boxes produced per hour by production method 3 or 4 is less than the mean number of defective boxes produced per hour by production method 1 or 2. The sample block means are the average defective box counts obtained by machine operators 1, 2, and 3. Denoting these sample block means as $\bar{x}_{\bullet 1}$, $\bar{x}_{\bullet 2}$, and $\bar{x}_{\bullet 3}$, we see from Table 12.7 that $\bar{x}_{\bullet 1} = 6.0$, $\bar{x}_{\bullet 2} = 7.75$, and $\bar{x}_{\bullet 3} = 9.0$. Because $\bar{x}_{\bullet 1}$, $\bar{x}_{\bullet 2}$, and $\bar{x}_{\bullet 3}$ differ, we have evidence that the abilities of the machine operators differ and thus that using the machine operators as blocks is reasonable.

TABLE 12.7 Numbers of Defective Cardboard Boxes Obtained by Production Methods 1, 2, 3, and 4 and Machine Operators 1, 2, and 3 DS CardBox

Treatment (Production Method)	Block (Machine Operator) 1	2	3	Sample Treatment Mean
1	9	10	12	10.3333
2	8	11	12	10.3333
3	3	5	7	5.0
4	4	5	5	4.6667
Sample Block Mean	6.0	7.75	9.0	$\bar{x} = 7.5833$

In general, a **randomized block design** compares p treatments (for example, production methods) by using b blocks (for example, machine operators). Each block is used exactly once to measure the effect of each and every treatment. The advantage of the randomized block design over the completely randomized design is that we are comparing the treatments by using the *same* experimental units. Thus any true differences in the treatments will not be concealed by differences in the experimental units.

In order to analyze the data obtained in a randomized block design, we define

x_{ij} = the value of the response variable observed when block j uses treatment i

$\bar{x}_{i\bullet}$ = the mean of the b values of the response variable observed when using treatment i

$\bar{x}_{\bullet j}$ = the mean of the p values of the response variable observed when using block j

$\bar{x}$ = the mean of the total of the bp values of the response variable that we have observed in the experiment

The ANOVA procedure for a randomized block design partitions the **total sum of squares (*SSTO*)** into three components: the **treatment sum of squares (*SST*),** the **block sum of squares (*SSB*),** and the **error sum of squares (*SSE*).** The formula for this partitioning is

$$SSTO = SST + SSB + SSE$$

We define each of these sums of squares and show how they are calculated for the defective cardboard box data as follows (note that $p = 4$ and $b = 3$):

Step 1: Calculate SST, which measures the amount of between-treatment variability:

$$\begin{aligned} SST &= b\sum_{i=1}^{p}(\bar{x}_{i\bullet} - \bar{x})^2 \\ &= 3[(\bar{x}_{1\bullet} - \bar{x})^2 + (\bar{x}_{2\bullet} - \bar{x})^2 + (\bar{x}_{3\bullet} - \bar{x})^2 + (\bar{x}_{4\bullet} - \bar{x})^2] \\ &= 3[(10.3333 - 7.5833)^2 + (10.3333 - 7.5833)^2 \\ &\quad + (5.0 - 7.5833)^2 + (4.6667 - 7.5833)^2] \\ &= 90.9167 \end{aligned}$$

Step 2: Calculate SSB, which measures the amount of variability due to the blocks:

$$\begin{aligned} SSB &= p\sum_{j=1}^{b}(\bar{x}_{\bullet j} - \bar{x})^2 \\ &= 4[(\bar{x}_{\bullet 1} - \bar{x})^2 + (\bar{x}_{\bullet 2} - \bar{x})^2 + (\bar{x}_{\bullet 3} - \bar{x})^2] \\ &= 4[(6.0 - 7.5833)^2 + (7.75 - 7.5833)^2 + (9.0 - 7.5833)^2] \\ &= 18.1667 \end{aligned}$$

Step 3: Calculate $SSTO$, which measures the total amount of variability:

$$\begin{aligned} SSTO &= \sum_{i=1}^{p}\sum_{j=1}^{b}(x_{ij} - \bar{x})^2 \\ &= (9 - 7.5833)^2 + (10 - 7.5833)^2 + (12 - 7.5833)^2 \\ &\quad + (8 - 7.5833)^2 + (11 - 7.5833)^2 + (12 - 7.5833)^2 \\ &\quad + (3 - 7.5833)^2 + (5 - 7.5833)^2 + (7 - 7.5833)^2 \\ &\quad + (4 - 7.5833)^2 + (5 - 7.5833)^2 + (5 - 7.5833)^2 \\ &= 112.9167 \end{aligned}$$

Step 4: Calculate SSE, which measures the amount of variability due to the error:

$$\begin{aligned} SSE &= SSTO - SST - SSB \\ &= 112.9167 - 90.9167 - 18.1667 \\ &= 3.8333 \end{aligned}$$

These sums of squares are shown in Table 12.8, which is the ANOVA table for a randomized block design. This table also gives the degrees of freedom, mean squares, and F statistics used to

TABLE 12.8 Randomized Block ANOVA Table for the Defective Box Data

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Treatments	$p - 1 = 3$	$SST = 90.9167$	$MST = \frac{SST}{p-1} = 30.3056$	$F(\text{treatments}) = \frac{MST}{MSE} = 47.4348$
Blocks	$b - 1 = 2$	$SSB = 18.1667$	$MSB = \frac{SSB}{b-1} = 9.0833$	$F(\text{blocks}) = \frac{MSB}{MSE} = 14.2174$
Error	$(p-1)(b-1) = 6$	$SSE = 3.8333$	$MSE = \frac{SSE}{(p-1)(b-1)} = .6389$	
Total	$pb - 1 = 11$	$SSTO = 112.9167$		

test the hypotheses of interest in a randomized block experiment, as well as the values of these quantities for the defective cardboard box data.

Of main interest is the test of the null hypothesis H_0 that **no differences exist between the treatment effects** on the mean value of the response variable versus the alternative hypothesis H_a that **at least two treatment effects differ.** We can reject H_0 in favor of H_a at level of significance α if F(treatments) is greater than the F_α point based on $p - 1$ numerator and $(p - 1)(b - 1)$ denominator degrees of freedom. In the defective cardboard box case, $F_{.05}$ based on $p - 1 = 3$ numerator and $(p - 1)(b - 1) = 6$ denominator degrees of freedom is 4.76 (see Table A.7, page 796). Because F(treatments) $= 47.4348$ (see Table 12.8) is greater than $F_{.05} = 4.76$, we reject H_0 at the .05 level of significance. Therefore, we have strong evidence that at least two production methods have different effects on the mean number of defective boxes produced per hour.

It is also of interest to test the null hypothesis H_0 that **no differences exist between the block effects** on the mean value of the response variable versus the alternative hypothesis H_a that **at least two block effects differ.** We can reject H_0 in favor of H_a at level of significance α if F(blocks) is greater than the F_α point based on $b - 1$ numerator and $(p - 1)(b - 1)$ denominator degrees of freedom. In the defective cardboard box case, $F_{.05}$ based on $b - 1 = 2$ numerator and $(p - 1)(b - 1) = 6$ denominator degrees of freedom is 5.14 (see Table A.7, page 796). Because F(blocks) $= 14.2174$ (see Table 12.8) is greater than $F_{.05} = 5.14$, we reject H_0 at the .05 level of significance. Therefore, we have strong evidence that at least two machine operators have different effects on the mean number of defective boxes produced per hour.

Figure 12.7 gives the MINITAB and Excel outputs of a randomized block ANOVA of the defective cardboard box data. The p-value of .000 ($<.001$) related to F(treatments) provides extremely strong evidence of differences in production method effects. The p-value of .0053 related to F(blocks) provides very strong evidence of differences in machine operator effects.

If, in a randomized block design, we conclude that at least two treatment effects differ, we can perform pairwise comparisons to determine how they differ.

Point Estimates and Confidence Intervals in a Randomized Block ANOVA

Consider the **difference between the effects of treatments i and h on the mean value of the response variable.** Then:

1. A **point estimate** of this difference is $\bar{x}_{i\cdot} - \bar{x}_{h\cdot}$.
2. A **Tukey simultaneous 100(1 − α) percent confidence interval** for this difference is

$$\left[(\bar{x}_{i\cdot} - \bar{x}_{h\cdot}) \pm q_\alpha \frac{s}{\sqrt{b}}\right]$$

Here the value q_α is obtained from Table A.10 (pages 799–800), which is a table of percentage points of the studentized range. In this table q_α is listed corresponding to values of p and $(p - 1)(b - 1)$.

FIGURE 12.7 MINITAB and Excel Outputs of a Randomized Block ANOVA of the Defective Box Data

(a) The MINITAB Output

Rows: Method Columns: Operator

	1	2	3	All
1	9.000	10.000	12.000	10.333
2	8.000	11.000	12.000	10.333
3	3.000	5.000	7.000	5.000
4	4.000	5.000	5.000	4.667
All	6.000	7.750	9.000	7.583

Method	Mean	
1	10.3333	12
2	10.3333	13
3	5.0000	14
4	4.6667	15

Operator	Mean	
1	6.00	16
2	7.75	17
3	9.00	18

Two-way ANOVA: Rejects versus Method, Operator

Source	DF	SS	MS	F	P
Method	3	90.917 1	30.3056 5	47.43 8	0.000 9
Operator	2	18.167 2	9.0833 6	14.22 10	0.005 11
Error	6	3.833 3	0.6389 7		
Total	11	112.917 4			

(b) The Excel output

ANOVA: Two-Factor Without Replication

Summary	Count	Sum	Average	Variance
Method1	3	31	10.3333 12	2.3333
Method2	3	31	10.3333 13	4.3333
Method3	3	5	5 14	4
Method4	3	14	4.6667 15	0.3333
Operator1	4	24	6 16	8.6667
Operator2	4	31	7.75 17	10.25
Operator3	4	36	9 18	12.6667

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Method	90.9167 1	3	30.3056 5	47.4348 8	0.0001 9	4.7571
Operator	18.1667 2	2	9.0833 6	14.2174 10	0.0053 11	5.1433
Error	3.8333 3	6	0.6389 7			
Total	112.9167 4	11				

1 SST 2 SSB 3 SSE 4 SSTO 5 MST 6 MSB 7 MSE 8 F(treatments) 9 p-value for F(treatments)
10 F(blocks) 11 p-value for F(blocks) 12 $\bar{x}_{1\bullet}$ 13 $\bar{x}_{2\bullet}$ 14 $\bar{x}_{3\bullet}$ 15 $\bar{x}_{4\bullet}$ 16 $\bar{x}_{\bullet 1}$ 17 $\bar{x}_{\bullet 2}$ 18 $\bar{x}_{\bullet 3}$

