12

Analysis of Variance

STATISTICS TODAY

Is Seeing Really Believing?

Many adults look on the eyewitness testimony of children with skepticism. They believe that young witnesses' testimony is less accurate than the testimony of adults in court cases. Several statistical studies have been done on this subject.

In a preliminary study, three researchers randomly selected fourteen 8-year-olds, fourteen 12-year-olds, and fourteen adults. The researchers showed each group the same video of a crime being committed. The next day, each witness responded to direct and cross-examination questioning. Then the researchers, using statistical methods explained in this chapter, were able to determine if there were differences in the accuracy of the testimony of the three groups on direct examination and on cross-examination. The statistical methods used here differ from the ones explained in Chapter 9 because there are three groups rather than two. See Statistics Today—Revisited at the end of this chapter.

Source: C. Luus, G. Wells, and J. Turtle, "Child Eyewitnesses: Seeing Is Believing," *Journal of Applied Psychology* 80, no. 2, pp. 317–26.

OUTLINE

OBJECTIVES

After completing this chapter, you should be able to

1. Use the one-way ANOVA technique to determine if there is a significant difference among three or more means.
2. Determine which means differ, using the Scheffé or Tukey test if the null hypothesis is rejected in the ANOVA.
3. Use the two-way ANOVA technique to determine if there is a significant difference in the main effects or interaction.

Historical Note

The methods of analysis of variance were developed by R. A. Fisher in the early 1920s.

Introduction

The F test, used to compare two variances as shown in Chapter 9, can also be used to compare three or more means. This technique is called *analysis of variance,* or *ANOVA*. It is used to test claims involving three or more means. (*Note:* The F test can also be used to test the equality of two means. But since it is equivalent to the t test in this case, the t test is usually used instead of the F test when there are only two means.) For example, suppose a researcher wishes to see whether the means of the time it takes three groups of students to solve a computer problem using HTML, Java, and PHP are different. The researcher will use the ANOVA technique for this test. The z and t tests should not be used when three or more means are compared, for reasons given later in this chapter.

For three groups, the F test can show only whether a difference exists among the three means. It cannot reveal where the difference lies—that is, between $\overline{X}_1$ and $\overline{X}_2$, or $\overline{X}_1$ and $\overline{X}_3$, or $\overline{X}_2$ and $\overline{X}_3$. If the F test indicates that there is a difference among the means, other statistical tests are used to find where the difference exists. The most commonly used tests are the Scheffé test and the Tukey test, which are also explained in this chapter.

The analysis of variance that is used to compare three or more means is called a *one-way analysis of variance* since it contains only one variable. In the previous example, the variable is the type of computer language used. The analysis of variance can be extended to studies involving two variables, such as type of computer language used and mathematical background of the students. These studies involve a *two-way analysis of variance.* Section 12–3 explains the two-way analysis of variance.

12–1 One-Way Analysis of Variance

OBJECTIVE 1

Use the one-way ANOVA technique to determine if there is a significant difference among three or more means.

When an F test is used to test a hypothesis concerning the means of three or more populations, the technique is called **analysis of variance** (commonly abbreviated as **ANOVA**).

The **one-way analysis of variance** test is used to test the equality of three or more means using sample variances.

The procedure used in this section is called the **one-way analysis of variance** because there is only one independent variable that distinguishes between the different populations in the study. The independent variable is also called a *factor*.

At first glance, you might think that to compare the means of three or more samples, you can use the t test, comparing two means at a time. But there are several reasons why the t test should not be done.

First, when you are comparing two means at a time, the rest of the means under study are ignored. With the F test, all the means are compared simultaneously. Second, when you are comparing two means at a time and making all pairwise comparisons, the probability of rejecting the null hypothesis when it is true is increased, since the more t tests that are conducted, the greater is the likelihood of getting significant differences by chance alone. Third, the more means there are to compare, the more t tests are needed. For example, for the comparison of 3 means two at a time, 3 t tests are required. For the comparison of 5 means two at a time, 10 tests are required. And for the comparison of 10 means two at a time, 45 tests are required.

As the number of populations to be compared increases, the probability of making a type I error using multiple t tests for a given level of significance α also increases. To address this problem, the technique of analysis of variance is used. This technique involves a comparison of two estimates of the same population variance.

Recall that the characteristics of the F distribution are as follows:

1. The values of F cannot be negative, because variances are always positive or zero.
2. The distribution is positively skewed.

3. The mean value of F is approximately equal to 1.

4. The F distribution is a family of curves based on the degrees of freedom of the variance of the numerator and the degrees of freedom of the variance of the denominator.

Even though you are comparing three or more means in this use of the F test, *variances* are used in the test instead of means.

With the F test, two different estimates of the population variance are made. The first estimate is called the **between-group variance,** and it involves finding the variance of the means. The second estimate, the **within-group variance,** is made by computing the variance using all the data and is not affected by differences in the means. If there is no difference in the means, the between-group variance estimate will be approximately equal to the within-group variance estimate, and the F test value will be approximately equal to one. The null hypothesis will not be rejected. However, when the means differ significantly, the between-group variance will be much larger than the within-group variance; the F test value will be significantly greater than one; and the null hypothesis will be rejected. Since variances are compared, this procedure is called *analysis of variance* (ANOVA).

The formula for the F test is

$$F = \frac{\text{variance between groups}}{\text{variance within groups}}$$

The variance between groups measures the differences in the means that result from the different treatments given to each group. To calculate this value, it is necessary to find the *grand mean* $\overline{X}_{\mathrm{GM}}$, which is the mean of all the values in all of the samples. The formula for the grand mean is

$$\overline{X}_{\mathrm{GM}} = \frac{\Sigma X}{N}$$

This value is used to find the between-group variance s_B^2. This is the variance among the means using the sample sizes as weights.

The formula for the between-group variance, denoted by s_B^2, is

$$s_B^2 = \frac{\Sigma n_i(\overline{X}_i - \overline{X}_{\mathrm{GM}})}{k - 1}$$

where k = number of groups
n_i = sample size
$\overline{X}_i$ = sample mean

This formula can be written out as

$$s_B^2 = \frac{n_1(\overline{X}_1 - \overline{X}_{\mathrm{GM}})^2 + n_2(\overline{X}_2 - \overline{X}_{\mathrm{GM}})^2 + \cdots + n_k(\overline{X}_k - \overline{X}_{\mathrm{GM}})^2}{k - 1}$$

Next find the within group variance, denoted by s_W^2. The formula finds the overall variance by calculating a weighted average of the individual variances. It does not involve using differences of means. The formula for the within-group variance is

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}$$

where n_i = sample size
s_i^2 = variance of sample

This formula can be written out as

$$s_W^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{(n_1 - 1) + (n_2 - 2) + \cdots + (n_k - 1)}$$

Finally, the F test value is computed. The formula can now be written using the symbols s_B^2 and s_W^2.

The formula for the F test for one-way analysis of variance is

$$F = \frac{s_B^2}{s_W^2}$$

where s_B^2 = between-group variance
s_W^2 = within-group variance

As stated previously, a significant test value means that there is a high probability that this difference in means is not due to chance, but it does not indicate where the difference lies.

The degrees of freedom for this F test are d.f.N. $= k - 1$, where k is the number of groups, and d.f.D. $= N - k$, where N is the sum of the sample sizes of the groups $N = n_1 + n_2 + \cdots + n_k$. The sample sizes need not be equal. The F test to compare means is always right-tailed.

The results of the one-way analysis of variance can be summarized by placing them in an **ANOVA summary table.** The numerator of the fraction of the s_B^2 term is called the **sum of squares between groups,** denoted by SS_B. The numerator of the s_W^2 term is called the **sum of squares within groups,** denoted by SS_W. This statistic is also called the *sum of squares for the error.* SS_B is divided by d.f.N. to obtain the between-group variance. SS_W is divided by $N - k$ to obtain the within-group or error variance. These two variances are sometimes called **mean squares,** denoted by MS_B and MS_W. These terms are used to summarize the analysis of variance and are placed in a summary table, as shown in Table 12–1.

TABLE 12–1 Analysis of Variance Summary Table

Source	Sum of squares	d.f.	Mean square	F
Between	SS_B	$k - 1$	MS_B	
Within (error)	SS_W	$N - k$	MS_W	
Total				

Unusual Stat

The *Journal of the American College of Nutrition* reports that a study found no correlation between body weight and the percentage of calories eaten after 5:00 P.M.

In the table,

$$SS_B = \text{sum of squares between groups}$$
$$SS_W = \text{sum of squares within groups}$$
$$k = \text{number of groups}$$
$$N = n_1 + n_2 + \cdots + n_k = \text{sum of sample sizes for groups}$$
$$MS_B = \frac{SS_B}{k - 1}$$
$$MS_W = \frac{SS_W}{N - k}$$
$$F = \frac{MS_B}{MS_W}$$

To use the F test to compare two or more means, the following assumptions must be met.

Assumptions for the *F* Test for Comparing Three or More Means

1. The populations from which the samples were obtained must be normally or approximately normally distributed.
2. The samples must be independent of one another.
3. The variances of the populations must be equal.
4. The samples must be simple random samples, one from each of the populations.

In this book, the assumptions will be stated in the exercises; however, when encountering statistics in other situations, you must check to see that these assumptions have been met before proceeding.

The steps for computing the F test value for the ANOVA are summarized in this Procedure Table.

Procedure Table

Finding the *F* Test Value for the Analysis of Variance

Step 1 Find the mean and variance of each sample.

$$(\bar{X}_1, s_1^2), (\bar{X}_2, s_2^2), \ldots, (\bar{X}_k, s_k^2)$$

Step 2 Find the grand mean.

$$\bar{X}_{\text{GM}} = \frac{\Sigma X}{N}$$

Step 3 Find the between-group variance.

$$s_B^2 = \frac{\Sigma n_i(\bar{X}_i - \bar{X}_{\text{GM}})^2}{k - 1}$$

Step 4 Find the within-group variance.

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}$$

Step 5 Find the F test value.

$$F = \frac{s_B^2}{s_W^2}$$

The degrees of freedom are

$$\text{d.f.N.} = k - 1$$

where k is the number of groups, and

$$\text{d.f.D.} = N - k$$

where N is the sum of the sample sizes of the groups

$$N = n_1 + n_2 + \cdots + n_k$$

The one-way analysis of variance follows the regular five-step hypothesis-testing procedure.

Step 1 State the hypotheses.
Step 2 Find the critical values.
Step 3 Compute the test value.
Step 4 Make the decision.
Step 5 Summarize the results.

Examples 12–1 and 12–2 illustrate the computational procedure for the ANOVA technique for comparing three or more means, and the steps are summarized in the Procedure Table.

EXAMPLE 12–1 Miles per Gallon

A researcher wishes to see if there is a difference in the fuel economy for city driving for three different types of automobiles: small automobiles, sedans, and luxury automobiles. He randomly samples four small automobiles, five sedans, and three luxury automobiles. The miles per gallon for each is shown. At $\alpha = 0.05$, test the claim that there is no difference among the means. The data are shown.

Small	Sedans	Luxury
36	43	29
44	35	25
34	30	24
35	29	
	40	

Source: U.S. Environmental Protection Agency.

Step 1 State the hypotheses and identify the claim.

H_0: $\mu_1 = \mu_2 = \mu_3$ (claim)
H_1: At least one mean is different from the others

Step 2 Find the critical value.

$$N = 12 \qquad k = 3$$
$$\text{d.f.N.} = k - 1 = 3 - 1 = 2$$
$$\text{d.f.D.} = N - k = 12 - 3 = 9$$

The critical value from Table H in Appendix A with $\alpha = 0.05$ is 4.26.

Step 3 Compute the test value.

a. Find the mean and variance for each sample. (Use the formulas in Chapter 3.)

For the small cars: $\bar{X} = 37.25$ $\quad s^2 = 20.917$
For the sedans: $\bar{X} = 35.4$ $\quad s^2 = 37.3$
For the luxury cars: $\bar{X} = 26$ $\quad s^2 = 7$

b. Find the grand mean.

$$\bar{X}_{\text{GM}} = \frac{\Sigma X}{N} = \frac{36 + 44 + 34 + \cdots + 24}{12} = \frac{404}{12} = 33.667$$

c. Find the between-group variance.

$$s_B^2 = \frac{\Sigma n(\bar{X}_i - \bar{X}_{\text{GM}})^2}{k - 1}$$
$$= \frac{4(37.25 - 33.667)^2 + 5(35.4 - 33.667)^2 + 3(26 - 33.667)^2}{3 - 1}$$
$$= \frac{242.717}{2} = 121.359$$

d. Find the within-group variance.