EXAMPLE 12.6 The Cardboard Box Case: Comparing Production Methods

C

We have previously concluded that we have extremely strong evidence that at least two production methods have different effects on the mean number of defective boxes produced per hour. We have also seen that the sample treatment means are $\bar{x}_{1\bullet} = 10.3333$, $\bar{x}_{2\bullet} = 10.3333$, $\bar{x}_{3\bullet} = 5.0$, and $\bar{x}_{4\bullet} = 4.6667$. Because $\bar{x}_{4\bullet}$ is the smallest sample treatment mean, we will use Tukey simultaneous 95 percent confidence intervals to compare the effect of production method 4 with the effects of production methods 1, 2, and 3. To compute these intervals, we first note that $q_{.05} = 4.90$ is the entry in Table A.10 (page 799) corresponding to $p = 4$ and $(p - 1)(b - 1) = 6$. Also, note that the MSE found in the randomized block ANOVA table is .6389 (see Figure 12.7), which implies that $s = \sqrt{.6389} = .7993$. It follows that a Tukey simultaneous 95 percent confidence interval for the difference between the effects of production methods 4 and 1 on the mean number of defective boxes produced per hour is

$$\left[(\bar{x}_{4\bullet} - \bar{x}_{1\bullet}) \pm q_{.05}\frac{s}{\sqrt{b}}\right] = \left[(4.6667 - 10.3333) \pm 4.90\left(\frac{.7993}{\sqrt{3}}\right)\right]$$
$$= [-5.6666 \pm 2.2615]$$
$$= [-7.9281, -3.4051]$$

Furthermore, it can be verified that a Tukey simultaneous 95 percent confidence interval for the difference between the effects of production methods 4 and 2 on the mean number of defective boxes produced per hour is also [−7.9281, −3.4051]. Therefore, we can be 95 percent confident that changing from production method 1 or 2 to production method 4 decreases the mean number of defective boxes produced per hour by a machine operator by between 3.4051 and 7.9281 boxes. A Tukey simultaneous 95 percent confidence interval for the difference between the effects of production methods 4 and 3 on the mean number of defective boxes produced per hour is

BI

$$[(\bar{x}_{4\cdot} - \bar{x}_{3\cdot}) \pm 2.2615] = [(4.6667 - 5) \pm 2.2615]$$
$$= [-2.5948,\ 1.9282]$$

This interval tells us (with 95 percent confidence) that changing from production method 3 to production method 4 might decrease the mean number of defective boxes produced per hour by as many as 2.5948 boxes or might increase this mean by as many as 1.9282 boxes. In other words, because this interval contains 0, we cannot conclude that the effects of production methods 4 and 3 differ.

Exercises for Section 12.3

CONCEPTS

12.10 In your own words, explain why we sometimes employ the randomized block design.

12.11 Describe what *SSTO*, *SST*, *SSB*, and *SSE* measure.

12.12 How can we test to determine if the blocks we have chosen are reasonable?

METHODS AND APPLICATIONS

12.13 A marketing organization wishes to study the effects of four sales methods on weekly sales of a product. The organization employs a randomized block design in which three salesman use each sales method. The results obtained are given in Figure 12.8, along with the Excel output of a randomized block ANOVA of these data. DS SaleMeth

a Test the null hypothesis H_0 that no differences exist between the effects of the sales methods (treatments) on mean weekly sales. Set $\alpha = .05$. Can we conclude that the different sales methods have different effects on mean weekly sales?

b Test the null hypothesis H_0 that no differences exist between the effects of the salesmen (blocks) on mean weekly sales. Set $\alpha = .05$. Can we conclude that the different salesmen have different effects on mean weekly sales?

c Use Tukey simultaneous 95 percent confidence intervals to make pairwise comparisons of the sales method effects on mean weekly sales. Which sales method(s) maximize mean weekly sales?

FIGURE 12.8 The Sales Method Data and the Excel Output of a Randomized Block ANOVA DS SaleMeth

Sales Method, *i*	Salesman, *j*: *A*	*B*	*C*
1	32	29	30
2	32	30	28
3	28	25	23
4	25	24	23

ANOVA: Two-Factor without Replication

SUMMARY	Count	Sum	Average	Variance
Method 1	3	91	30.3333	2.3333
Method 2	3	90	30	4
Method 3	3	76	25.3333	6.3333
Method 4	3	72	24	1
Salesman A	4	117	29.25	11.5833
Salesman B	4	108	27	8.6667
Salesman C	4	104	26	12.6667

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	93.5833	3	31.1944	36.2258	0.0003	4.7571
Columns	22.1667	2	11.0833	12.8710	0.0068	5.1433
Error	5.1667	6	0.8611			
Total	120.9167	11				

TABLE 12.9 Results of a Bottle Design Experiment DS BottleDes2

Bottle Design, i	Supermarket, j: 1	2	3	4
A	16	14	1	6
B	33	30	19	23
C	23	21	8	12

TABLE 12.10 Results of a Keyboard Experiment DS Keyboard

Data Entry Specialist	Keyboard Brand: *A*	*B*	*C*
1	77	67	63
2	71	62	59
3	74	63	59
4	67	57	54

12.14 A consumer preference study involving three different bottle designs (A, B, and C) for the jumbo size of a new liquid laundry detergent was carried out using a randomized block experimental design, with supermarkets as blocks. Specifically, four supermarkets were supplied with all three bottle designs, which were priced the same. Table 12.9 gives the number of bottles of each design sold in a 24-hour period at each supermarket. If we use these data, SST, SSB, and SSE can be calculated to be 586.1667, 421.6667, and 1.8333, respectively. DS BottleDes2

a Test the null hypothesis H_0 that no differences exist between the effects of the bottle designs on mean daily sales. Set $\alpha = .05$. Can we conclude that the different bottle designs have different effects on mean sales?

b Test the null hypothesis H_0 that no differences exist between the effects of the supermarkets on mean daily sales. Set $\alpha = .05$. Can we conclude that the different supermarkets have different effects on mean sales?

c Use Tukey simultaneous 95 percent confidence intervals to make pairwise comparisons of the bottle design effects on mean daily sales. Which bottle design(s) maximize mean sales?

12.15 To compare three brands of computer keyboards, four data entry specialists were randomly selected. Each specialist used all three keyboards to enter the same kind of text material for 10 minutes, and the number of words entered per minute was recorded. The data obtained are given in Table 12.10. If we use these data, SST, SSB, and SSE can be calculated to be 392.6667, 143.5833, and 2.6667, respectively. DS Keyboard

a Test the null hypothesis H_0 that no differences exist between the effects of the keyboard brands on the mean number of words entered per minute. Set $\alpha = .05$.

b Test the null hypothesis H_0 that no differences exist between the effects of the data entry specialists on the mean number of words entered per minute. Set $\alpha = .05$.

c Use Tukey simultaneous 95 percent confidence intervals to make pairwise comparisons of the keyboard brand effects on the mean number of words entered per minute. Which keyboard brand maximizes the mean number of words entered per minute?

12.16 The Coca-Cola Company introduced New Coke in 1985. Within three months of this introduction, negative consumer reaction forced Coca-Cola to reintroduce the original formula of Coke as Coca-Cola Classic. Suppose that two years later, in 1987, a marketing research firm in Chicago compared the sales of Coca-Cola Classic, New Coke, and Pepsi in public building vending machines. To do this, the marketing research firm randomly selected 10 public buildings in Chicago having both a Coke machine (selling Coke Classic and New Coke) and a Pepsi machine.

FIGURE 12.9 The Coca-Cola Data and a MINITAB Output of a Randomized Block ANOVA of the Data

	Building: 1	2	3	4	5	6	7	8	9	10
Coke Classic	45	136	134	41	146	33	71	224	111	87
New Coke	6	114	56	14	39	20	42	156	61	140
Pepsi	24	90	100	43	51	42	68	131	74	107

Two-way ANOVA: Cans versus Drink, Building

Source	DF	SS	MS	F	P
Drink	2	7997.6	3998.80	5.78	0.011
Building	9	55573.5	6174.83	8.93	0.000
Error	18	12443.7	691.32		
Total	29	76014.8			

Descriptive Statistics: Cans

Variable	Drink	Mean
Cans	Coke Classic	102.8
	New Coke	64.8
	Pepsi	73.0

The data—in number of cans sold over a given period of time—and a MINITAB randomized block ANOVA of the data are given in Figure 12.9. DS Coke

a Test the null hypothesis H_0 that no differences exist between the mean sales of Coca-Cola Classic, New Coke, and Pepsi in Chicago public building vending machines. Set $\alpha = .05$.

b Make pairwise comparisons of the mean sales of Coca-Cola Classic, New Coke, and Pepsi in Chicago public building vending machines by using Tukey simultaneous 95 percent confidence intervals.

c By the mid-1990s the Coca-Cola Company had discontinued making New Coke and had returned to making only its original product. Is there evidence in the 1987 study that this might happen? Explain your answer.

12.4 Two-Way Analysis of Variance

LO12-4 Assess the effects of two factors on a response variable by using a two-way analysis of variance.

Many response variables are affected by more than one factor. Because of this we must often conduct experiments in which we study the effects of several factors on the response. In this section we consider studying the effects of **two factors** on a response variable. To begin, recall that in Example 12.2 we discussed an experiment in which the Tastee Bakery Company investigated the effect of shelf display height on monthly demand for one of its bakery products. This one-factor experiment is actually a simplification of a two-factor experiment carried out by the Tastee Bakery Company. We discuss this two-factor experiment in the following example.

EXAMPLE 12.7 The Supermarket Case: Comparing Display Heights and Widths C

The Tastee Bakery Company supplies a bakery product to many metropolitan supermarkets. The company wishes to study the effects of two factors—**shelf display height** and **shelf display width**—on **monthly demand** (measured in cases of 10 units each) for this product. The factor "display height" is defined to have three levels: B (bottom), M (middle), and T (top). The factor "display width" is defined to have two levels: R (regular) and W (wide). The **treatments** in this experiment are **display height and display width combinations.** These treatments are

$$BR \quad BW \quad MR \quad MW \quad TR \quad TW$$

Here, for example, the notation BR denotes the treatment "bottom display height and regular display width." For each display height and width combination the company randomly selects a sample of $m = 3$ metropolitan area supermarkets (all supermarkets used in the study will be of equal sales potential). Each supermarket sells the product for one month using its assigned display height and width combination, and the month's demand for the product is recorded. The six samples obtained in this experiment are given in Table 12.11 on the next page. We let $x_{ij,k}$ denote the monthly demand obtained at the kth supermarket that used display height i and display width j. For example, $x_{MW,2} = 78.4$ is the monthly demand obtained at the second supermarket that used a middle display height and a wide display.

In addition to giving the six samples, Table 12.11 gives the **sample treatment mean** for each display height and display width combination. For example, $\bar{x}_{BR} = 55.9$ is the mean of the sample of three demands observed at supermarkets using a bottom display height and a regular display width. The table also gives the sample mean demand for each level of display height (B, M, and T) and for each level of display width (R and W). Specifically,

$\bar{x}_{B\bullet} = 55.8$ = the mean of the six demands observed when using a bottom display height

$\bar{x}_{M\bullet} = 77.2$ = the mean of the six demands observed when using a middle display height

$\bar{x}_{T\bullet} = 51.5$ = the mean of the six demands observed when using a top display height

$\bar{x}_{\bullet R} = 60.8$ = the mean of the nine demands observed when using a regular display width

$\bar{x}_{\bullet W} = 62.2$ = the mean of the nine demands observed when using a wide display

Finally, Table 12.11 gives $\bar{x} = 61.5$, which is the overall mean of the total of 18 demands observed in the experiment. Because $\bar{x}_{M\bullet} = 77.2$ is considerably larger than $\bar{x}_{B\bullet} = 55.8$ and $\bar{x}_{T\bullet} = 51.5$, we estimate that mean monthly demand is highest when using a middle display height. Because $\bar{x}_{\bullet R} = 60.8$ and $\bar{x}_{\bullet W} = 62.2$ do not differ by very much, we estimate there is little difference between the effects of a regular display width and a wide display on mean monthly demand.

TABLE 12.11 Six Samples of Monthly Demands for a Bakery Product DS BakeSale2

	Display Width		
Display Height	***R***	***W***	
B	58.2	55.7	
	53.7	52.5	
	55.8	58.9	
	$\bar{x}_{BR} = 55.9$	$\bar{x}_{BW} = 55.7$	$\bar{x}_{B\cdot} = 55.8$
M	73.0	76.2	
	78.1	78.4	
	75.4	82.1	
	$\bar{x}_{MR} = 75.5$	$\bar{x}_{MW} = 78.9$	$\bar{x}_{M\cdot} = 77.2$
T	52.4	54.0	
	49.7	52.1	
	50.9	49.9	
	$\bar{x}_{TR} = 51.0$	$\bar{x}_{TW} = 52.0$	$\bar{x}_{T\cdot} = 51.5$
	$\bar{x}_{\cdot R} = 60.8$	$\bar{x}_{\cdot W} = 62.2$	$\bar{x} = 61.5$

FIGURE 12.10 Graphical Analysis of the Bakery Demand Data

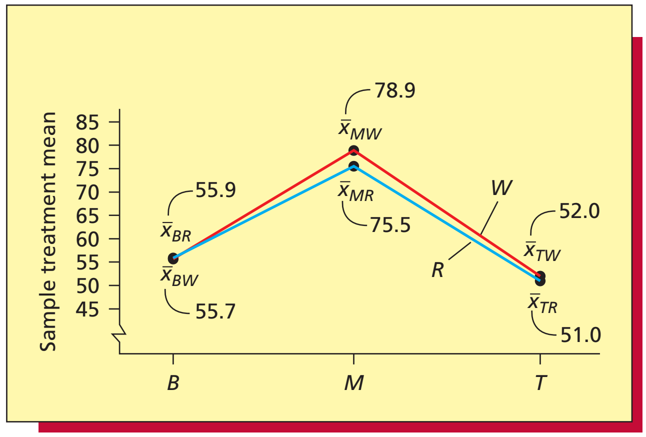
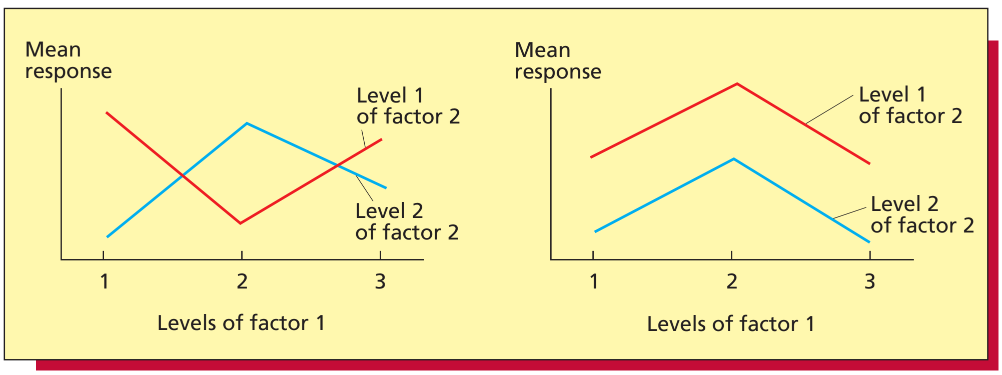


Figure 12.10 presents a graphical analysis of the bakery demand data. In this figure we plot, for each display width (*R* and *W*), the change in the sample treatment mean demand associated with changing the display height from bottom (*B*) to middle (*M*) to top (*T*). Note that, for either a regular display width (*R*) or a wide display (*W*), the middle display height (*M*) gives the highest mean monthly demand. Also, note that, for either a bottom, middle, or top display height, there is little difference between the effects of a regular display width and a wide display on mean monthly demand. This sort of graphical analysis is useful for determining whether a condition called **interaction** exists. In general, for two factors that might affect a response variable, we say that **interaction exists if the relationship between the mean response and one factor depends on the other factor.** This is clearly true in the leftmost figure below:

LO12-5 Describe what happens when two factors interact.



Specifically, this figure shows that at levels 1 and 3 of factor 1, level 1 of factor 2 gives the highest mean response, while at level 2 of factor 1, level 2 of factor 2 gives the highest mean response. On the other hand, the **parallel** line plots in the rightmost figure indicate a lack of interaction between factors 1 and 2. Because the sample mean plots in Figure 12.10 look nearly parallel, we might intuitively conclude that there is little or no interaction between display height and display width.

Suppose we wish to study the effects of two factors on a response variable. We assume that the first factor, which we refer to as **factor 1,** has ***a*** **levels** (levels 1, 2, . . . , *a*). Further, we assume that the second factor, which we will refer to as **factor 2,** has ***b*** **levels** (levels 1, 2, . . . , *b*). Here a **treatment** is considered to be a **combination of a level of factor 1 and a level of factor 2.** It follows that there are a total of *ab* treatments, and we assume that we will employ a *completely randomized experimental design* in which we will assign *m* randomly selected experimental units to each treatment. This procedure results in our observing *m* values of the response variable for each of the *ab* treatments, and in this case we say that we are performing a **two-factor factorial experiment.**

In addition to graphical analysis, **two-way analysis of variance (two-way ANOVA)** is a useful tool for analyzing the data from a two-factor factorial experiment. To explain the ANOVA approach for analyzing such an experiment, we define

$x_{ij,k}$ = the kth value of the response variable observed when using level i of factor 1 and level j of factor 2

$\bar{x}_{ij}$ = the mean of the m values observed when using the ith level of factor 1 and the jth level of factor 2

$\bar{x}_{i\bullet}$ = the mean of the bm values observed when using the ith level of factor 1

$\bar{x}_{\bullet j}$ = the mean of the am values observed when using the jth level of factor 2

$\bar{x}$ = the mean of the abm values that we have observed in the experiment

The ANOVA procedure for a two-factor factorial experiment partitions the **total sum of squares (*SSTO*)** into four components: the **factor 1 sum of squares–*SS*(1),** the **factor 2 sum of squares–*SS*(2),** the **interaction sum of squares–*SS*(int),** and the **error sum of squares–*SSE*.** The formula for this partitioning is as follows:

$$SSTO = SS(1) + SS(2) + SS(\text{int}) + SSE$$

We define each of these sums of squares and show how they are calculated for the bakery demand data as follows (note that $a = 3$, $b = 2$, and $m = 3$):

Step 1: Calculate *SSTO*, which measures the total amount of variability:

$$\begin{aligned} SSTO &= \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{m}(x_{ij,k} - \bar{x})^2 \\ &= (58.2 - 61.5)^2 + (53.7 - 61.5)^2 + \cdots + (49.9 - 61.5)^2 = 2{,}366.28 \end{aligned}$$

Step 2: Calculate *SS*(1), which measures the amount of variability due to the different levels of factor 1:

$$\begin{aligned} SS(1) &= bm\sum_{i=1}^{a}(\bar{x}_{i\bullet} - \bar{x})^2 \\ &= 2 \cdot 3[(\bar{x}_{B\bullet} - \bar{x})^2 + (\bar{x}_{M\bullet} - \bar{x})^2 + (\bar{x}_{T\bullet} - \bar{x})^2] \\ &= 6[(55.8 - 61.5)^2 + (77.2 - 61.5)^2 + (51.5 - 61.5)^2] = 2{,}273.88 \end{aligned}$$

Step 3: Calculate *SS*(2), which measures the amount of variability due to the different levels of factor 2:

$$\begin{aligned} SS(2) &= am\sum_{j=1}^{b}(\bar{x}_{\bullet j} - \bar{x})^2 \\ &= 3 \cdot 3[(\bar{x}_{\bullet R} - \bar{x})^2 + (\bar{x}_{\bullet W} - \bar{x})^2] \\ &= 9[(60.8 - 61.5)^2 + (62.2 - 61.5)^2] = 8.82 \end{aligned}$$

Step 4: Calculate *SS*(int), which measures the amount of variability due to the interaction between factors 1 and 2:

$$\begin{aligned} SS(\text{int}) &= m\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}_{ij} - \bar{x}_{i\bullet} - \bar{x}_{\bullet j} + \bar{x})^2 \\ &= 3[(\bar{x}_{BR} - \bar{x}_{B\bullet} - \bar{x}_{\bullet R} + \bar{x})^2 + (\bar{x}_{BW} - \bar{x}_{B\bullet} - \bar{x}_{\bullet W} + \bar{x})^2 \\ &\quad + (\bar{x}_{MR} - \bar{x}_{M\bullet} - \bar{x}_{\bullet R} + \bar{x})^2 + (\bar{x}_{MW} - \bar{x}_{M\bullet} - \bar{x}_{\bullet W} + \bar{x})^2 \\ &\quad + (\bar{x}_{TR} - \bar{x}_{T\bullet} - \bar{x}_{\bullet R} + \bar{x})^2 + (\bar{x}_{TW} - \bar{x}_{T\bullet} - \bar{x}_{\bullet W} + \bar{x})^2] \\ &= 3[(55.9 - 55.8 - 60.8 + 61.5)^2 + (55.7 - 55.8 - 62.2 + 61.5)^2 \\ &\quad + (75.5 - 77.2 - 60.8 + 61.5)^2 + (78.9 - 77.2 - 62.2 + 61.5)^2 \\ &\quad + (51.0 - 51.5 - 60.8 + 61.5)^2 + (52.0 - 51.5 - 62.2 + 61.5)^2] = 10.08 \end{aligned}$$