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)} = \frac{(4 - 1)(20.917) + (5 - 1)(37.3) + (3 - 1)7}{(4 - 1) + (5 - 1) + (3 - 1)}$$
$$= \frac{225.951}{9} = 25.106$$

e. Find the F test value.

$$F = \frac{s_B^2}{s_W^2} = \frac{121.359}{25.106} = 4.83$$

Step 4 Make the decision. The test value $4.83 > 4.26$, so the decision is to reject the null hypothesis. See Figure 12–1.

FIGURE 12–1 Critical Value and Test Value for Example 12–1

Step 5 Summarize the results. There is enough evidence to conclude that at least one mean is different from the others.

The **ANOVA summary table** is shown in Table 12–2.

TABLE 12–2 Analysis of Variance Summary Table for Example 12–1

Source	Sum of squares	d.f.	Mean square	*F*
Between	242.717	2	121.359	4.83
Within (error)	225.954	9	25.106	
Total	468.671	11		

The P-values for ANOVA are found by using the same procedure shown in Section 9–5. For Example 12–1, the F test value is 4.83. In Table H with d.f.N. = 2 and d.f.D. = 9, the F test value falls between $\alpha = 0.025$ with an F value of 5.71 and $\alpha = 0.05$ with an F value of 4.26. Hence, $0.025 < P\text{-value} < 0.05$. In this case, the null hypothesis is rejected at $\alpha = 0.05$ since the $P\text{-value} < 0.05$. The TI-84 P-value is 0.0375.

EXAMPLE 12–2 Employees at Toll Road Interchanges

A state employee wishes to see if there is a significant difference in the number of employees at the interchanges of three state toll roads. The data are shown. At $\alpha = 0.05$, can it be concluded that there is a significant difference in the average number of employees at each interchange?

Pennsylvania Turnpike	Greensburg Bypass/ Mon-Fayette Expressway	Beaver Valley Expressway
7	10	1
14	1	12
32	1	1
19	0	9
10	11	1
11	1	11

Source: Pennsylvania Turnpike Commission.

SOLUTION

Step 1 State the hypotheses and identify the claim.

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

H_1: At least one mean is different from the others (claim).

Step 2 Find the critical value. Since $k = 3$, $N = 18$, and $\alpha = 0.05$,

d.f.N. $= k - 1 = 3 - 1 = 2$

d.f.D. $= N - k = 18 - 3 = 15$

The critical value is 3.68.

Step 3 Compute the test value.

a. Find the mean and variance of each sample. The mean and variance for each sample are

Turnpike	$\bar{X}_1 = 15.5$	$s_1^2 = 81.9$
Mon-Fayette	$\bar{X}_2 = 4.0$	$s_2^2 = 25.6$
Beaver Valley	$\bar{X}_3 = 5.8$	$s_3^2 = 29.0$

b. Find the grand mean.

$$\bar{X}_{\text{GM}} = \frac{\Sigma X}{N} = \frac{7 + 14 + 32 + \cdots + 11}{18} = \frac{152}{18} = 8.44$$

c. Find the between-group variance.

$$s_B^2 = \frac{\Sigma n_i(\bar{X}_i - \bar{X}_{\text{GM}})^2}{k - 1}$$

$$= \frac{6(15.5 - 8.44)^2 + 6(4 - 8.44)^2 + 6(5.8 - 8.44)^2}{3 - 1}$$

$$= \frac{459.16}{2} = 229.58$$

d. Find the within-group variance.

$$s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}$$

$$= \frac{(6 - 1)(81.9) + (6 - 1)(25.6) + (6 - 1)(29.0)}{(6 - 1) + (6 - 1) + (6 - 1)}$$

$$= \frac{682.50}{15} = 45.50$$

e. Find the F test value.

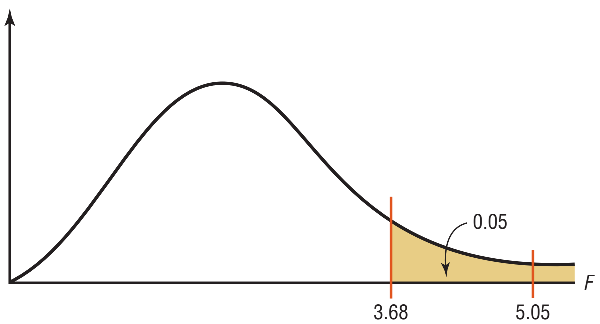
$$F = \frac{s_B^2}{s_W^2} = \frac{229.58}{45.50} = 5.05$$

Step 4 Make the decision. Since $5.05 > 3.68$, the decision is to reject the null hypothesis. See Figure 12–2.

Interesting Facts

The weight of 1 cubic foot of wet snow is about 10 pounds while the weight of 1 cubic foot of dry snow is about 3 pounds.

FIGURE 12–2 Critical Value and Test Value for Example 12–2



Step 5 Summarize the results. There is enough evidence to support the claim that there is a difference among the means. The ANOVA summary table for this example is shown in Table 12–3.

TABLE 12–3 Analysis of Variance Summary Table for Example 12–2

Source	Sum of squares	d.f.	Mean square	*F*
Between	459.16	2	229.58	5.05
Within	682.50	15	45.50	
Total	1141.66	17		

The *P*-values for ANOVA are found by using the procedure shown in Section 9–2. For Example 12–2, find the two α values in the tables for the F distribution (Table H), using d.f.N. = 2 and d.f.D. = 15, where $F = 5.05$ falls between. In this case, 5.05 falls between 4.77 and 6.36, corresponding, respectively, to $\alpha = 0.025$ and $\alpha = 0.01$; hence, $0.01 < P\text{-value} < 0.025$. Since the *P*-value is between 0.01 and 0.025 and since *P*-value < 0.05 (the originally chosen value for α), the decision is to reject the null hypothesis. (The *P*-value obtained from a calculator is 0.021.)

When the null hypothesis is rejected in ANOVA, it only means that at least one mean is different from the others. To locate the difference or differences among the means, it is necessary to use other tests such as the Tukey or the Scheffé test.

Applying the Concepts 12–1

Colors That Make You Smarter

The following set of data values was obtained from a study of people's perceptions on whether the color of a person's clothing is related to how intelligent the person looks. The subjects rated the person's intelligence on a scale of 1 to 10. Randomly selected group 1 subjects were shown people with clothing in shades of blue and gray. Randomly selected group 2 subjects were shown people with clothing in shades of brown and yellow. Randomly selected group 3 subjects were shown people with clothing in shades of pink and orange. The results follow.

Group 1	Group 2	Group 3
8	7	4
7	8	9
7	7	6
7	7	7
8	5	9
8	8	8
6	5	5
8	8	8
8	7	7
7	6	5
7	6	4
8	6	5
8	6	4

1. Use ANOVA to test for any significant differences between the means.
2. What is the purpose of this study?
3. Explain why separate *t* tests are not accepted in this situation.

See page 686 for the answers.

Exercises 12–1

1. What test is used to compare three or more means?
2. State three reasons why multiple t tests cannot be used to compare three or more means.
3. What are the assumptions for ANOVA?
4. Define between-group variance and within-group variance.
5. State the hypotheses used in the ANOVA test.
6. When there is no significant difference among three or more means, the value of F will be close to what number?

For Exercises 7 through 20, assume that all variables are normally distributed, that the samples are independent, that the population variances are equal, and that the samples are simple random samples, one from each of the populations. Also, for each exercise, perform the following steps.

a. State the hypotheses and identify the claim.
b. Find the critical value.
c. Compute the test value.
d. Make the decision.
e. Summarize the results, and explain where the differences in the means are.

Use the traditional method of hypothesis testing unless otherwise specified.

7. **Tire Prices** A large tire company held an end-of-season clearance sale. Listed are sale prices for random samples of different models for three different brands. Is there sufficient evidence at $\alpha = 0.05$ to conclude a difference in mean prices for the three brands?

Brand A	Brand B	Brand C
112	125	113
100	150	119
120	103	136
93	120	151
119	131	162
108	166	141
103	158	150

8. **Sodium Contents of Foods** The amount of sodium (in milligrams) in one serving for a random sample of three different kinds of foods is listed. At the 0.05 level of significance, is there sufficient evidence to conclude that a difference in mean sodium amounts exists among condiments, cereals, and desserts?

Condiments	Cereals	Desserts
270	260	100
130	220	180
230	290	250
180	290	250
80	200	300
70	320	360
200	140	300
		160

Source: *The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.*

9. **Hybrid Vehicles** A study was done before the recent surge in gasoline prices to compare the cost to drive 25 miles for different types of hybrid vehicles. The cost of a gallon of gas at the time of the study was approximately \$2.50. Based on the information given for different models of hybrid cars, trucks, and SUVs, is there sufficient evidence to conclude a difference in the mean cost to drive 25 miles? Use $\alpha = 0.05$. (The information in this exercise will be used in Exercise 3 in Section 12–2.)

Hybrid cars	Hybrid SUVs	Hybrid trucks
2.10	2.10	3.62
2.70	2.42	3.43
1.67	2.25	
1.67	2.10	
1.30	2.25	

Source: www.fueleconomy.com

10. **Healthy Eating** Americans appear to be eating healthier. Between 1970 and 2007 the per capita consumption of broccoli increased 1000% from 0.5 to 5.5 pounds. A nutritionist followed a group of people randomly assigned to one of three groups and noted their monthly broccoli intake (in pounds). At $\alpha = 0.05$, is there a difference in means?

Group A	Group B	Group C
2.0	2.0	3.7
1.5	1.5	2.5
0.75	4.0	4.0
1.0	3.0	5.1
1.3	2.5	3.8
3.0	2.0	2.9

Source: *World Almanac.*

11. **Student Loans** The average undergraduate student loan for a recent year was \$8500. A random sample of students from three different schools revealed the following loan amounts for the last school year. Based on the $\alpha = 0.05$ level of significance, is there a difference in means?

College A	College B	College C
9,000	10,000	12,000
10,500	15,000	15,000
12,600	16,000	16,500
10,900	14,500	15,500
15,000	12,000	14,000
11,000		12,800

Source: *World Almanac.*

12. **Weight Gain of Athletes** A researcher wishes to see whether there is any difference in the weight gains of athletes following one of three special diets. Athletes are randomly assigned to three groups and placed on the diet

for 6 weeks. The weight gains (in pounds) are shown here. At $\alpha = 0.05$, can the researcher conclude that there is a difference in the diets?

Diet A	Diet B	Diet C
3	10	8
6	12	3
7	11	2
4	14	5
	8	
	6	

A computer printout for this problem is shown. Use the P-value method and the information in this printout to test the claim. (The information in this exercise will be used in Exercise 4 of Section 12–2.)

Computer Printout for Exercise 12

ANALYSIS OF VARIANCE SOURCE TABLE

Source	df	Sum of Squares	Mean Square	F	P-value
Bet Groups	2	101.095	50.548	7.740	0.00797
W/I Groups	11	71.833	6.530		
Total	13	172.929			

DESCRIPTIVE STATISTICS

Condit	N	Means	St Dev
diet A	4	5.000	1.826
diet B	6	10.167	2.858
diet C	4	4.500	2.646

13. Expenditures per Pupil The per-pupil costs (in thousands of dollars) for cyber charter school tuition for school districts in three areas of southwestern Pennsylvania are shown. At $\alpha = 0.05$, is there a difference in the means? If so, give a possible reason for the difference. (The information in this exercise will be used in Exercise 5 of Section 12–2.)

Area I	Area II	Area III
6.2	7.5	5.8
9.3	8.2	6.4
6.8	8.5	5.6
6.1	8.2	7.1
6.7	7.0	3.0
6.9	9.3	3.5

Source: *Tribune-Review.*

14. Cell Phone Bills The average local cell phone monthly bill is \$50.07. A random sample of monthly bills from three different providers is listed below. At $\alpha = 0.05$, is there a difference in mean bill amounts among providers?

Provider X	Provider Y	Provider Z
48.20	105.02	59.27
60.59	85.73	65.25
72.50	61.95	70.27
55.62	75.69	42.19
89.47	82.11	52.34

Source: *World Almanac.*

15. Television Viewing Time The average U.S. television viewing time (2010–2011) for all viewers is 34 hours and 16 minutes per week. Random samples of three different groups indicated their weekly viewing habits (in hours) as listed below. At the 0.05 level of significance, is there evidence of a difference in means between the groups?

Men 21+ years	Women 21+ years	"Teens" 12–20 years
28	32	44
26	31	37
20	47	40
25	40	31
31	34	28
		34

Source: *World Almanac.*

16. Annual Child Care Costs Annual child care costs for infants are considerably higher than for older children. At $\alpha = 0.05$, can you conclude a difference in mean infant day care costs for different regions of the United States? (Annual costs per infant are given in dollars.)