TABLE 12.12 Two-Way ANOVA Table for the Bakery Demand Data

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F
Factor 1	$a - 1 = 2$	$SS(1) = 2{,}273.88$	$MS(1) = \frac{SS(1)}{a-1} = 1136.94$	$F(1) = \frac{MS(1)}{MSE} = 185.6229$
Factor 2	$b - 1 = 1$	$SS(2) = 8.82$	$MS(2) = \frac{SS(2)}{b-1} = 8.82$	$F(2) = \frac{MS(2)}{MSE} = 1.44$
Interaction	$(a-1)(b-1) = 2$	$SS(\text{int}) = 10.08$	$MS(\text{int}) = \frac{SS(\text{int})}{(a-1)(b-1)} = 5.04$	$F(\text{int}) = \frac{MS(\text{int})}{MSE} = .8229$
Error	$ab(m-1) = 12$	$SSE = 73.50$	$MSE = \frac{SSE}{ab(m-1)} = 6.125$	
Total	$abm - 1 = 17$	$SSTO = 2{,}366.28$		

Step 5: Calculate SSE, which measures the amount of variability due to the error:

$$
\begin{aligned}
SSE &= SSTO - SS(1) - SS(2) - SS(\text{int}) \\
&= 2{,}366.28 - 2{,}273.88 - 8.82 - 10.08 = 73.50
\end{aligned}
$$

These sums of squares are shown in Table 12.12, which is called a **two-way analysis of variance (ANOVA) table.** This table also gives the degrees of freedom, mean squares, and F statistics used to test the hypotheses of interest in a two-factor factorial experiment, as well as the values of these quantities for the shelf display data.

We first test the null hypothesis $\boldsymbol{H_0}$ that **no interaction exists between factors 1 and 2** versus the alternative hypothesis $\boldsymbol{H_a}$ that **interaction does exist.** We can reject H_0 in favor of H_a at level of significance α if $F(\text{int})$ is greater than the F_α point based on $(a - 1)(b - 1)$ numerator and $ab(m - 1)$ denominator degrees of freedom. In the supermarket case, $F_{.05}$ based on $(a - 1)(b - 1) = 2$ numerator and $ab(m - 1) = 12$ denominator degrees of freedom is 3.89 (see Table A.7, page 796). Because $F(\text{int}) = .8229$ (see Table 12.12) is less than $F_{.05} = 3.89$, we cannot reject H_0 at the .05 level of significance. We conclude that little or no interaction exists between shelf display height and shelf display width. That is, we conclude that the relationship between mean demand for the bakery product and shelf display height depends little (or not at all) on the shelf display width. Further, we conclude that the relationship between mean demand and shelf display width depends little (or not at all) on the shelf display height. Therefore, we can test the significance of each factor separately.

To test the significance of factor 1, we test the null hypothesis $\boldsymbol{H_0}$ **that no differences exist between the effects of the different levels of factor 1** on the mean response versus the alternative hypothesis $\boldsymbol{H_a}$ **that at least two levels of factor 1 have different effects.** We can reject H_0 in favor of H_a at level of significance α if $F(1)$ is greater than the F_α point based on $a - 1$ numerator and $ab(m - 1)$ denominator degrees of freedom. In the supermarket case, $F_{.05}$ based on $a - 1 = 2$ numerator and $ab(m - 1) = 12$ denominator degrees of freedom is 3.89. Because $F(1) = 185.6229$ (see Table 12.12) is greater than $F_{.05} = 3.89$, we can reject H_0 at the .05 level of significance. Therefore, we have strong evidence that at least two of the bottom, middle, and top display heights have different effects on mean monthly demand.

To test the significance of factor 2, we test the null hypothesis $\boldsymbol{H_0}$ **that no differences exist between the effects of the different levels of factor 2** on the mean response versus the alternative hypothesis $\boldsymbol{H_a}$ **that at least two levels of factor 2 have different effects.** We can reject H_0 in favor of H_a at level of significance α if $F(2)$ is greater than the F_α point based on $b - 1$ numerator and $ab(m - 1)$ denominator degrees of freedom. In the supermarket case, $F_{.05}$ based on $b - 1 = 1$ numerator and $ab(m - 1) = 12$ denominator degrees of freedom is 4.75. Because $F(2) = 1.44$ (see Table 12.12) is less than $F_{.05} = 4.75$, we cannot reject H_0 at the .05 level of significance. Therefore, we do not have strong evidence that the regular display width and the wide display have different effects on mean monthly demand.

FIGURE 12.11 MINITAB and Excel Outputs of a Two-Way ANOVA of the Bakery Demand Data

(a) The MINITAB Output

Rows : Height Columns : Width
Cell Contents : Demand : Mean

	Regular	Wide	All
Bottom	55.90	55.70	55.80
Middle	75.50	78.90	77.20
Top	51.00	52.00	51.50
All	60.80	62.20	61.50

Height	Mean
Bottom	55.8 $\boxed{16}$
Middle	77.2 $\boxed{17}$
Top	51.5 $\boxed{18}$

Width	Mean
Regular	60.8 $\boxed{19}$
Wide	62.2 $\boxed{20}$

Two-way ANOVA: Demand versus Height, Width

Source	DF	SS	MS	F	P
Height	2	2273.88 $\boxed{1}$	1136.94 $\boxed{6}$	185.62 $\boxed{10}$	0.000 $\boxed{11}$
Width	1	8.82 $\boxed{2}$	8.82 $\boxed{7}$	1.44 $\boxed{12}$	0.253 $\boxed{13}$
Interaction	2	10.08 $\boxed{3}$	5.04 $\boxed{8}$	0.82 $\boxed{14}$	0.462 $\boxed{15}$
Error	12	73.50 $\boxed{4}$	6.12 $\boxed{9}$		
Total	17	2366.28 $\boxed{5}$			

(b) The Excel Output

ANOVA: Two-Factor With Replication

SUMMARY	Regular	Wide	Total
Bottom			
Count	3	3	6
Sum	167.7	167.1	334.8
Average	55.9	55.7	55.8 $\boxed{16}$
Variance	5.07	10.24	6.136
Middle			
Count	3	3	6
Sum	226.5	236.7	463.2
Average	75.5	78.9	77.2 $\boxed{17}$
Variance	6.51	8.89	9.628
Top			
Count	3	3	6
Sum	153.0	156.0	309.0
Average	51.0	52.0	51.5 $\boxed{18}$
Variance	1.8	4.2	2.7
Total			
Count	9	9	
Sum	547.2	559.8	
Average	60.8 $\boxed{19}$	62.2 $\boxed{20}$	
Variance	129.405	165.277	

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Height	2273.88 $\boxed{1}$	2	1136.94 $\boxed{6}$	185.6229 $\boxed{10}$	0.0000 $\boxed{11}$	3.8853
Width	8.82 $\boxed{2}$	1	8.82 $\boxed{7}$	1.4400 $\boxed{12}$	0.2533 $\boxed{13}$	4.7472
Interaction	10.08 $\boxed{3}$	2	5.04 $\boxed{8}$	0.8229 $\boxed{14}$	0.4625 $\boxed{15}$	3.8853
Within	73.5 $\boxed{4}$	12	6.125 $\boxed{9}$			
Total	2366.28 $\boxed{5}$	17				

$\boxed{1}$ SS(1) $\boxed{2}$ SS(2) $\boxed{3}$ SS(int) $\boxed{4}$ SSE $\boxed{5}$ SSTO $\boxed{6}$ MS(1) $\boxed{7}$ MS(2) $\boxed{8}$ MS(int) $\boxed{9}$ MSE $\boxed{10}$ $F(1)$ $\boxed{11}$ p-value for $F(1)$ $\boxed{12}$ $F(2)$ $\boxed{13}$ p-value for $F(2)$ $\boxed{14}$ $F(\text{int})$ $\boxed{15}$ p-value for $F(\text{int})$ $\boxed{16}$ $\bar{x}_{B\cdot}$ $\boxed{17}$ $\bar{x}_{M\cdot}$ $\boxed{18}$ $\bar{x}_{T\cdot}$ $\boxed{19}$ $\bar{x}_{\cdot R}$ $\boxed{20}$ $\bar{x}_{\cdot W}$

Noting that Figure 12.11 gives MINITAB and Excel outputs of a two-way ANOVA for the bakery demand data, we next discuss how to make pairwise comparisons.

Point Estimates and Confidence Intervals in Two-Way ANOVA

1 Consider the **difference between the effects of levels *i* and *i′* of factor 1 on the mean value of the response variable.**

- **a** A **point estimate** of this difference is $\bar{x}_{i\bullet} - \bar{x}_{i'\bullet}$
- **b** A **Tukey simultaneous 100(1 − α) percent confidence interval** for this difference (in the set of all possible paired differences between the effects of the different levels of factor 1) is

$$\left[(\bar{x}_{i\bullet} - \bar{x}_{i'\bullet}) \pm q_\alpha\sqrt{MSE\left(\frac{1}{bm}\right)}\right]$$

where q_α is obtained from Table A.10 (pages 799–800), which is a table of percentage points of the studentized range. Here q_α is listed corresponding to values of a and $ab(m-1)$.

2 Consider the **difference between the effects of levels *j* and *j′* of factor 2 on the mean value of the response variable.**

- **a** A **point estimate** of this difference is $\bar{x}_{\bullet j} - \bar{x}_{\bullet j'}$
- **b** A **Tukey simultaneous 100(1 − α) percent confidence interval** for this difference (in the set of all possible paired differences between the effects of the different levels of factor 2) is

$$\left[(\bar{x}_{\bullet j} - \bar{x}_{\bullet j'}) \pm q_\alpha\sqrt{MSE\left(\frac{1}{am}\right)}\right]$$

where q_α is obtained from Table A.10 and is listed corresponding to values of b and $ab(m-1)$.

3 Let μ_{ij} denote the **mean value of the response variable obtained when using level *i* of factor 1 and level *j* of factor 2.** A **point estimate of μ_{ij} is $\bar{x}_{ij}$,** and an **individual 100(1 − α) percent confidence interval** for μ_{ij} is

$$\left[\bar{x}_{ij} \pm t_{\alpha/2}\sqrt{\frac{MSE}{m}}\right]$$

where the $t_{\alpha/2}$ point is based on $ab(m-1)$ degrees of freedom.