New England	Midwest	Southwest
10,390	9,449	7,644
7,592	6,985	9,691
8,755	6,677	5,996
9,464	5,400	5,386
7,328	8,372	

Source: www.naccrra.org (National Association of Child Care Resources and Referral Agencies: "Breaking the Piggy Bank").

17. Microwave Oven Prices A research organization tested microwave ovens. At $\alpha = 0.10$, is there a significant difference in the average prices of the three types of oven?

Watts		
1000	900	800
270	240	180
245	135	155
190	160	200
215	230	120
250	250	140
230	200	180
	200	140
	210	130

A computer printout for this exercise is shown. Use the P-value method and the information in this printout to test the claim. (The information in this exercise will be used in Exercise 6 of Section 12–2.)

Computer Printout for Exercise 17

ANALYSIS OF VARIANCE SOURCE TABLE

Source	df	Sum of Squares	Mean Square	F	P-value
Bet Groups	2	21729.735	10864.867	10.118	0.00102
W/I Groups	19	20402.083	1073.794		
Total	21	42131.818			

DESCRIPTIVE STATISTICS

Condit	N	Means	St Dev
1000	6	233.333	28.23
900	8	203.125	39.36
800	8	155.625	28.21

18. Calories in Fast-Food Sandwiches Three popular fast-food restaurant franchises specializing in burgers were surveyed to find out the number of calories in their frequently ordered sandwiches. At the 0.05 level of significance, can it be concluded that a difference in mean number of calories per burger exists? The information in this exercise will be used for Exercise 7 in Section 12–2.

FF#1	FF#2	FF#3
970	1010	740
880	970	540
840	920	510
710	850	510
	820	

Source: www.fatcalories.com

19. Number of Pupils in a Class A large school district has several middle schools. Three schools were randomly chosen, and four classes were selected from each. The numbers of pupils in each class are shown here. At $\alpha = 0.10$, is there sufficient evidence that the mean number of students per class differs among schools?

MS 1	MS 2	MS 3
21	28	25
25	22	20
19	25	23
17	30	22

20. Average Debt of College Graduates Kiplinger's listed the top 100 public colleges based on many factors. From that list, here is the average debt at graduation for various schools in four selected states. At $\alpha = 0.05$, can it be concluded that the average debt at graduation differs for these four states?

New York	Virginia	California	Pennsylvania
14,734	14,524	13,171	18,105
16,000	15,176	14,431	17,051
14,347	12,665	14,689	16,103
14,392	12,591	13,788	22,400
12,500	18,385	15,297	17,976

Source: www.Kiplinger.com

Technology Step by Step

TI-84 Plus Step by Step

One-Way Analysis of Variance (ANOVA)

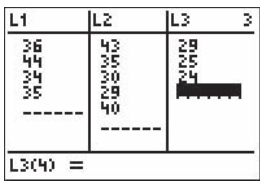
1. Enter the data into L_1, L_2, L_3, etc.
2. Press **STAT** and move the cursor to TESTS.
3. Press **H (ALPHA^)** for ANOVA(.
4. Type each list followed by a comma. End with) and press **ENTER.**

Example TI12–1

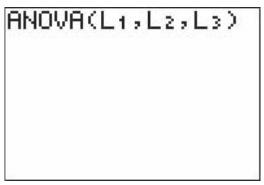
Test the claim H_0: $\mu_1 = \mu_2 = \mu_3 =$ at $\alpha = 0.05$ for these data from Example 12–1.

Small	Sedans	Luxury
36	43	29
44	35	25
34	30	24
35	29	
	40	

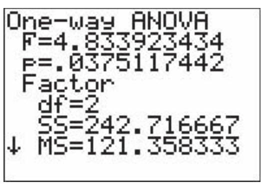
Input



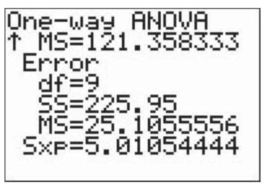
Input



Output



Output



The F test value is 4.833923434. The P-value is 0.0375117442, which is significant at $\alpha = 0.05$. The factor variable has

$$\text{d.f.} = 2$$
$$\text{SS} = 242.716667$$
$$\text{MS} = 121.358333$$

The error has

$$\text{d.f.} = 9$$
$$\text{SS} = 225.95$$
$$\text{MS} = 25.1055556$$

EXCEL
Step by Step

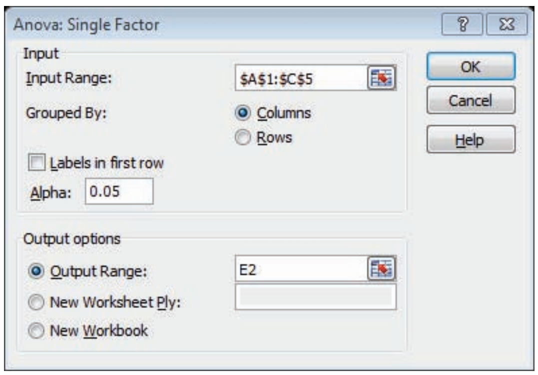
One-Way Analysis of Variance (ANOVA)

Example XL12–1

1. Enter the data below in columns A, B, and C of an Excel worksheet.

10	6	5
12	8	9
9	3	12
15	0	8
13	2	4

2. From the toolbar, select **Data,** then **Data Analysis.**
3. Select **Anova: Single Factor.**
4. Type in **A1:C5** in the **Input Range** box.
5. Check **Grouped By: Columns.**
6. Type **0.05** for the **Alpha level.**
7. Under the **Output options,** check the **Output Range** and type **E2**.
8. Click [OK].



The results of the ANOVA are shown below.

Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	5	59	11.8	5.7		
Column 2	5	19	3.8	10.2		
Column 3	5	38	7.6	10.3		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	160.1333333	2	80.06666667	9.167938931	0.003831317	3.885293835
Within Groups	104.8	12	8.733333333			
Total	264.9333333	14				

MINITAB
Step by Step

One-Way Analysis of Variance (ANOVA)

Example 12–1

Is there a difference in the average city MPG rating by type of vehicle?

1. Enter the MPG ratings in C1 Small, C2 Sedan, and C3 Luxury.

↓	C1	C2	C3
	Small	Sedan	Luxury
1	36	43	29
2	44	35	25
3	34	30	24
4	35	29	
5		40	

2. Select **Stat>ANOVA>One-way (unstacked).**
   - **a.** Drag the mouse over the three columns of Observed counts.
   - **b.** Click on [Select].
   - **c.** Click [OK].

The results are displayed in the session window.

One-way ANOVA: Small, Sedan, Luxury

```
Source   DF      SS      MS      F      P
Factor    2   242.7   121.4   4.83  0.038
Error     9   226.0    25.1
Total    11   468.7

S = 5.011    R-Sq = 51.79%    R-Sq(adj) = 41.08%

                                 Individual 95% CIs For Mean Based on Pooled StDev
Level    N    Mean    StDev   ----------+------------+------------+------------+-
Small    4  37.250    4.573                           (------------*------------)
Sedan    5  35.400    6.107                        (------------*------------)
Luxury   3  26.000    2.646   (--------------*--------------)
                              ------------+------------+------------+------------+-
                                        24.0         30.0         36.0         42.0

Pooled StDev = 5.011
```

12–2 The Scheffé Test and the Tukey Test

When the null hypothesis is rejected using the *F* test, the researcher may want to know where the difference among the means is. Several procedures have been developed to determine where the significant differences in the means lie after the ANOVA procedure has been performed. Among the most commonly used tests are the *Scheffé test* and the *Tukey test.*

OBJECTIVE 2

Determine which means differ, using the Scheffé or Tukey test if the null hypothesis is rejected in the ANOVA.

Scheffé Test

To conduct the **Scheffé test,** you must compare the means two at a time, using all possible combinations of means. For example, if there are three means, the following comparisons must be done:

$$\overline{X}_1 \text{ versus } \overline{X}_2 \qquad \overline{X}_1 \text{ versus } \overline{X}_3 \qquad \overline{X}_2 \text{ versus } \overline{X}_3$$

Unusual Stat

According to the *British Medical Journal,* the body's circadian rhythms produce drowsiness during the midafternoon, matched only by the 2:00 A.M. to 7:00 A.M. period for sleep-related traffic accidents.

Formula for the Scheffé Test

$$F_S = \frac{(\bar{X}_i - \bar{X}_j)^2}{s_W^2[(1/n_i) + (1/n_j)]}$$

where $\bar{X}_i$ and $\bar{X}_j$ are the means of the samples being compared, n_i and n_j are the respective sample sizes, and s_W^2 is the within-group variance.

To find the critical value F' for the Scheffé test, multiply the critical value for the F test by $k - 1$:

$$F' = (k - 1)(\text{C.V.})$$

There is a significant difference between the two means being compared when the F test value, F_S, is greater than the critical value, F'. Example 12–3 illustrates the use of the Scheffé test.

EXAMPLE 12–3

Use the Scheffé test to test each pair of means in Example 12–1 to see if a significant difference exists between each pair of means. Use $\alpha = 0.05$.

SOLUTION

The F critical value for Example 12–1 is 4.26. Then the critical value for the individual tests with d.f.N. = 2 and d.f.D. = 9 is

$$F' = (k - 1)(\text{C.V.}) = (3 - 1)(4.26) = 8.52$$

a. For $\bar{X}_1$ versus $\bar{X}_2$,

$$F_S = \frac{(\bar{X}_1 - \bar{X}_2)^2}{s_W^2[(1/n_1) + (1/n_2)]} = \frac{(37.25 - 35.4)^2}{25.106(\frac{1}{4} + \frac{1}{5})} = 0.30$$

Since $0.30 < 8.52$, the decision is that μ_1 is not significantly different from μ_2.

b. For $\bar{X}_1$ versus $\bar{X}_3$,

$$F_S = \frac{(\bar{X}_1 - \bar{X}_3)^2}{s_W^2[(1/n_1) + (1/n_3)]} = \frac{(37.25 - 26)^2}{25.106(\frac{1}{4} + \frac{1}{3})} = 8.64$$

Since $8.64 > 8.52$, the decision is that μ_1 is significantly different from μ_3.

c. For $\bar{X}_2$ versus $\bar{X}_3$,

$$F_S = \frac{(\bar{X}_2 - \bar{X}_3)^2}{s_W^2[(1/n_2) + (1/n_3)]} = \frac{(35.4 - 26)^2}{25.106(\frac{1}{5} + \frac{1}{3})} = 6.60$$

Since $6.60 < 8.64$, the decision is that μ_2 is not significantly different from μ_3. Hence, only the mean of the small cars is not equal to the mean of luxury cars.

On occasion, when the F test value is greater than the critical value, the Scheffé test may not show any significant differences in the pairs of means. This result occurs because the difference may actually lie in the average of two or more means when compared with the other mean. The Scheffé test can be used to make these types of comparisons, but the technique is beyond the scope of this book.

SPEAKING OF STATISTICS Tricking Knee Pain

This study involved three groups. The results showed that patients in all three groups felt better after 2 years. State possible null and alternative hypotheses for this study. Was the null hypothesis rejected? Explain how the statistics could have been used to arrive at the conclusion.

HEALTH

TRICKING KNEE PAIN

You sign up for a clinical trial of arthroscopic surgery used to relieve knee pain caused by arthritis. You're sedated and wake up with tiny incisions. Soon your bum knee feels better. Two years later you find out you had "placebo" surgery. In a study at the Houston VA Medical Center, researchers divided 180 patients into three groups: two groups had damaged cartilage removed, while the third got simulated surgery. Yet an equal number of patients in all groups felt better after two years. Some 650,000 people have the surgery annually, but they're wasting their money, says Dr. Nelda P. Wray, who led the study. And the patients who got fake surgery? "They aren't angry at us," she says. "They still report feeling better."

— STEPHEN P. WILLIAMS

Source: From *Newsweek* July 22, 2002 © Newsweek, Inc.

Tukey Test

The **Tukey test** can also be used after the analysis of variance has been completed to make pairwise comparisons between means when the groups have the same sample size. The symbol for the test value in the Tukey test is q.

Formula for the Tukey Test

$$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{s_W^2/n}}$$

where $\bar{X}_i$ and $\bar{X}_j$ are the means of the samples being compared, n is the size of the samples, and s_W^2 is the within-group variance.

When the absolute value of q is greater than the critical value for the Tukey test, there is a significant difference between the two means being compared.

The critical value for the Tukey test is found using Table N in Appendix A, where k is the number of means in the original problem and v is the degrees of freedom for s_W^2, which is $N - k$. The value of k is found across the top row, and v is found in the left column.