EXAMPLE 12.8 The Supermarket Case: Comparing Display Heights and Widths C

We have previously concluded that at least two of the bottom, middle, and top display heights have different effects on mean monthly demand. Because $\bar{x}_{M\bullet} = 77.2$ is greater than $\bar{x}_{B\bullet} = 55.8$ and $\bar{x}_{T\bullet} = 51.5$, we will use Tukey simultaneous 95 percent confidence intervals to compare the effect of a middle display height with the effects of the bottom and top display heights. To compute these intervals, we first note that $q_{.05} = 3.77$ is the entry in Table A.10 (page 799) corresponding to $a = 3$ and $ab(m-1) = 12$. Also note that the *MSE* found in the two-way ANOVA table is 6.125 (see Table 12.12 on page 448). It follows that a Tukey simultaneous 95 percent confidence interval for the difference between the effects of a middle and bottom display height on mean monthly demand is

$$\left[(\bar{x}_{M\bullet} - \bar{x}_{B\bullet}) \pm q_{.05}\sqrt{MSE\left(\frac{1}{bm}\right)}\right] = \left[(77.2 - 55.8) \pm 3.77\sqrt{6.125\left(\frac{1}{2(3)}\right)}\right]$$
$$= [21.4 \pm 3.8091]$$
$$= [17.5909, 25.2091]$$

This interval says we are 95 percent confident that changing from a bottom display height to a middle display height will increase the mean demand for the bakery product by between 17.5909 and 25.2091 cases per month. Similarly, a Tukey simultaneous 95 percent confidence interval for the difference between the effects of a middle and top display height on mean monthly demand is

$$[(\bar{x}_{M\bullet} - \bar{x}_{T\bullet}) \pm 3.8091] = [(77.2 - 51.5) \pm 3.8091]$$
$$= [21.8909, 29.5091]$$

This interval says we are 95 percent confident that changing from a top display height to a middle display height will increase mean demand for the bakery product by between 21.8909 and 29.5091 cases per month. Together, these intervals make us 95 percent confident that a middle shelf display height is, on average, at least 17.5909 cases sold per month better than a bottom shelf display height and at least 21.8909 cases sold per month better than a top shelf display height.

BI

Next, recall that previously conducted *F*-tests suggest that there is little or no interaction between display height and display width and that there is little difference between using a regular display width and a wide display. However, noting that $\bar{x}_{MW} = 78.9$ is slightly larger than $\bar{x}_{MR} = 75.5$, we now find an individual 95 percent confidence interval for μ_{MW}, the mean demand obtained when using a middle display height and a wide display:

$$\left[\bar{x}_{MW} \pm t_{.025}\sqrt{\frac{MSE}{m}}\right] = \left[78.9 \pm 2.179\sqrt{\frac{6.125}{3}}\right]$$
$$= [75.7865, 82.0135]$$

Here $t_{.025} = 2.179$ is based on $ab(m - 1) = 12$ degrees of freedom. This interval says that, when we use a middle display height and a wide display, we can be 95 percent confident that mean demand for the bakery product will be between 75.7865 and 82.0135 cases per month.

If we conclude that (substantial) interaction exists between factors 1 and 2, the effects of changing the level of one factor will depend on the level of the other factor. In this case, we cannot analyze the levels of the two factors separately. One simple alternative procedure is to use one-way ANOVA (see Section 12.2) to compare all of the treatment means (the μ_{ij}'s) with the possible purpose of finding the best combination of levels of factors 1 and 2.

Exercises for Section 12.4

CONCEPTS

12.17 What is a treatment in the context of a two-factor factorial experiment?

12.18 Explain what we mean when we say that interaction exists between two factors.

METHODS AND APPLICATIONS

12.19 A study compared three display panels used by air traffic controllers. Each display panel was tested for four different simulated emergency conditions. Twenty-four highly trained air traffic controllers were used in the study. Two controllers were randomly assigned to each display panel–emergency condition combination. The time (in seconds) required to stabilize the emergency condition was recorded. Figure 12.12 gives the resulting data and the MINITAB output of a two-way ANOVA of the data. DS Display2

- **a** Interpret the interaction plot in Figure 12.12. Then test for interaction with $\alpha = .05$.
- **b** Test the significance of display panel effects with $\alpha = .05$.
- **c** Test the significance of emergency condition effects with $\alpha = .05$.
- **d** Make pairwise comparisons of display panels A, B, and C by using Tukey simultaneous 95 percent confidence intervals.
- **e** Make pairwise comparisons of emergency conditions 1, 2, 3, and 4 by using Tukey simultaneous 95 percent confidence intervals.
- **f** Which display panel minimizes the time required to stabilize an emergency condition? Does your answer depend on the emergency condition? Why?
- **g** Calculate a 95 percent (individual) confidence interval for the mean time required to stabilize emergency condition 4 using display panel B.

12.20 A telemarketing firm has studied the effects of two factors on the response to its television advertisements. The first factor is the time of day at which the ad is run, while the second is the position of the ad within the hour. The data in Figure 12.13, which were obtained by using a completely randomized experimental design, give the number of calls placed to an 800 number following a sample broadcast of the advertisement. If we use Excel to analyze these data, we obtain the output in Figure 12.13. DS TelMktResp

- **a** Perform graphical analysis to check for interaction between time of day and position of advertisement. Explain your conclusion. Then test for interaction with $\alpha = .05$.
- **b** Test the significance of time of day effects with $\alpha = .05$.
- **c** Test the significance of position of advertisement effects with $\alpha = .05$.
- **d** Make pairwise comparisons of the morning, afternoon, and evening times by using Tukey simultaneous 95 percent confidence intervals.
- **e** Make pairwise comparisons of the four ad positions by using Tukey simultaneous 95 percent confidence intervals.
- **f** Which time of day and advertisement position maximizes consumer response? Compute a 95 percent (individual) confidence interval for the mean number of calls placed for this time of day/ad position combination.

FIGURE 12.12 The Display Panel Data and the MINITAB Output of a Two-Way ANOVA DS Display2

	Emergency Condition			
Display Panel	1	2	3	4
A	17	25	31	14
	14	24	34	13
B	15	22	28	9
	12	19	31	10
C	21	29	32	15
	24	28	37	19

Two-way ANOVA: Time versus Panel, Condition

Source	DF	SS	MS	F	P
Panel	2	218.58	109.292	26.49	0.000
Condition	3	1247.46	415.819	100.80	0.000
Interaction	6	16.42	2.736	0.66	0.681
Error	12	49.50	4.125		
Total	23	1531.96			

Tabulated statistics: Panel, Condition

Rows: Panel Columns: Condition

	1	2	3	4	All
A	15.50	24.50	32.50	13.50	21.50
B	13.50	20.50	29.50	9.50	18.25
C	22.50	28.50	34.50	17.00	25.63
All	17.17	24.50	32.17	13.33	21.79

Cell Contents: Time : Mean

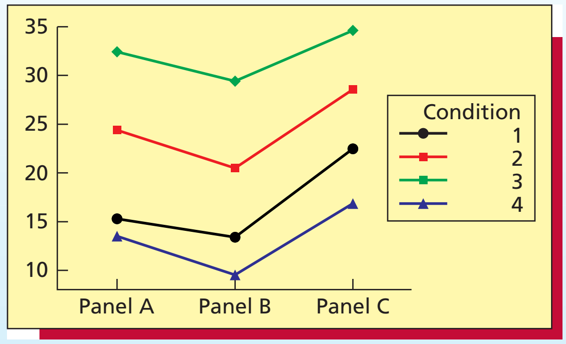
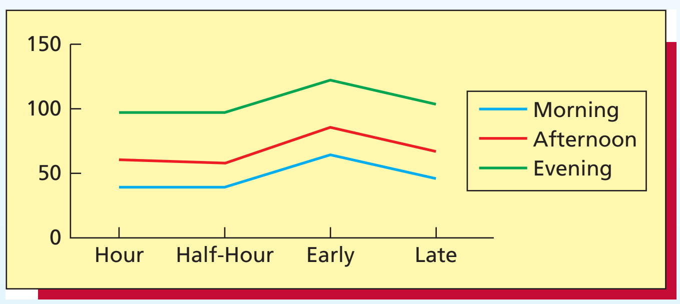


FIGURE 12.13 The Telemarketing Data and the Excel Output of a Two-Way ANOVA DS TelMktResp

	Position of Advertisement			
Time of Day	On the Hour	On the Half-Hour	Early in Program	Late in Program
10:00 morning	42	36	62	51
	37	41	68	47
	41	38	64	48
4:00 afternoon	62	57	88	67
	60	60	85	60
	58	55	81	66
9:00 evening	100	97	127	105
	96	96	120	101
	103	101	126	107

ANOVA: Two-Factor With Replication

Summary	Hour	Half-Hour	Early	Late	Total
Morning					
Count	3	3	3	3	12
Sum	120	115	194	146	575
Average	40	38.3	64.7	48.7	47.9
Variance	7	6.3	9.3	4.3	123.7
Afternoon					
Count	3	3	3	3	12
Sum	180	172	254	193	799
Average	60	57.3	84.7	64.3	66.6
Variance	4	6.3	12.3	14.3	132.4
Evening					
Count	3	3	3	3	12
Sum	299	294	373	313	1279
Average	99.67	98	124.3	104.3	106.6
Variance	12.33	7	14.3	9.3	128.3
Total					
Count	9	9	9	9	
Sum	599	581	821	652	
Average	66.56	64.56	91.22	72.44	
Variance	697.53	701.78	700.69	625.03	



ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	21560.89	2	10780.444	1209.02	8.12E-25	3.403
Columns	3989.42	3	1329.B06	149.14	1.19E-15	3.009
Interaction	25.33	6	4.222	0.47	0.8212	2.508
Within	214	24	8.917			
Total	25789.64	35				

Chapter Summary

We began this chapter by introducing some basic concepts of **experimental design.** We saw that we carry out an experiment by setting the values of one or more **factors** before the values of the **response variable** are observed. The different values (or levels) of a factor are called **treatments,** and the purpose of most experiments is to compare and estimate the effects of the various treatments on the response variable. We saw that the different treatments are assigned to **experimental units,** and we discussed the **completely randomized experimental design.** This design assigns independent, random samples of experimental units to the treatments.

We began studying how to analyze experimental data by discussing **one-way analysis of variance (one-way ANOVA).** Here we study how one factor (having p levels) affects the response variable. In particular, we learned how to use this methodology to test for differences between the **treatment means** and to estimate the size of pairwise differences between the treatment means.

Sometimes, even if we randomly select the experimental units, differences between the experimental units conceal differences between the treatments. In such a case, we learned that we can employ a **randomized block design.** Each **block** (experimental unit or set of experimental units) is used exactly once to measure the effect of each and every treatment. Because we are comparing the treatments by using the same experimental units, any true differences between the treatments will not be concealed by differences between the experimental units.

The last technique we studied in this chapter was **two-way analysis of variance (two-way ANOVA).** Here we study the effects of two factors by carrying out a **two-factor factorial experiment.** If there is little or no interaction between the two factors, then we are able to study the significance of each of the two factors separately. On the other hand, if substantial interaction exists between the two factors, we study the nature of the differences between the treatment means.

Glossary of Terms

analysis of variance table: A table that summarizes the sums of squares, mean squares, F statistic(s), and p-value(s) for an analysis of variance. (pages 434, 441, and 448)
completely randomized experimental design: An experimental design in which independent, random samples of experimental units are assigned to the treatments. (page 428)
experimental units: The entities (objects, people, and so on) to which the treatments are assigned. (page 427)
factor: A variable that might influence the response variable; an independent variable. (page 427)
interaction: When the relationship between the mean response and one factor depends on the level of the other factor. (page 446)
one-way ANOVA: A method used to estimate and compare the effects of the different levels of a single factor on a response variable. (page 429)
randomized block design: An experimental design that compares p treatments by using b blocks (experimental units or sets of experimental units). Each block is used exactly once to measure the effect of each and every treatment. (page 439)
replication: When a treatment is applied to more than one experimental unit. (page 427)
response variable: The variable of interest in an experiment; the dependent variable. (page 427)
treatment: A value (or level) of a factor (or combination of factors). (page 427)
treatment mean: The mean value of the response variable obtained by using a particular treatment. (page 429)
two-factor factorial experiment: An experiment in which we randomly assign m experimental units to each combination of levels of two factors. (page 446)
two-way ANOVA: A method used to study the effects of two factors on a response variable. (page 447)

Important Formulas and Tests

One-way ANOVA sums of squares: pages 431–432
One-way ANOVA F-test: page 432
One-way ANOVA table: page 434
Estimation in one-way ANOVA: page 435
Randomized block sums of squares: page 440
Randomized block ANOVA table: page 441
Estimation in a randomized block ANOVA: page 441
Two-way ANOVA sums of squares: pages 447–448
Two-way ANOVA table: page 448
Estimation in two-way ANOVA: page 450

Supplementary Exercises

12.21 An experiment is conducted to study the effects of two sales approaches—high-pressure (H) and low-pressure (L)—and to study the effects of two sales pitches (1 and 2) on the weekly sales of a product. The data in Table 12.13 on the next page are obtained by using a completely randomized

TABLE 12.13 Results of the Sales Approach Experiment DS SaleMeth2

Sales Pressure	Sales Pitch 1	Sales Pitch 2
H	32	32
	29	30
	30	28
L	28	25
	25	24
	23	23

TABLE 12.14 Reduction of Cholesterol Levels DS CholRed

Drug *X*	Drug *Y*	Drug *Z*
22	40	15
31	35	9
19	47	14
27	41	11
25	39	21
18	33	5

TABLE 12.15 Results of the House Profitability Study DS HouseProf

Foreman	House Design *A*		*B*		*C*	
1	10.2	10.65	12.2	11.95	19.4	18.80
	11.1		11.7		18.2	
2	9.7	10.25	11.6	11.80	13.6	13.15
	10.8		12.0		12.7	

FIGURE 12.14 Excel Output of a Two-Way ANOVA of the Sales Approach Data

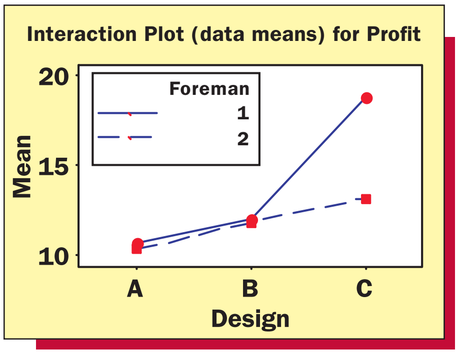
ANOVA: Two-Factor With Replication

SUMMARY	Pitch 1	Pitch 2	Total
High Pressure			
Count	3	3	6
Sum	91	90	181
Average	30.3333	30	30.1667
Variance	2.3333	4	2.5667
Low Pressure			
Count	3	3	6
Sum	76	72	148
Average	25.3333	24	24.6667
Variance	6.3333	1	3.4667
Total			
Count	6	6	
Sum	167	162	
Average	27.8333	27	
Variance	10.9667	12.8	

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Pressure	90.75	1	90.75	26.5610	0.0009	5.3177
Pitch	2.0833	1	2.0833	0.6098	0.4574	5.3177
Interaction	0.75	1	0.75	0.2195	0.6519	5.3177
Within	27.3333	8	3.4167			
Total	120.917	11				

FIGURE 12.15 An Interaction Plot for the House Profitability Data



design, and Figure 12.14 gives the Excel output of a two-way ANOVA of the sales experiment data. DS SaleMeth2

a Perform graphical analysis to check for interaction between sales pressure and sales pitch.
b Test for interaction by setting $\alpha = .05$.
c Test for differences in the effects of the levels of sales pressure by setting $\alpha = .05$.
d Test for differences in the effects of the levels of sales pitch by setting $\alpha = .05$.

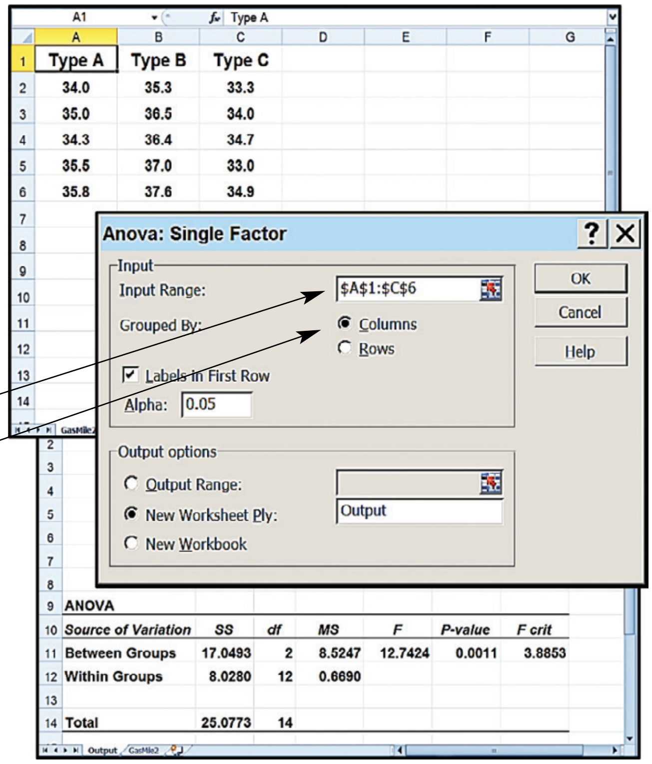
12.22 A drug company wishes to compare the effects of three different drugs (X, Y, and Z) that are being developed to reduce cholesterol levels. Each drug is administered to six patients at the recommended dosage for six months. At the end of this period the reduction in cholesterol level is recorded for each patient. The results are given in Table 12.14. Using these data we obtain $SSTO = 2547.8$, $SSE = 395.7$, $\bar{x}_X = 23.67$, $\bar{x}_Y = 39.17$, and $\bar{x}_Z = 12.50$. Completely analyze these data using one-way ANOVA. DS CholRed

12.23 A small builder of speculative homes builds three basic house designs and employs two foremen. The builder has used each foreman to build two houses of each design and has obtained the profits given in Table 12.15 (the profits are given in thousands of dollars, and the sample means are enclosed in blue rectangles). If we use two-way ANOVA, we find that the p-value related to $F(\text{int})$ is .001. Is this consistent with what you see in Figure 12.15? Explain your answer. Using the fact that $MSE = .39$, find an individual 95 percent confidence interval for the true mean profit when foreman 1 builds house design 3. DS HouseProf

Appendix 12.1 ■ Experimental Design and Analysis of Variance Using Excel

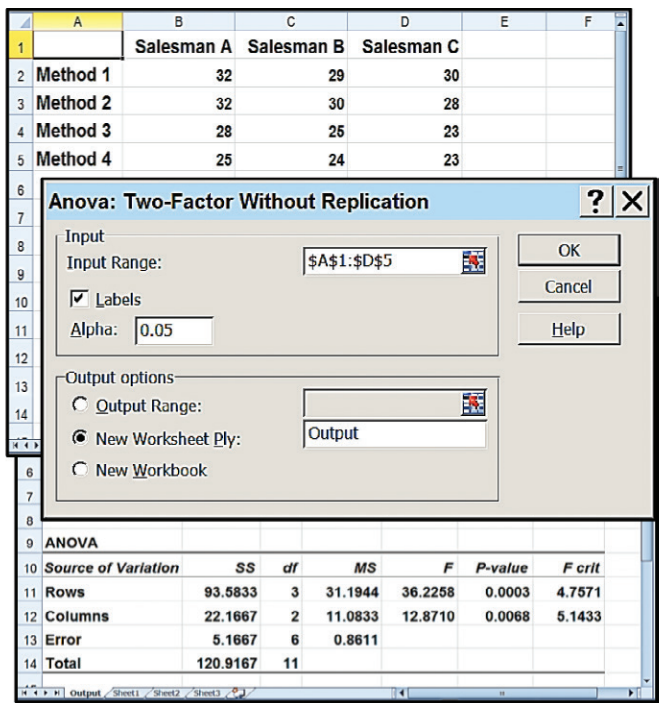
One-way ANOVA in Figure 12.2(b) on page 434 (data file: GasMile2.xlsx):

- Enter the gasoline mileage data from Table 12.1 (page 428) as follows: type the label "Type A" in cell A1 with its five mileage values in cells A2 to A6; type the label "Type B" in cell B1 with its five mileage values in cells B2 to B6; type the label "Type C" in cell C1 with its five mileage values in cells C2 to C6.
- Select **Data : Data Analysis : Anova : Single Factor** and click OK in the Data Analysis dialog box.
- In the "Anova: Single Factor" dialog box, enter A1:C6 into the "Input Range" window.
- Select the "Grouped by: Columns" option.
- Place a checkmark in the "Labels in first row" checkbox.
- Enter 0.05 into the Alpha box.
- Under output options, select "New Worksheet Ply" to have the output placed in a new worksheet and enter the name "Output" for the new worksheet.
- Click OK in the "Anova: Single Factor" dialog box.