EXAMPLE 12–4

Using the Tukey test, test each pair of means in Example 12–2 to see whether a specific difference exists, at $\alpha = 0.05$.

SOLUTION

a. For $\overline{X}_1$ versus $\overline{X}_2$,

$$q = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s_W^2/n}} = \frac{15.5 - 4.0}{\sqrt{45.50/5}} = 3.812$$

b. For $\overline{X}_1$ versus $\overline{X}_3$,

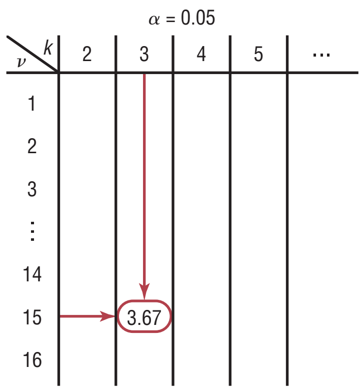
$$q = \frac{\overline{X}_1 - \overline{X}_3}{\sqrt{s_W^2/n}} = \frac{15.5 - 5.8}{\sqrt{45.50/5}} = 3.216$$

c. For $\overline{X}_2$ versus $\overline{X}_3$,

$$q = \frac{\overline{X}_2 - \overline{X}_3}{\sqrt{s_W^2/n}} = \frac{4.0 - 5.8}{\sqrt{45.50/5}} = -0.597$$

To find the critical value for the Tukey test, use Table N in Appendix A. The number of means k is found in the row at the top, and the degrees of freedom for s_W^2 are found in the left column (denoted by v). Since $k = 3$, d.f. $= 18 - 3 = 15$, and $\alpha = 0.05$, the critical value is 3.67. See Figure 12–3. Hence, the only q value that is greater in absolute value than the critical value is the one for the difference between $\overline{X}_1$ and $\overline{X}_2$. The conclusion, then, is that there is a significant difference in means for the turnpike and the Mon-Fayette Expressway.

FIGURE 12–3 Finding the Critical Value in Table N for the Tukey Test (Example 12–4)



You might wonder why there are two different tests that can be used after the ANOVA. Actually, there are several other tests that can be used in addition to the Scheffé and Tukey tests. It is up to the researcher to select the most appropriate test. The Scheffé test is the most general, and it can be used when the samples are of different sizes. Furthermore, the Scheffé test can be used to make comparisons such as the average of $\overline{X}_1$ and $\overline{X}_2$ compared with $\overline{X}_3$. However, the Tukey test is more powerful than the Scheffé test for making pairwise comparisons for the means. A rule of thumb for pairwise comparisons is to use the Tukey test when the samples are equal in size and the Scheffé test when the samples differ in size. This rule will be followed in this textbook.

Applying the Concepts 12–2

Colors That Make You Smarter

The following set of data values was obtained from a study of people's perceptions on whether the color of a person's clothing is related to how intelligent the person looks. The subjects rated the person's intelligence on a scale of 1 to 10. Randomly selected group 1 subjects were shown people with clothing in shades of blue and gray. Randomly selected group 2 subjects were shown people with clothing in shades of brown and yellow. Randomly selected group 3 subjects were shown people with clothing in shades of pink and orange. The results follow.

Group 1	Group 2	Group 3
8	7	4
7	8	9
7	7	6
7	7	7
8	5	9
8	8	8
6	5	5
8	8	8
8	7	7
7	6	5
7	6	4
8	6	5
8	6	4

1. Use the Tukey test to test all possible pairwise comparisons.
2. Are there any contradictions in the results?
3. Explain why separate *t* tests are not accepted in this situation.
4. When would Tukey's test be preferred over the Scheffé method? Explain.

See page 686 for the answers.

Exercises 12–2

1. What two tests can be used to compare two means when the null hypothesis is rejected using the one-way ANOVA *F* test?

2. Explain the difference between the two tests used to compare two means when the null hypothesis is rejected using the one-way ANOVA *F* test.

For Exercises 3 through 8, the null hypothesis was rejected. Use the Scheffé test when sample sizes are unequal or the Tukey test when sample sizes are equal, to test the differences between the pairs of means. Assume all variables are normally distributed, samples are independent, and the population variances are equal.

3. Exercise 9 in Section 12–1.

4. Exercise 12 in Section 12–1.

5. Exercise 13 in Section 12–1.

6. Exercise 17 in Section 12–1.

7. Exercise 18 in Section 12–1.

8. Exercise 20 in Section 12–1.

For Exercises 9 through 13, do a complete one-way ANOVA. If the null hypothesis is rejected, use either the Scheffé or Tukey test to see if there is a significant difference in the pairs of means. Assume all assumptions are met.

9. Emergency Room Visits Fractures accounted for 2.7% of all U.S. emergency room visits for a total of 454,000 visits for a recent year. A random sample of weekly ER visits is recorded for three hospitals in a large metropolitan area during the summer months. At $\alpha = 0.05$, is there sufficient evidence to conclude a difference in means?

Hospital X	Hospital Y	Hospital Z
28	30	25
27	18	20
40	34	30
45	28	22
29	26	18
25	31	20

Source: *World Almanac.*

10. Weights of Digital Cameras The data consist of the weights in ounces of three different types of digital camera. Use $\alpha = 0.05$ to see if the means are equal.

2–3 Megapixels	4–5 Megapixels	6–8 Megapixels
6	14	19
8	11	27
7	15	21
11	24	23
4	17	24
8	10	33

11. Fiber Content of Foods The number of grams of fiber per serving for a random sample of three different kinds of foods is listed. Is there sufficient evidence at the 0.05 level of significance to conclude that there is a difference in mean fiber content among breakfast cereals, fruits, and vegetables?

Breakfast cereals	Fruits	Vegetables
3	5.5	10
4	2	1.5
6	4.4	3.5
4	1.6	2.7
10	3.8	2.5
5	4.5	6.5
6	2.8	4
8		3
5		

Source: *The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.*

12. Per-Pupil Expenditures The expenditures (in dollars) per pupil for states in three sections of the country are listed. Using $\alpha = 0.05$, can you conclude that there is a difference in means?

Eastern third	Middle third	Western third
4946	6149	5282
5953	7451	8605
6202	6000	6528
7243	6479	6911
6113		

Source: *New York Times Almanac.*

13. Weekly Unemployment Benefits The average weekly unemployment benefit for the entire United States is \$297. Three states are randomly selected, and a sample of weekly unemployment benefits is recorded for each. At $\alpha = 0.05$, is there sufficient evidence to conclude a difference in means? If so, perform the appropriate test to find out where the difference exists.

Florida	Pennsylvania	Maine
200	300	250
187	350	195
192	295	275
235	362	260
260	280	220
175	340	290

Source: *World Almanac.*

12–3 Two-Way Analysis of Variance

OBJECTIVE 3

Use the two-way ANOVA technique to determine if there is a significant difference in the main effects or interaction.

The analysis of variance technique shown previously is called a **one-way ANOVA** since there is only *one independent variable.* The **two-way ANOVA** is an extension of the one-way analysis of variance; it involves *two independent variables.* The independent variables are also called **factors.**

The two-way analysis of variance is quite complicated, and many aspects of the subject should be considered when you are using a research design involving a two-way ANOVA. For the purposes of this textbook, only a brief introduction to the subject will be given.

In doing a study that involves a two-way analysis of variance, the researcher is able to test the effects of two independent variables or factors on one *dependent variable.* In addition, the interaction effect of the two variables can be tested.

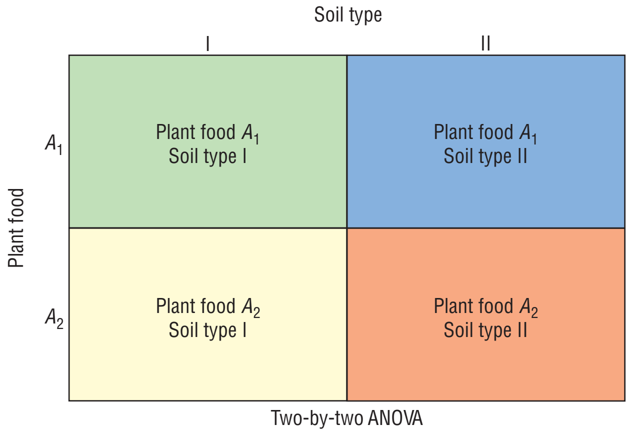


FIGURE 12–4
Treatment Groups for the Plant Food–Soil Type Experiment

For example, suppose a researcher wishes to test the effects of two different types of plant food and two different types of soil on the growth of certain plants. The two independent variables are the type of plant food and the type of soil, while the dependent variable is the plant growth. Other factors, such as water, temperature, and sunlight, are held constant.

To conduct this experiment, the researcher sets up four groups of plants. See Figure 12–4. Assume that the plant food type is designated by the letters A_1 and A_2 and the soil type by the Roman numerals I and II. The groups for such a two-way ANOVA are sometimes called **treatment groups.** The four groups are

Group 1	Plant food A_1, soil type I
Group 2	Plant food A_1, soil type II
Group 3	Plant food A_2, soil type I
Group 4	Plant food A_2, soil type II

The plants are assigned to the groups at random. This design is called a 2×2 (read "two-by-two") design, since each variable consists of two **levels,** that is, two different treatments.

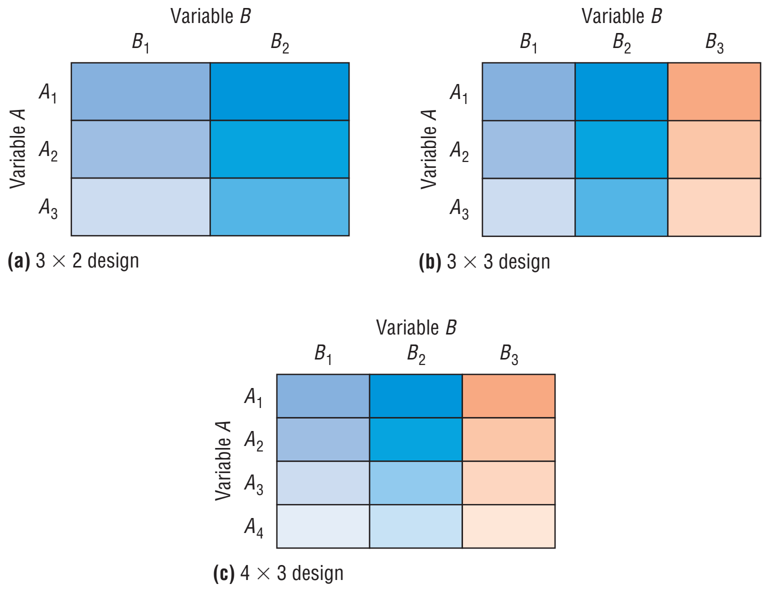
The two-way ANOVA enables the researcher to test the effects of the plant food and the soil type in a single experiment rather than in separate experiments involving the plant food alone and the soil type alone.

In this case, the effect of the plant food is the change in the response variable that results from changing the level or the type of food. The effect of soil type is the change in the response variable that results from changing the level or type of soil. These two effects of the independent variable are called the **main effects.** Furthermore, the researcher can test an additional hypothesis about the effect of the *interaction* of the two variables—plant food and soil type—on plant growth. For example, is there a difference between the growth of plants using plant food A_1 and soil type II and the growth of plants using plant food A_2 and soil type I? When a difference of this type occurs, the experiment is said to have a significant **interaction effect.** The interaction effect represents the joint effect of the two factors over and above the effects of each factor considered separately. That is, the types of plant food affect the plant growth differently in different soil types. When the interaction effect is statistically significant, the researcher should not consider the effects of the individual factors without considering the interaction effect.

There are many different kinds of two-way ANOVA designs, depending on the number of levels of each variable. Figure 12–5 shows a few of these designs. As stated previously, the plant food–soil type experiment uses a 2×2 ANOVA.

The design in Figure 12–5(a) is called a 3×2 design, since the factor in the rows has three levels and the factor in the columns has two levels. Figure 12–5(b) is a 3×3 design,

FIGURE 12–5
Some Types of Two-Way ANOVA Designs



Interesting Facts

As unlikely as it sounds, lightning can travel through phone wires. You should probably hold off on taking a bath or shower as well during an electrical storm. According to the *Annals of Emergency Medicine*, lightning can also travel through water pipes.

since each factor has three levels. Figure 12–5(c) is a 4 × 3 design, since the factor in the rows has four levels and the factor in the columns has three levels.