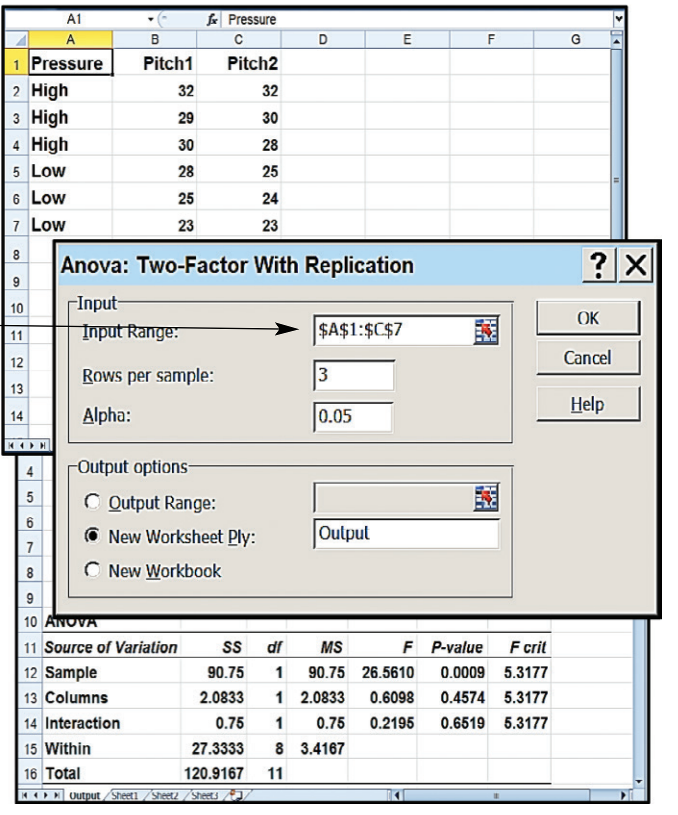
Randomized block ANOVA in Figure 12.8 on page 443 (data file: SaleMeth.xlsx):

- Enter the sales methods data from Figure 12.8 (page 443) as shown in the screen.
- Select **Data : Data Analysis : Anova: Two-Factor Without Replication** and click OK in the Data Analysis dialog box.
- In the "Anova: Two Factor Without Replication" dialog box, enter A1:D5 into the "Input Range" window.
- Place a checkmark in the "Labels" checkbox.
- Enter 0.05 in the Alpha box.
- Under output options, select "New Worksheet Ply" to have the output placed in a new worksheet and enter the name "Output" for the new worksheet.
- Click OK in the "Anova: Two-Factor Without Replication" dialog box.



Two-way ANOVA in Figure 12.14 on page 454 (data file: SaleMeth2.xlsx):

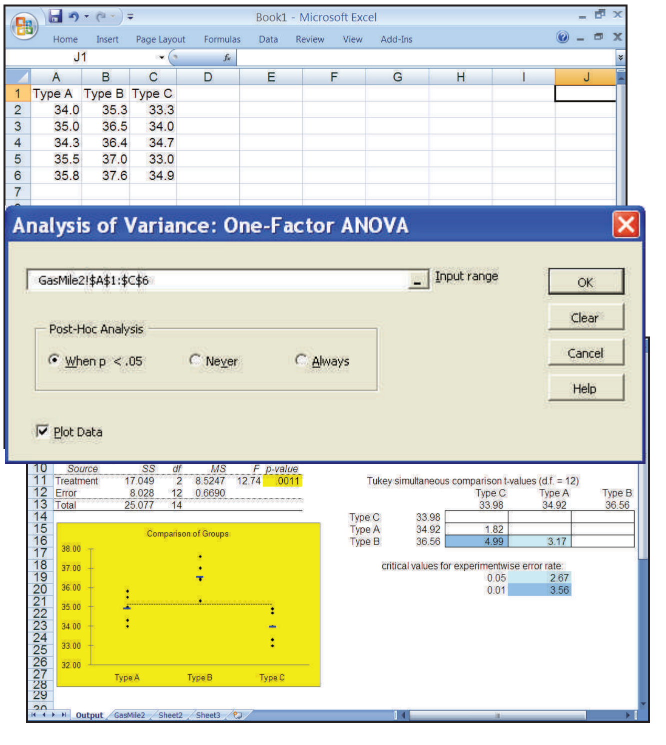
- Enter the sales approach experiment data from Table 12.13 (page 454) as shown in the screen.
- Select **Data : Data Analysis : Anova : Two-Factor With Replication** and click OK in the Data Analysis dialog box.
- In the "Anova: Two-Factor With Replication" dialog box, enter A1:C7 into the "Input Range" window.
- Enter the value 3 into the "Rows per Sample" box (this indicates the number of replications).
- Enter 0.05 in the Alpha box.
- Under output options, select "New Worksheet Ply" to have the output placed in a new worksheet and enter the name "Output" for the new worksheet.
- Click OK in the "Anova: Two-Factor With Replication" dialog box.



Appendix 12.2 ■ Experimental Design and Analysis of Variance Using MegaStat

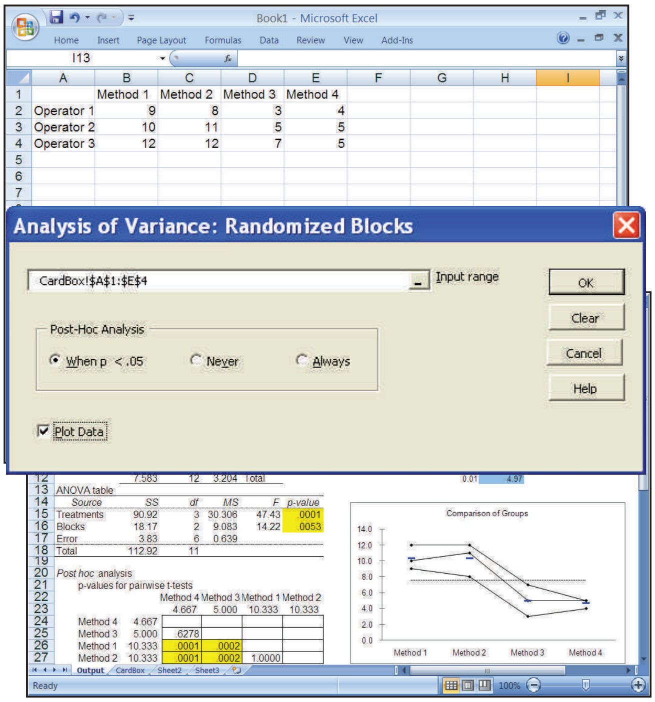
One-way ANOVA similar to Figure 12.2(b) on page 434 (data file: GasMile2.xlsx):

- Enter the gas mileage data in Table 12.1 (page 428) into columns A, B, and C—Type A mileages in column A (with label "Type A"), Type B mileages in column B (with label "Type B"), and Type C mileages in column C (with label "Type C"). Note that the input columns for the different groups must be side by side. However, the number of observations in each group can be different.
- Select **Add-Ins : MegaStat : Analysis of Variance : One-Factor ANOVA.**
- In the One-Factor ANOVA dialog box, use the AutoExpand feature to enter the range A1:C6 into the Input Range window.
- If desired, request "Post-hoc Analysis" to obtain Tukey simultaneous comparisons and pairwise *t* tests. Select from the options: "Never," "Always," or "When $p < .05$." The option "When $p < .05$" gives post-hoc analysis when the *p*-value for the *F* statistic is less than .05.
- Check the Plot Data checkbox to obtain a plot comparing the groups.
- Click OK in the One-Factor ANOVA dialog box.



Randomized block ANOVA similar to Figure 12.7(b) on page 442 (data file: CardBox.xlsx):

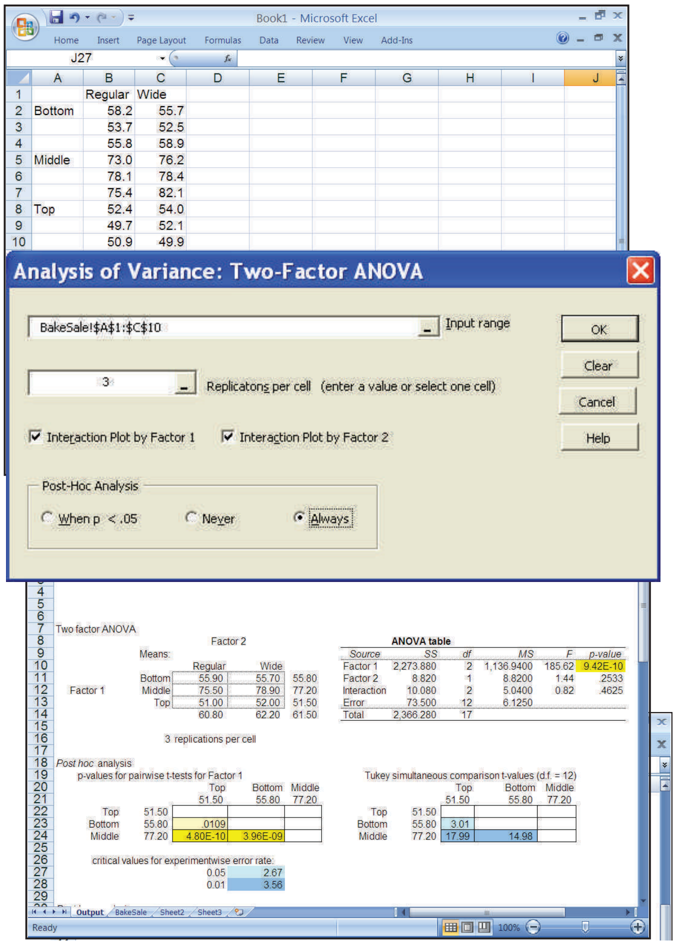
- Enter the cardboard box data in Table 12.7 (page 439) in the arrangement shown in the screen. Here each column corresponds to a **treatment** (in this case, a production method) and each row corresponds to a **block** (in this case, a machine operator). Identify the production methods using the labels Method 1, Method 2, Method 3, and Method 4 in cells B1, C1, D1, and E1. Identify the blocks using the labels Operator 1, Operator 2, and Operator 3 in cells A2, A3, and A4.
- Select **Add-Ins : MegaStat : Analysis of Variance : Randomized Blocks ANOVA.**
- In the Randomized Blocks ANOVA dialog box, click in the Input Range window and enter the range A1:E4.
- If desired, request "Post-hoc Analysis" to obtain Tukey simultaneous comparisons and pairwise *t*-tests. Select from the options: "Never," "Always," or "When $p < .05$." The option "When $p < .05$" gives post-hoc analysis when the *p*-value related to the *F* statistic for the treatments is less than .05.
- Check the Plot Data checkbox to obtain a plot comparing the treatments.



- Click OK in the Randomized Blocks ANOVA dialog box.

Two-way ANOVA similar to Figure 12.11(b) on page 449 (data file: BakeSale2.xlsx):

- Enter the bakery demand data in Table 12.11 (page 446) in the arrangement shown in the screen. Here the row labels Bottom, Middle, and Top are the levels of factor 1 (in this case, shelf display height) and the column labels Regular and Wide are the levels of factor 2 (in this case, shelf display width). The arrangement of the data is as laid out in Table 12.11.
- Select **Add-Ins : MegaStat : Analysis of Variance: Two-Factor ANOVA.**
- In the Two-Factor ANOVA dialog box, enter the range A1:C10 into the Input Range window.
- Type 3 into the "Replications per Cell" window.
- Check the "Interaction Plot by Factor 1" and "Interaction Plot by Factor 2" checkboxes to obtain interaction plots.
- If desired, request "Post-hoc Analysis" to obtain Tukey simultaneous comparisons and pairwise *t*-tests. Select from the options: "Never," "Always," and "When $p < .05$." The option "When $p < .05$" gives post-hoc analysis when the *p*-value related to the *F* statistic for a factor is less than .05. Here we have selected "Always."
- Click OK in the Two-Factor ANOVA dialog box.



Appendix 12.3 ■ Experimental Design and Analysis of Variance Using MINITAB

One-way ANOVA in Figure 12.2(a) on page 434 (data file: GasMile2.MTW):

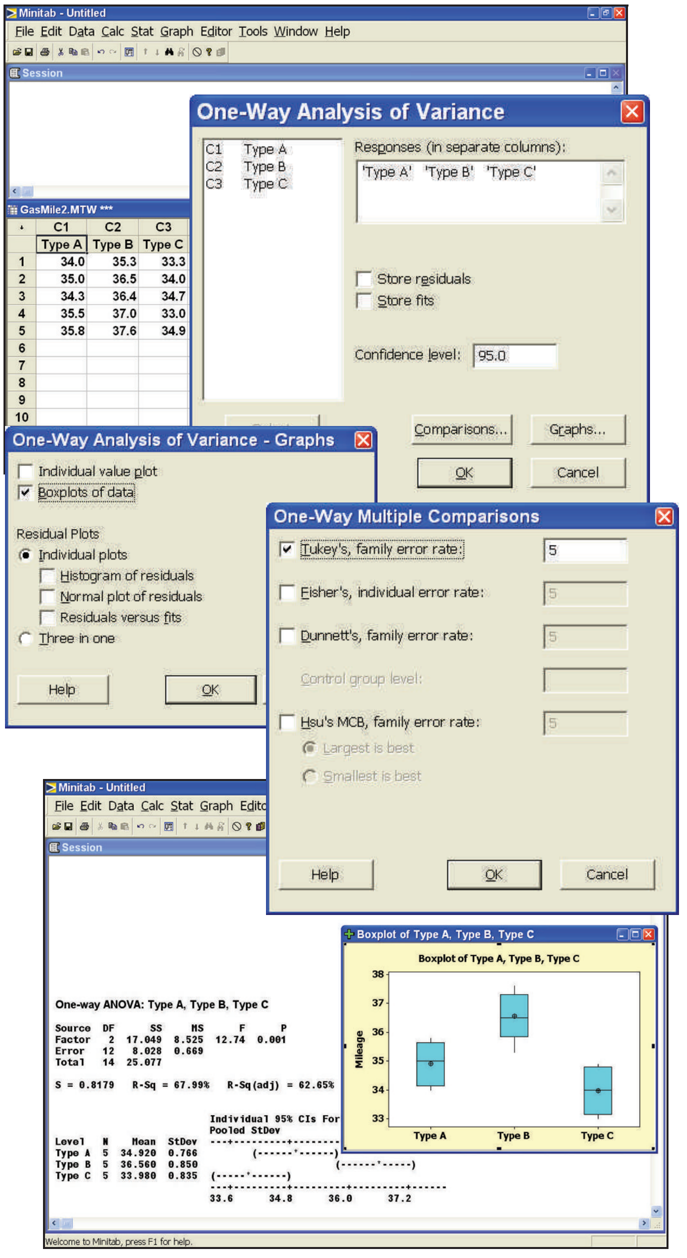
- In the Data window, enter the data from Table 12.1 (page 428) into three columns with variable names Type A, Type B, and Type C.
- Select **Stat : ANOVA : One-way (Unstacked).**
- In the "One-Way Analysis of Variance" dialog box, enter 'Type A' 'Type B' 'Type C' into the "Responses (in separate columns)" window. (The single quotes are necessary because of the blank spaces in the variable names. The quotes will be added automatically if the names are selected from the variable list or if they are selected by double clicking.)
- Click OK in the "One-Way Analysis of Variance" dialog box.

To produce mileage by gasoline type boxplots similar to those shown in Table 12.1 (page 428):

- Click the Graphs . . . button in the "One-Way Analysis of Variance" dialog box.
- Check the "Boxplots of data" checkbox and click OK in the "One-Way Analysis of Variance—Graphs" dialog box.
- Click OK in the "One-Way Analysis of Variance" dialog box.

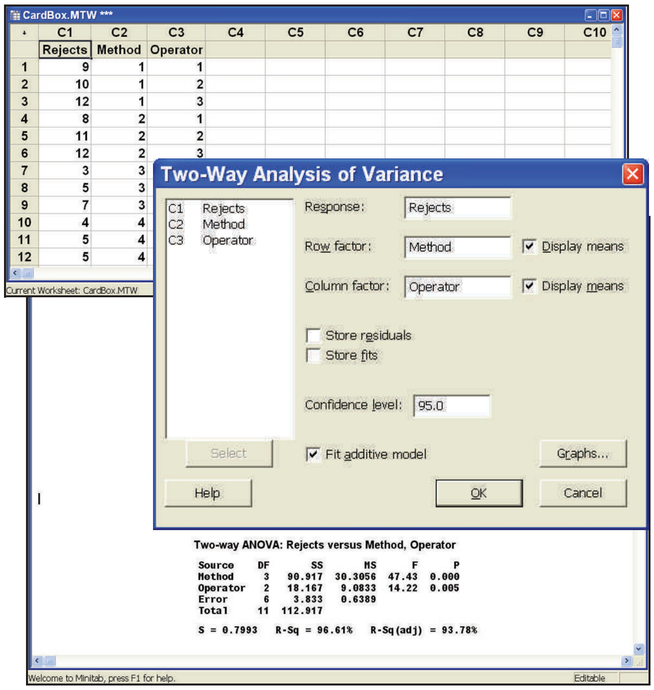
To produce Tukey pairwise comparisons:

- Click on the Comparisons . . . button in the "One-Way Analysis of Variance" dialog box.
- Check the "Tukey's family error rate" checkbox.
- In the "Tukey's family error rate" box, enter the desired experimentwise error rate (here we have entered 5, which denotes 5%—alternatively, we could enter the decimal fraction .05).
- Click OK in the "One-Way Multiple Comparisons" dialog box.
- Click OK in the "One-Way Analysis of Variance" dialog box.
- The one-way ANOVA output and the Tukey multiple comparisons will be given in the Session window, and the box plots will appear in a graphics window.



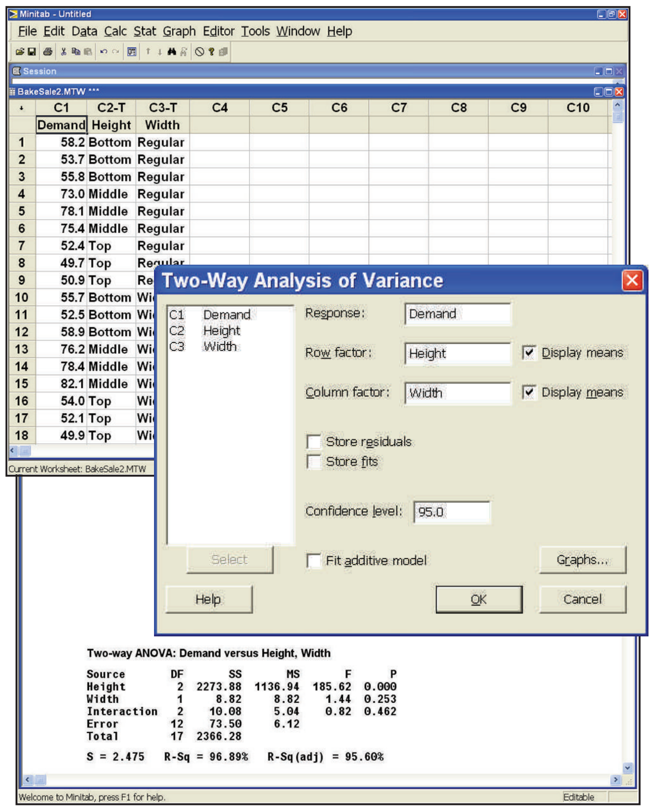
Randomized Block ANOVA in Figure 12.7(a) on page 442 (data File: CardBox.MTW):

- In the data window, enter the observed number of defective boxes from Table 12.7 (page 439) into column C1 with variable name "Rejects"; enter the corresponding production method (1,2,3,or 4) into column C2 with variable name "Method"; and enter the corresponding machine operator (1,2,or 3) into column C3 with variable name "Operator."
- Select **Stat : ANOVA : Two-way.**
- In the "Two-Way Analysis of Variance" dialog box, select Rejects into the Response window.
- Select Method into the Row Factor window and check the "Display Means" checkbox.
- Select Operator into the Column Factor window and check the "Display Means" checkbox.
- Check the "Fit additive model" checkbox.
- Click OK in the "Two-way Analysis of Variance" dialog box to display the randomized block ANOVA in the Session window.



Two-way ANOVA in Figure 12.11(a) on page 449 (data file: BakeSale2.MTW):

- In the data window, enter the observed demands from Table 12.11 (page 446) into column C1 with variable name "Demand"; enter the corresponding shelf display heights (Bottom, Middle, or Top) into column C2 with variable name "Height"; and enter the corresponding shelf display widths (Regular or Wide) into column C3 with variable name "Width."
- Select **Stat : ANOVA : Two-Way.**
- In the "Two-Way Analysis of Variance" dialog box, select Demand into the Response window.
- Select Height into the "Row Factor" window.
- Select Width into the "Column Factor" window.
- To produce tables of means by Height and Width, check the "Display means" checkboxes next to the "Row factor" and "Column factor" windows. This will also produce individual confidence intervals for each level of the row factor and each level of the column factor—these intervals are not shown in Figure 12.11.
- Enter the desired level of confidence for the individual confidence intervals in the "Confidence level" box.
- Click OK in the "Two-Way Analysis of Variance" dialog box.



To produce an interaction plot similar to the one in Figure 12.10 on page 446:

- Select **Stat : ANOVA : Interactions plot.**
- In the Interactions Plot dialog box, select Demand into the Responses window.
- Select Width and Height into the Factors window.
- Click OK in the Interactions Plot dialog box to obtain the plot in a graphics window.