The two-way ANOVA design has several null hypotheses. There is one for each independent variable and one for the interaction. In the plant food–soil type problem, the hypotheses are as follows:

1. The hypotheses regarding the plant food–soil type interaction effect are stated as follows.

H_0: There is no interaction effect between type of plant food used and type of soil used on plant growth.

H_1: There is an interaction effect between food type and soil type on plant growth.

2. The hypotheses regarding plant food are stated as follows.

H_0: There is no difference in means of heights of plants grown using different foods.

H_1: There is a difference in means of heights of plants grown using different foods.

3. The hypotheses regarding soil type are stated as follows:

H_0: There is no difference in means of heights of plants grown in different soil types.

H_1: There is a difference in means of heights of plants grown in different soil types.

As with the one-way ANOVA, a between-group variance estimate is calculated, and a within-group variance estimate is calculated. An F test is then performed for each of the independent variables and the interaction. The results of the two-way ANOVA are summarized in a two-way table, as shown in Table 12–4 for the plant experiment.

TABLE 12–4 ANOVA Summary Table for Plant Food and Soil Type

Source	Sum of squares	d.f.	Mean square	*F*
Plant food				
Soil type				
Interaction				
Within (error)				
Total				

In general, the two-way ANOVA summary table is set up as shown in Table 12–5.

TABLE 12–5 ANOVA Summary Table

Source	Sum of squares	d.f.	Mean square	F
A	SS_A	$a - 1$	MS_A	F_A
B	SS_B	$b - 1$	MS_B	F_B
$A \times B$	$SS_{A\times B}$	$(a-1)(b-1)$	$MS_{A\times B}$	$F_{A\times B}$
Within (error)	SS_W	$ab(n-1)$	MS_W	
Total				

In the table,

$$\begin{aligned}
SS_A &= \text{sum of squares for factor } A \\
SS_B &= \text{sum of squares for factor } B \\
SS_{A\times B} &= \text{sum of squares for interaction} \\
SS_W &= \text{sum of squares for within-group term or error term} \\
a &= \text{number of levels of factor } A \\
b &= \text{number of levels of factor } B \\
n &= \text{number of subjects in each group} \\
MS_A &= \frac{SS_A}{a-1} \\
MS_B &= \frac{SS_B}{b-1} \\
MS_{A\times B} &= \frac{SS_{A\times B}}{(a-1)(b-1)} \\
MS_W &= \frac{SS_W}{ab(n-1)} \\
F_A &= \frac{MS_A}{MS_W} \quad \text{with d.f.N.} = a-1\text{, d.f.D.} = ab(n-1) \\
F_B &= \frac{MS_B}{MS_W} \quad \text{with d.f.N.} = b-1\text{, d.f.D.} = ab(n-1) \\
F_{A\times B} &= \frac{MS_{A\times B}}{MS_W} \quad \text{with d.f.N.} = (a-1)(b-1)\text{, d.f.D.} = ab(n-1)
\end{aligned}$$

The assumptions for the two-way analysis of variance are basically the same as those for the one-way ANOVA, except for sample size.

Assumptions for the Two-Way ANOVA

1. The populations from which the samples were obtained must be normally or approximately normally distributed.
2. The samples must be independent.
3. The variances of the populations from which the samples were selected must be equal.
4. The groups must be equal in sample size.

In this book, the assumptions will be stated in the exercises; however, when encountering statistics in other situations, you must check to see that these assumptions have been met before proceeding.

The two-way analysis of variance follows the regular five-step hypothesis-testing procedure.

Step 1 State the hypotheses.

Step 2 Find the critical values.

Step 3 Compute the test value.

Step 4 Make the decision.

Step 5 Summarize the results.

The computational procedure for the two-way ANOVA is quite lengthy. For this reason, it will be omitted in Example 12–5, and only the two-way ANOVA summary table will be shown. The table used in Example 12–5 is similar to the one generated by most computer programs. You should be able to interpret the table and summarize the results.

EXAMPLE 12–5 Gasoline Consumption

A researcher wishes to see whether the type of gasoline used and the type of automobile driven have any effect on gasoline consumption. Two types of gasoline, regular and high-octane, will be used, and two types of automobiles, two-wheel- and all-wheel-drive, will be used in each group. There will be two automobiles in each group, for a total of eight automobiles used. Using a two-way analysis of variance, determine if there is an interactive effect, an effect due to the type of gasoline used, and an effect due to the type of vehicle driven.

The data (in miles per gallon) are shown here, and the summary table is given in Table 12–6.

	Type of automobile	
Gas	**Two-wheel-drive**	**All-wheel-drive**
Regular	26.7 25.2	28.6 29.3
High-octane	32.3 32.8	26.1 24.2

TABLE 12–6 ANOVA Summary Table for Example 12–5

Source	SS	d.f.	MS	F
Gasoline A	3.920			
Automobile B	9.680			
Interaction ($A \times B$)	54.080			
Within (error)	3.300			
Total	70.980			

Unusual Stats

Of Americans born today, one-third of the women will reach age 100, compared to only 10% of the men, according to Ronald Klatz, M.D., president of the American Academy of Anti-Aging Medicine.

SOLUTION

Step 1 State the hypotheses. The hypotheses for the interaction are

H_0: There is no interaction effect between type of gasoline used and type of automobile a person drives on gasoline consumption.

H_1: There is an interaction effect between type of gasoline used and type of automobile a person drives on gasoline consumption.

The hypotheses for the gasoline types are

H_0: There is no difference between the means of gasoline consumption for two types of gasoline.

H_1: There is a difference between the means of gasoline consumption for two types of gasoline.

The hypotheses for the types of automobile driven are

H_0: There is no difference between the means of gasoline consumption for two-wheel-drive and all-wheel-drive automobiles.

H_1: There is a difference between the means of gasoline consumption for two-wheel-drive and all-wheel-drive automobiles.

Step 2 Find the critical values for each F test. In this case, each independent variable, or factor, has two levels. Hence, a 2×2 ANOVA table is used. Factor A is designated as the gasoline type. It has two levels, regular and high-octane; therefore, $a = 2$. Factor B is designated as the automobile type. It also has two levels; therefore, $b = 2$. The degrees of freedom for each factor are as follows:

$$\begin{aligned}
\text{Factor } A\text{:} &\quad \text{d.f.N.} = a - 1 = 2 - 1 = 1\\
\text{Factor } B\text{:} &\quad \text{d.f.N.} = b - 1 = 2 - 1 = 1\\
\text{Interaction } (A \times B)\text{:} &\quad \text{d.f.N.} = (a - 1)(b - 1)\\
&\quad \phantom{\text{d.f.N.}} = (2 - 1)(2 - 1) = 1 \cdot 1 = 1\\
\text{Within (error):} &\quad \text{d.f.D.} = ab(n - 1)\\
&\quad \phantom{\text{d.f.D.}} = 2 \cdot 2(2 - 1) = 4
\end{aligned}$$

where n is the number of data values in each group. In this case, $n = 2$.

The critical value for the F_A test is found by using $\alpha = 0.05$, d.f.N. $= 1$, and d.f.D. $= 4$. In this case, $F_A = 7.71$. The critical value for the F_B test is found by using $\alpha = 0.05$, d.f.N. $= 1$, and d.f.D. $= 4$; also F_B is 7.71. Finally, the critical value for the $F_{A \times B}$ test is found by using d.f.N. $= 1$ and d.f.D. $= 4$; it is also 7.71.

Note: If there are different levels of the factors, the critical values will not all be the same. For example, if factor A has three levels and factor b has four levels, and if there are two subjects in each group, then the degrees of freedom are as follows:

$$\begin{aligned}
&\text{d.f.N.} = a - 1 = 3 - 1 = 2 && \text{factor } A\\
&\text{d.f.N.} = b - 1 = 4 - 1 = 3 && \text{factor } B\\
&\text{d.f.N.} = (a - 1)(b - 1) = (3 - 1)(4 - 1) &&\\
&\phantom{\text{d.f.N.}} = 2 \cdot 3 = 6 && \text{factor } A \times B\\
&\text{d.f.D.} = ab(n - 1) = 3 \cdot 4(2 - 1) = 12 && \text{within (error) factor}
\end{aligned}$$

Step 3 Complete the ANOVA summary table to get the test values. The mean squares are computed first.

$$\text{MS}_A = \frac{\text{SS}_A}{a - 1} = \frac{3.920}{2 - 1} = 3.920$$

$$\text{MS}_B = \frac{\text{SS}_B}{b - 1} = \frac{9.680}{2 - 1} = 9.680$$

$$\text{MS}_{A \times B} = \frac{\text{SS}_{A \times B}}{(a - 1)(b - 1)} = \frac{54.080}{(2 - 1)(2 - 1)} = 54.080$$

$$\text{MS}_W = \frac{\text{SS}_W}{ab(n - 1)} = \frac{3.300}{4} = 0.825$$

The F values are computed next.

$$F_A = \frac{\text{MS}_A}{\text{MS}_W} = \frac{3.920}{0.825} = 4.752 \qquad \text{d.f.N.} = a - 1 = 1 \qquad \text{d.f.D.} = ab(n - 1) = 4$$

$$F_B = \frac{\text{MS}_B}{\text{MS}_W} = \frac{9.680}{0.825} = 11.733 \qquad \text{d.f.N.} = b - 1 = 1 \qquad \text{d.f.D.} = ab(n - 1) = 4$$

$$F_{A \times B} = \frac{\text{MS}_{A \times B}}{\text{MS}_W} = \frac{54.080}{0.825} = 65.552 \qquad \text{d.f.N.} = (a - 1)(b - 1) = 1 \qquad \text{d.f.D.} = ab(n - 1) = 4$$

The completed ANOVA table is shown in Table 12–7.

TABLE 12–7 Completed ANOVA Summary Table for Example 12–5

Source	SS	d.f.	MS	*F*
Gasoline *A*	3.920	1	3.920	4.752
Automobile *B*	9.680	1	9.680	11.733
Interaction ($A \times B$)	54.080	1	54.080	65.552
Within (error)	3.300	4	0.825	
Total	70.980	7		

Step 4 Make the decision. Since $F_B = 11.733$ and $F_{A\times B} = 65.552$ are greater than the critical value 7.71, the null hypotheses concerning the type of automobile driven and the interaction effect should be rejected. Since the interaction effect is statistically significant, no decision should be made about the automobile type without further investigation.

Step 5 Summarize the results. Since the null hypothesis for the interaction effect was rejected, it can be concluded that the combination of type of gasoline and type of automobile does affect gasoline consumption.

Interesting Fact

Some birds can fly as high as 5 miles.

In the preceding analysis, the effect of the type of gasoline used and the effect of the type of automobile driven are called the *main effects*. If there is no significant interaction effect, the main effects can be interpreted independently. However, if there is a significant interaction effect, the main effects must be interpreted cautiously, if at all.

To interpret the results of a two-way analysis of variance, researchers suggest drawing a graph, plotting the means of each group, analyzing the graph, and interpreting the results. In Example 12–5, find the means for each group or cell by adding the data values in each cell and dividing by n. The means for each cell are shown in the chart here.

	Type of automobile	
Gas	**Two-wheel-drive**	**All-wheel-drive**
Regular	$\bar{X} = \dfrac{26.7 + 25.2}{2} = 25.95$	$\bar{X} = \dfrac{28.6 + 29.3}{2} = 28.95$
High-octane	$\bar{X} = \dfrac{32.3 + 32.8}{2} = 32.55$	$\bar{X} = \dfrac{26.1 + 24.2}{2} = 25.15$

The graph of the means for each of the variables is shown in Figure 12–6. In this graph, the lines cross each other. When such an intersection occurs and the interaction is significant, the interaction is said to be a **disordinal interaction.** When there is a disordinal interaction, you should not interpret the main effects without considering the interaction effect.

The other type of interaction that can occur is an *ordinal interaction.* Figure 12–7 shows a graph of means in which an ordinal interaction occurs between two variables. The lines do not cross each other, nor are they parallel. If the F test value for the interaction is significant and the lines do not cross each other, then the interaction is said to be an **ordinal interaction** and the main effects can be interpreted independently of each other.

Finally, when there is no significant interaction effect, the lines in the graph will be parallel or approximately parallel. When this situation occurs, the main effects can be interpreted independently of each other because there is no significant interaction. Figure 12–8 shows the graph of two variables when the interaction effect is not significant; the lines are parallel.

FIGURE 12–6
Graph of the Means of the Variables in Example 12–5

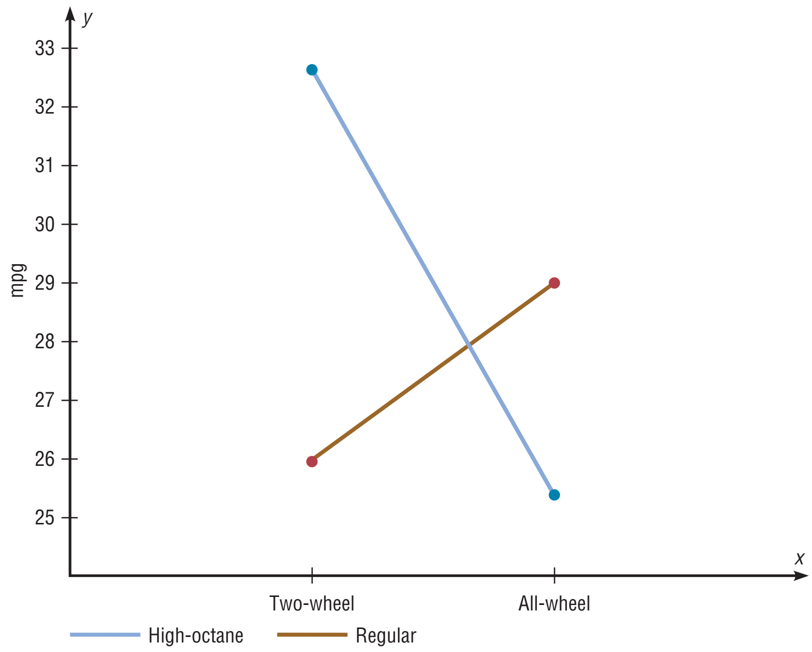
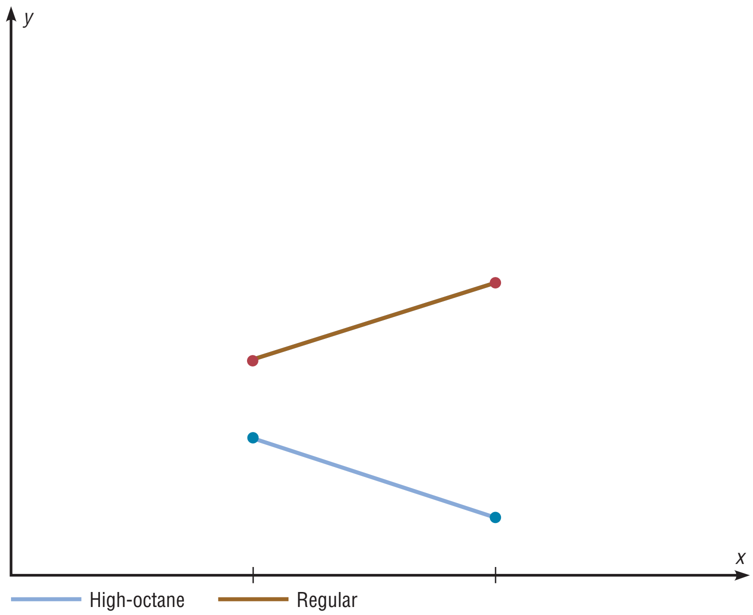
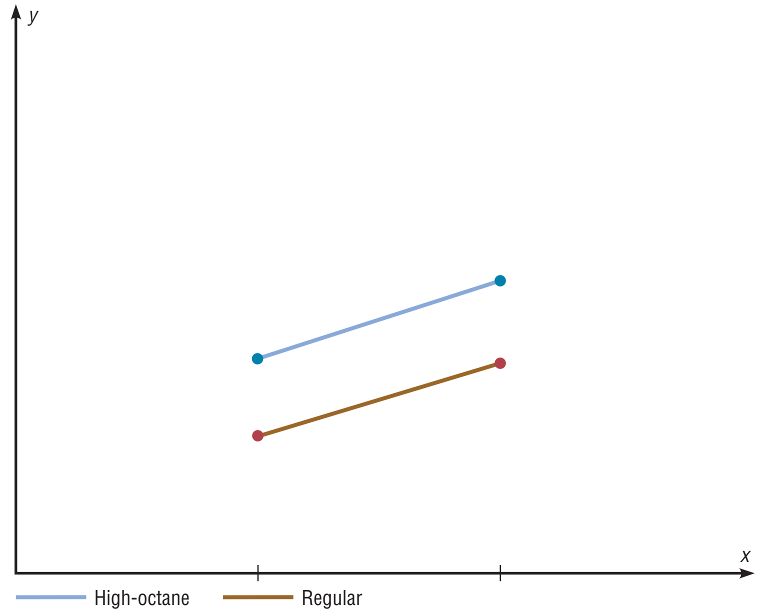


FIGURE 12–7
Graph of Two Variables Indicating an Ordinal Interaction



Example 12–5 was an example of a 2 × 2 two-way analysis of variance, since each independent variable had two levels. For other types of variance problems, such as a 3 × 2 or a 4 × 3 ANOVA, interpretation of the results can be quite complicated. Procedures using tests such as the Tukey and Scheffé tests for analyzing the cell means exist and are similar to the tests shown for the one-way ANOVA, but they are beyond the scope of this textbook. Many other designs for analysis of variance are available to researchers, such as three-factor designs and repeated-measure designs; they are also beyond the scope of this book.

FIGURE 12–8
Graph of Two Variables Indicating No Interaction



In summary, the two-way ANOVA is an extension of the one-way ANOVA. The former can be used to test the effects of two independent variables and a possible interaction effect on a dependent variable.

Applying the Concepts 12–3

Automobile Sales Techniques

The following outputs are from the result of an analysis of how car sales are affected by the experience of the salesperson and the type of sales technique used. Experience was broken up into four levels, and two different sales techniques were used. Analyze the results and draw conclusions about level of experience with respect to the two different sales techniques and how they affect car sales.

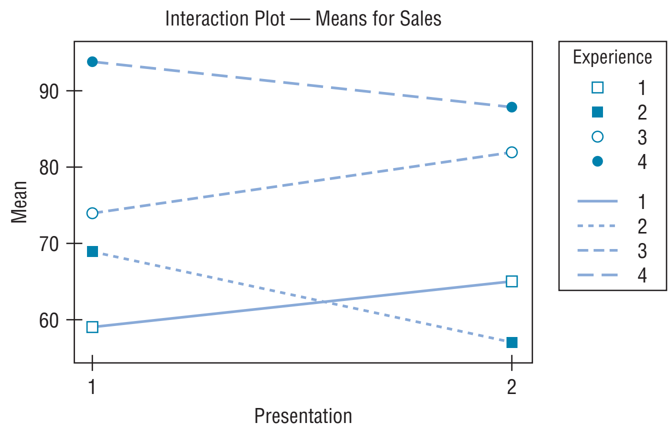
Two-Way Analysis of Variance

Analysis of Variance for Sales

Source	DF	SS	MS
Experience	3	3414.0	1138.0
Presentation	1	6.0	6.0
Interaction	3	414.0	138.0
Error	16	838.0	52.4
Total	23	4672.0	

```
                              Individual 95% CI
Experience          Mean      ------------+------------+------------+------------+------------
1                   62.0       (----------*----------)
2                   63.0         (----------*----------)
3                   78.0                        (-----------*-----------)
4                   91.0                                  (-----------*-----------)
                              -----------+-----------+-----------+-----------+-----------
                                    60.0        70.0        80.0        90.0

                              Individual 95% CI
Presentation        Mean      -----------+-----------+-----------+-----------+-----------
1                   74.0           (--------------------------*--------------------------)
2                   73.0      (-------------------------*------------------------)
                              -----------+-----------+-----------+-----------+-----------
```



See page 686 for the answers.

Exercises 12–3

1. How does the two-way ANOVA differ from the one-way ANOVA?

2. Explain what is meant by *main effects* and *interaction effect.*

3. How are the values for the mean squares computed?

4. How are the *F* test values computed?

5. In a two-way ANOVA, variable *A* has three levels and variable *B* has two levels. There are five data values in each cell. Find each degrees-of-freedom value.

a. d.f.N. for factor *A*

b. d.f.N. for factor *B*

c. d.f.N. for factor $A \times B$

d. d.f.D. for the within (error) factor

6. In a two-way ANOVA, variable *A* has six levels and variable *B* has five levels. There are seven data values in each cell. Find each degrees-of-freedom value.

a. d.f.N. for factor *A*

b. d.f.N. for factor *B*

c. d.f.N. for factor $A \times B$

d. d.f.D. for the within (error) factor

7. What are the two types of interactions that can occur in the two-way ANOVA?

8. When can the main effects for the two-way ANOVA be interpreted independently?

For Exercises 9 through 15, perform these steps. Assume that all variables are normally or approximately normally distributed, that the samples are independent, and that the population variances are equal.

a. State the hypotheses.

b. Find the critical value for each *F* test.

c. Complete the summary table and find the test value.

d. Make the decision.

e. Summarize the results. (Draw a graph of the cell means if necessary.)

9. Soap Bubble Experiments Hands-on soap bubble experiments are a great way to teach mathematics. In an effort to find the best possible bubble solution, two different soap concentrations were used along with two different amounts of glycerin additive. Students were then given a flat glass plate and a straw and were asked to blow their best bubble. The diameters of the resulting bubbles (in millimeters) are listed below. Can an interaction be concluded between the soap solution and the glycerin? Is there a difference in mean length of bubble diameter with respect to the concentration of soap to water? With respect to amount of glycerin additive? Use $\alpha = 0.05$.

	+1 Unit glycerin	**+2 Units glycerin**
Soap:water 13:25	115, 113, 105, 110	98, 100, 90, 95
Soap:water 1:2	90, 102, 100, 98	99, 100, 102, 95

10. Increasing Plant Growth A gardening company is testing new ways to improve plant growth. Twelve plants are randomly selected and exposed to a combination of two factors, a "Grow-light" in two different strengths and a plant food supplement with different mineral supplements. After a number of days, the plants are measured for growth, and the results (in inches) are put into the appropriate boxes.

	Grow-light 1	Grow-light 2
Plant food A	9.2, 9.4, 8.9	8.5, 9.2, 8.9
Plant food B	7.1, 7.2, 8.5	5.5, 5.8, 7.6

Can an interaction between the two factors be concluded? Is there a difference in mean growth with respect to light? With respect to plant food? Use $\alpha = 0.05$.

11. **Environmentally Friendly Air Freshener** As a new type of environmentally friendly, natural air freshener is being developed, it is tested to see whether the effects of temperature and humidity affect the length of time that the scent is effective. The numbers of days that the air freshener had a significant level of scent are listed below for two temperature and humidity levels. Can an interaction between the two factors be concluded? Is there a difference in mean length of effectiveness with respect to humidity? With respect to temperature? Use $\alpha = 0.05$.

	Temperature 1	Temperature 2
Humidity 1	35, 25, 26	35, 31, 37
Humidity 2	28, 22, 21	23, 19, 18

12. **Home-Building Times** A contractor wishes to see whether there is a difference in the time (in days) it takes two subcontractors to build three different types of homes. At $\alpha = 0.05$, analyze the data shown here, using a two-way ANOVA. See below for raw data.

Data for Exercise 12

	Home type		
Subcontractor	**I**	**II**	**III**
A	25, 28, 26, 30, 31	30, 32, 35, 29, 31	43, 40, 42, 49, 48
B	15, 18, 22, 21, 17	21, 27, 18, 15, 19	23, 25, 24, 17, 13

ANOVA Summary Table for Exercise 12

Source	SS	d.f.	MS	F
Subcontractor	1672.553			
Home type	444.867			
Interaction	313.267			
Within	328.800			
Total	2759.487			

13. **Durability of Paint** A pigment laboratory is testing both dry additives and solution-based additives to see their effect on the durability rating (a number from 1 to 10) of a finished paint product. The paint to be tested is divided into four equal quantities, and a different combination of the two additives is added to one-fourth of each quantity. After a prescribed number of hours, the durability rating is obtained for each of the 16 samples, and the results are recorded below in the appropriate space.

	Dry additive 1	Dry additive 2
Solution additive A	9, 8, 5, 6	4, 5, 8, 9
Solution additive B	7, 7, 6, 8	10, 8, 6, 7

Can an interaction be concluded between the dry and solution additives? Is there a difference in mean durability rating with respect to dry additive used? With respect to solution additive? Use $\alpha = 0.05$.

14. **Types of Outdoor Paint** Two types of outdoor paint, enamel and latex, were tested to see how long (in months) each lasted before it began to crack, flake, and peel. They were tested in four geographic locations in the United States to study the effects of climate on the paint. At $\alpha = 0.01$, analyze the data shown, using a two-way ANOVA shown below. Each group contained five test panels. See below for raw data.

Data for Exercise 14

	Geographic location			
Type of paint	**North**	**East**	**South**	**West**
Enamel	60, 53, 58, 62, 57	54, 63, 62, 71, 76	80, 82, 62, 88, 71	62, 76, 55, 48, 61
Latex	36, 41, 54, 65, 53	62, 61, 77, 53, 64	68, 72, 71, 82, 86	63, 65, 72, 71, 63

ANOVA Summary Table for Exercise 14

Source	SS	d.f.	MS	*F*
Paint type	12.1			
Location	2501.0			
Interaction	268.1			
Within	2326.8			
Total	5108.0			

15. Age and Sales A company sells three items: swimming pools, spas, and saunas. The owner decides to see whether the age of the sales representative and the type of item affect monthly sales. At $\alpha = 0.05$, analyze the data shown, using a two-way ANOVA. Sales are given in hundreds of dollars for a randomly selected month, and five salespeople were selected for each group.

ANOVA Summary Table for Exercise 15

Source	SS	d.f.	MS	*F*
Age	168.033			
Product	1,762.067			
Interaction	7,955.267			
Within	2,574.000			
Total	12,459.367			

Data for Exercise 15

	Product		
Age of salesperson	**Pool**	**Spa**	**Sauna**
Over 30	56, 23, 52, 28, 35	43, 25, 16, 27, 32	47, 43, 52, 61, 74
30 or under	16, 14, 18, 27, 31	58, 62, 68, 72, 83	15, 14, 22, 16, 27

Technology Step by Step

TI-84 Plus Step by Step

The TI-84 Plus does not have a built-in function for two-way analysis of variance. However, the downloadable program named TWOWAY is available on the Online Learning Center. Follow the instructions for downloading the program.

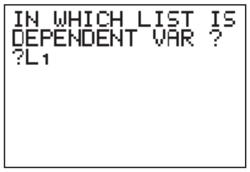
Performing a Two-Way Analysis of Variance

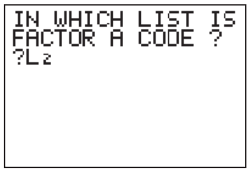
1. Enter the data values of the dependent variable into L_1 and the coded values for the levels of the factors into L_2 and L_3.
2. Press **PRGM,** move the cursor to the program named TWOWAY, and press **ENTER** twice.
3. Type $\mathbf{L_1}$ for the list that contains the dependent variable and press **ENTER.**
4. Type $\mathbf{L_2}$ for the list that contains the coded values for the first factor and press **ENTER.**
5. Type $\mathbf{L_3}$ for the list that contains the coded values for the second factor and press **ENTER.**
6. The program will show the statistics for the first factor.
7. Press **ENTER** to see the statistics for the second factor.
8. Press **ENTER** to see the statistics for the interaction.
9. Press **ENTER** to see the statistics for the error.
10. Press **ENTER** to clear the screen.

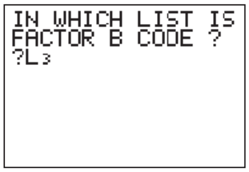
Example TI12–2

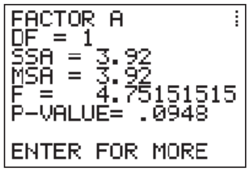
Perform a two-way analysis of variance for the gasoline data (Example 12–5 in the text). The gas mileages are the data values for the dependent variable. Factor *A* is the type of gasoline (1 for regular, 2 for high-octane). Factor *B* is the type of automobile (1 for two-wheel-drive, 2 for all-wheel-drive).

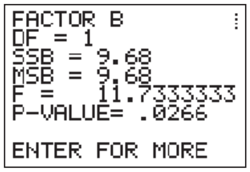
Gas mileages (L_1)	Type of gasoline (L_2)	Type of automobile (L_3)
26.7	1	1
25.2	1	1
32.3	2	1
32.8	2	1
28.6	1	2
29.3	1	2
26.1	2	2
24.2	2	2

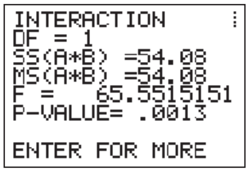


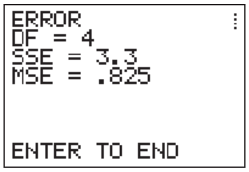












EXCEL
Step by Step

Two-Way Analysis of Variance (ANOVA)

This example pertains to Example 12–5 from the text.

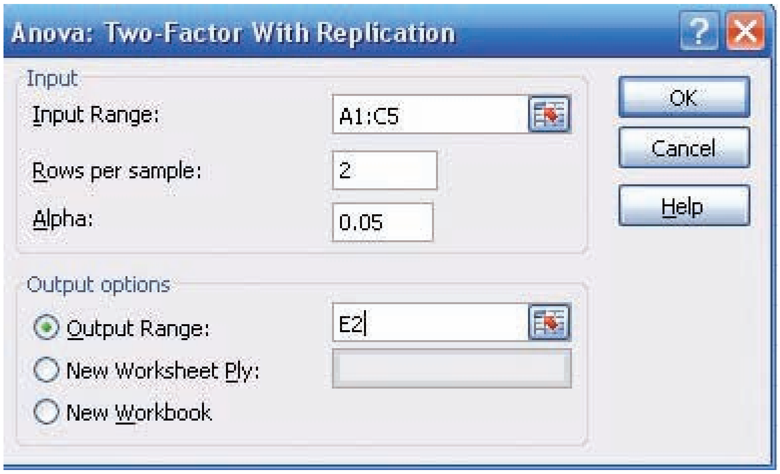
Example XL12–3

A researcher wishes to see if type of gasoline used and type of automobile driven have any effect on gasoline consumption. Use $\alpha = 0.05$.

1. Enter the data exactly as shown in the figure below in an Excel worksheet.

	A	B	C
1		2- wheel Drive	All-wheel drive
2	Regular	26.7	28.6
3		25.2	29.3
4	Hi-octane	32.3	26.1
5		32.8	24.2

2. From the toolbar, select **Data,** then **Data Analysis.**
3. Select **Anova: Two-Factor With Replication** under **Analysis tools,** then [OK].
4. In the **Anova: Single Factor dialog box,** type **A1:C5** for the **Input Range.**
5. Type **2** for the Rows per sample.
6. Type **0.05** for the **Alpha level.**
7. Under **Output options,** check **Output Range** and type **E2.**
8. Click [OK].



The two-way ANOVA table is shown below.

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	3.92	1	3.92	4.751515152	0.094766001	7.708647421
Columns	9.68	1	9.68	11.73333333	0.026647909	7.708647421
Interaction	54.08	1	54.08	65.55151515	0.00126491	7.708647421
Within	3.3	4	0.825			
Total	70.98	7				

MINITAB
Step by Step

Two-Way Analysis of Variance

For Example 12–5, how do gasoline type and vehicle type affect gasoline mileage?

1. Enter the data into three columns of a worksheet. The data for this analysis have to be "stacked" as shown.

a) All the gas mileage data are entered in a single column named MPG.

b) The second column contains codes identifying the gasoline type, a 1 for regular or a 2 for high-octane.

c) The third column will contain codes identifying the type of automobile, 1 for two-wheel-drive or 2 for all-wheel-drive.

↓	C1	C2	C3
	MPG	GasCode	TypeCode
1	26.7	1	1
2	25.2	1	1
3	32.3	2	1
4	32.8	2	1
5	28.6	1	2
6	29.3	1	2
7	26.1	2	2
8	24.2	2	2

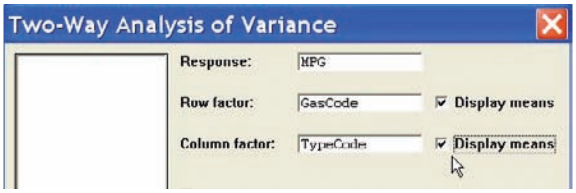
2. Select **Stat>ANOVA>Two-Way.**

a) Double-click MPG in the list box.

b) Double-click GasCode as Row factor.

c) Double-click TypeCode as Column factor.

d) Check the boxes for Display means, then click [OK].

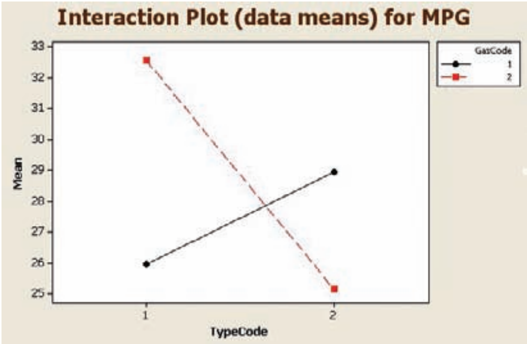


The session window will contain the results.

Two-Way ANOVA: MPG versus GasCode, TypeCode

```
Source        DF      SS       MS       F       P
GasCode        1    3.92    3.920    4.75   0.095
TypeCode       1    9.68    9.680   11.73   0.027
Interaction    1   54.08   54.080   65.55   0.001
Error          4    3.30    0.825
Total          7   70.98

                     Individual 95% CIs For Mean Based on
                     Pooled StDev
GasCode     Mean     --------+--------+--------+--------+-
1          27.45     (-------------*-------------)
2          28.85              (-------------*-------------)
                     --------+--------+--------+--------+-
                         27.0     28.0     29.0     30.0
                     Individual 95% CIs For Mean Based on
                     Pooled StDev
TypeCode    Mean     -----+---------+---------+---------+----
1          29.25                      (-----------*----------)
2          27.05     (----------*------------)
                     -----+---------+---------+---------+----
                        26.4      27.6      28.8      30.0
```



Plot Interactions

3. Select Stat>ANOVA>Interactions Plot.
   a) Double-click MPG for the response variable and GasCodes and TypeCodes for the factors.
   b) Click [OK].

Intersecting lines indicate a significant interaction of the two independent variables.

Summary

- The F test, as shown in Chapter 9, can be used to compare two sample variances to determine whether they are equal. It can also be used to compare three or more means. When three or more means are compared, the technique is called analysis of variance (ANOVA). The ANOVA technique uses two estimates of the population variance. The between-group variance is the variance of the sample means; the within-group variance is the overall variance of all the values. When there is no significant difference among the means, the two estimates will be approximately equal and the F test value will be close to 1. If there is a significant difference among the means, the between-group variance estimate will be larger than the within-group variance estimate and a significant test value will result. (12–1)
- If there is a significant difference among means, the researcher may wish to see where this difference lies. Several statistical tests can be used to compare the sample means after the ANOVA technique has been done. The most common are the Scheffé test and the Tukey test. When the sample sizes are the same, the Tukey test can be used. The Scheffé test is more general and can be used when the sample sizes are equal or not equal. (12–2)
- When there is one independent variable, the analysis of variance is called a one-way ANOVA. When there are two independent variables, the analysis of variance is called a two-way ANOVA. The two-way ANOVA enables the researcher to test the effects of two independent variables and a possible interaction effect on one dependent variable. If an interaction effect is found to be statistically significant, the researcher must investigate further to find out if the main effects can be examined. (12–3)

Important Terms

analysis of variance (ANOVA) 648

ANOVA summary table 650

between-group variance 649

disordinal interaction 671

factors 665

interaction effect 666

level 666

main effects 666

mean square 650

one-way ANOVA 648

ordinal interaction 671

Scheffé test 660

sum of squares between groups 650

sum of squares within groups 650

treatment groups 666

Tukey test 662

two-way ANOVA 665

within-group variance 649

Important Formulas

Formulas for the ANOVA test:

$$\bar{X}_{GM} = \frac{\Sigma X}{N}$$

$$F = \frac{s_B^2}{s_W^2}$$

where

$$s_B^2 = \frac{\Sigma n_i(\bar{X}_i - \bar{X}_{GM})^2}{k - 1} \qquad s_W^2 = \frac{\Sigma(n_i - 1)s_i^2}{\Sigma(n_i - 1)}$$

$$\text{d.f.N.} = k - 1 \qquad N = n_1 + n_2 + \cdots + n_k$$

$$\text{d.f.D.} = N - k \qquad k = \text{number of groups}$$

Formulas for the Scheffé test:

$$F_s = \frac{(\bar{X}_i - \bar{X}_j)^2}{s_W^2[(1/n_i) + (1/n_j)]} \quad \text{and} \quad F' = (k - 1)(\text{C.V.})$$

Formula for the Tukey test:

$$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{s_W^2/n}}$$

$$\text{d.f.N.} = k \quad \text{and} \quad \text{d.f.D.} = \text{degrees of freedom for } s_W^2$$

Formulas for the two-way ANOVA:

$$\text{MS}_A = \frac{\text{SS}_A}{a - 1} \qquad F_A = \frac{\text{MS}_A}{\text{MS}_W} \qquad \text{d.f.N.} = a - 1, \quad \text{d.f.D.} = ab(n - 1)$$

$$\text{MS}_B = \frac{\text{SS}_B}{b - 1} \qquad F_B = \frac{\text{MS}_B}{\text{MS}_W} \qquad \text{d.f.N.} = b - 1, \quad \text{d.f.D.} = ab(n - 1)$$

$$\text{MS}_{A\times B} = \frac{\text{SS}_{A\times B}}{(a - 1)(b - 1)} \qquad F_{A\times B} = \frac{\text{MS}_{A\times B}}{\text{MS}_W} \qquad \text{d.f.N.} = (a - 1)(b - 1), \quad \text{d.f.D.} = ab(n - 1)$$

$$\text{MS}_W = \frac{\text{SS}_W}{ab(n - 1)}$$

Review Exercises

If the null hypothesis is rejected in Exercises 1 through 8, use the Scheffé test when the sample sizes are unequal to test the differences between the means, and use the Tukey test when the sample sizes are equal. For these exercises, perform these steps. Assume the assumptions have been met.

a. State the hypotheses and identify the claim.

b. Find the critical value(s).

c. Compute the test value.

d. Make the decision.

e. Summarize the results.

Sections 12–1 and 12–2

Use the traditional method of hypothesis testing unless otherwise specified.

1. **Lengths of Various Types of Bridges** The data represent the lengths in feet of three types of bridges in the United States. At $\alpha = 0.01$, test the claim that there is no significant difference in the means of the lengths of the types of bridges.

Simple truss	Segmented concrete	Continuous plate
745	820	630
716	750	573
700	790	525
650	674	510
647	660	480
625	640	460
608	636	451
598	620	450
550	520	450
545	450	425
534	392	420
528	370	360

Source: *World Almanac and Book of Facts.*

2. **Number of State Parks** The numbers of state parks found in selected states in three different regions of

the country are listed. At $\alpha = 0.05$, can it be concluded that the average number of state parks differs by region?

South	West	New England
51	28	94
64	44	72
35	24	14
24	31	52
47	40	

Source: *Time Almanac.*

3. **Carbohydrates in Cereals** The number of carbohydrates per serving in randomly selected cereals from three manufacturers is shown. At the 0.05 level of significance, is there sufficient evidence to conclude a difference in the average number of carbohydrates?

Manufacturer 1	Manufacturer 2	Manufacturer 3
25	23	24
26	44	39
24	24	28
26	24	25
26	36	23
41	27	32
26	25	
43		

Source: *The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.*

4. **Grams of Fat per Serving of Pizza** The number of grams of fat per serving for three different kinds of pizza from several manufacturers is listed. At the 0.01 level of significance, is there sufficient evidence that a difference exists in mean fat content?

Cheese	Pepperoni	Supreme/Deluxe
18	20	16
11	17	27
19	15	17
20	18	17
16	23	12
21	23	27
16	21	20

Source: *The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.*

5. **Iron Content of Foods and Drinks** The iron content in three different types of food is shown. At the 0.10 level of significance, is there sufficient evidence to conclude that a difference in mean iron content exists for meats and fish, breakfast cereals, and nutritional high-protein drinks?

Meats and fish	Breakfast cereals	Nutritional drinks
3.4	8	3.6
2.5	2	3.6
5.5	1.5	4.5
5.3	3.8	5.5
2.5	3.8	2.7
1.3	6.8	3.6
2.7	1.5	6.3
	4.5	

Source: *The Doctor's Pocket Calorie, Fat, and Carbohydrate Counter.*

6. **Temperatures in January** The average January high temperatures (in degrees Fahrenheit) for selected tourist cities on different continents are listed. Is there sufficient evidence to conclude a difference in mean temperatures for the three areas? Use the 0.05 level of significance.

Europe	Central and South America	Asia
41	87	89
38	75	35
36	66	83
56	84	67
50	75	48

Source: *Time Almanac.*

7. **School Incidents Involving Police Calls** A researcher wishes to see if there is a difference in the average number of times local police were called in school incidents. Random samples of school districts were selected, and the numbers of incidents for a specific year were reported. At $\alpha = 0.05$, is there a difference in the means? If so, suggest a reason for the difference.

County A	County B	County C	County D
13	16	15	11
11	33	12	31
2	12	19	3
	2	2	
	2		

Source: U.S. Department of Education.

8. **Carbohydrates in Juices** Listed are the numbers of grams of carbohydrates in a random sample of eight-ounce servings of various types of juices. At the 0.01 level of significance, is there evidence of a difference in means?

Apple mix	Orange mix	Veggie mix
23	29	10
31	30	19
26	29	12
32	31	23
30	37	11

Section 12–3

9. **Review Preparation for Statistics** A statistics instructor wanted to see if student participation in review preparation methods led to higher examination scores. Five students were randomly selected and placed in each test group for a three-week unit on statistical inference. Everyone took the same examination at the end of the unit, and the resulting scores are shown. Is there sufficient evidence at $\alpha = 0.05$ to conclude an interaction between the two factors? Is there sufficient

evidence to conclude a difference in mean scores based on formula delivery system? Is there sufficient evidence to conclude a difference in mean scores based on the review organization technique?

	Formulas provided	Student-made formula cards
Student-led review	89, 76, 80, 90, 75	94, 86, 80, 79, 82
Instructor-led review	75, 80, 68, 65, 79	88, 78, 85, 65, 72

10. **Effects of Different Types of Diets** A medical researcher wishes to test the effects of two different diets and two different exercise programs on the glucose level in a person's blood. The glucose level is measured in milligrams per deciliter (mg/dl). Three subjects are randomly assigned to each group. Analyze the data shown here, using a two-way ANOVA with $\alpha = 0.05$.

Exercise program	Diet A	Diet B
I	62, 64, 66	58, 62, 53
II	65, 68, 72	83, 85, 91

ANOVA Summary Table for Exercise 10

Source	SS	d.f.	MS	*F*
Exercise	816.750			
Diet	102.083			
Interaction	444.083			
Within	108.000			
Total	1470.916			

STATISTICS TODAY

Is Seeing Really Believing? —Revisited

To see if there were differences in the testimonies of the witnesses in the three age groups, the witnesses responded to 17 questions, 10 on direct examination and 7 on cross-examination. These were then scored for accuracy. An analysis of variance test with age as the independent variable was used to compare the total number of questions answered correctly by the groups. The results showed no significant differences among the age groups for the direct examination questions. However, there was a significant difference among the groups on the cross-examination questions. Further analysis showed the 8-year-olds were significantly less accurate under cross-examination compared to the other two groups. The 12-year-old and adult eyewitnesses did not differ in the accuracy of their cross-examination responses.

Data Analysis

The Data Bank is found in Appendix B, or on the World Wide Web by following links from www.mhhe.com/math/stat/bluman

1. From the Data Bank, select a random sample of subjects, and test the hypothesis that the mean cholesterol levels of the nonsmokers, less-than-one-pack-a-day smokers, and one-pack-plus smokers are equal. Use an ANOVA test. If the null hypothesis is rejected, conduct the Scheffé test to find where the difference is. Summarize the results.
2. Repeat Exercise 1 for the mean IQs of the various educational levels of the subjects.
3. Using the Data Bank, randomly select 12 subjects and randomly assign them to one of the four groups in the following classifications.

	Smoker	Nonsmoker
Male		
Female		

Use one of these variables—weight, cholesterol, or systolic pressure—as the dependent variable, and perform a two-way ANOVA on the data. Use a computer program to generate the ANOVA table.

Chapter Quiz

Determine whether each statement is true or false. If the statement is false, explain why.

1. In analysis of variance, the null hypothesis should be rejected only when there is a significant difference among all pairs of means.
2. The F test does not use the concept of degrees of freedom.
3. When the F test value is close to 1, the null hypothesis should be rejected.
4. The Tukey test is generally more powerful than the Scheffé test for pairwise comparisons.

Select the best answer.

5. Analysis of variance uses the _______ test.

   a. z *c.* χ^2
   b. t *d.* F

6. The null hypothesis in ANOVA is that all the means are _______.

   a. Equal *c.* Variable
   b. Unequal *d.* None of the above

7. When you conduct an F test, _______ estimates of the population variance are compared.

   a. Two *c.* Any number of
   b. Three *d.* No

8. If the null hypothesis is rejected in ANOVA, you can use the _______ test to see where the difference in the means is found.

   a. z or t *c.* Scheffé or Tukey
   b. F or χ^2 *d.* Any of the above

Complete the following statements with the best answer.

9. When three or more means are compared, you use the _______ technique.
10. If the null hypothesis is rejected in ANOVA, the _______ test should be used when sample sizes are equal.

For Exercises 11 through 17, use the traditional method of hypothesis testing unless otherwise specified. Assume the assumptions have been met.

11. **Gasoline Prices** Random samples of summer gasoline prices per gallon are listed for three different states. Is there sufficient evidence of a difference in mean prices? Use $\alpha = 0.01$.

State 1	State 2	State 3
3.20	3.68	3.70
3.25	3.50	3.65
3.18	3.70	3.75
3.15	3.65	3.72

12. **Voters in Presidential Elections** In a recent Presidential election, a random sample of the percentage of voters who voted is shown. At $\alpha = 0.05$, is there a difference in the mean percentage of voters who voted?

Northeast	Southeast	Northwest	Southwest
65.3	54.8	60.5	42.3
59.9	61.8	61.0	61.2
66.9	49.6	74.0	54.7
64.2	58.6	61.4	56.7

Source: Committee for the Study of the American Electorate.

13. **Ages of Late-Night TV Talk Show Viewers** A media researcher wanted to see if there was a difference in the ages of viewers of three late-night television talk shows. Three random samples of viewers were selected, and the ages of the viewers are shown. At $\alpha = 0.01$, is there a difference in the means of the ages of the viewers? Why is the average age of a viewer important to a television show writer?

David Letterman	Jay Leno	Conan O'Brien
53	48	40
46	51	36
48	57	35
42	46	42
35	38	39

Source: Based on information from Nielsen Media Research.

14. **Prices of Body Soap** A consumer group desired to compare the mean price for 12-ounce bottles of liquid body soap from two nationwide brands and one store brand. Four different bottles of each were randomly selected at a large discount drug store, and the prices are noted. At the 0.05 level of significance, is there sufficient evidence to conclude a difference in mean prices? If so, perform the appropriate test to find out where.

Brand X	Brand Y	Store brand
5.99	8.99	4.99
6.99	7.99	3.99
8.59	6.29	5.29
6.49	7.29	4.49

15. **Air Pollution** A lot of different factors contribute to air pollution. One particular factor, particulate matter, was measured for prominent cities of three continents. Particulate matter includes smoke, soot, dust, and liquid droplets from combustion such that the particle is less than 10 microns in diameter and thus capable of reaching deep into the respiratory system. The measurements are listed here. At the 0.05 level of significance, is there sufficient evidence to conclude a

difference in means? If so, perform the appropriate test to find out where the differences in means are.

Asia	Europe	Africa
79	34	33
104	35	16
40	30	43
73	43	

Source: *World Almanac.*

16. Alumni Gift Solicitation Several students volunteered for an alumni phone-a-thon to solicit alumni gifts. The number of calls made by randomly selected students from each class is listed. At $\alpha = 0.05$, is there sufficient evidence to conclude a difference in means?

Freshmen	Sophomores	Juniors	Seniors
25	17	20	20
29	25	24	25
32	20	25	26
15	26	30	32
18	30	15	19
26	28	18	20
35			

17. Diets and Exercise Programs A researcher conducted a study of two different diets and two different exercise programs. Three randomly selected subjects were assigned to each group for one month. The values indicate the amount of weight each lost.

Exercise program \ Diet	A	B
I	5, 6, 4	8, 10, 15
II	3, 4, 8	12, 16, 11

Answer the following questions for the information in the printout shown.

a. What procedure is being used?

b. What are the names of the two variables?

c. How many levels does each variable contain?

d. What are the hypotheses for the study?

e. What are the *F* values for the hypotheses? State which are significant, using the *P*-values.

f. Based on the answers to part *e,* which hypotheses can be rejected?

Computer Printout for Problem 17

Datafile: NONAME.SST Procedure: Two-way ANOVA

TABLE OF MEANS:

	DIET A	B	Row Mean
EX PROG I	5.000	11.000	8.000
II	5.000	13.000	9.000
Col Mean	5.000	12.000	
Tot Mean	8.500		

SOURCE TABLE:

Source	df	Sums of Squares	Mean Square	F Ratio	p-value
DIET	1	147.000	147.000	21.000	0.00180
EX PROG	1	3.000	3.000	0.429	0.53106
DIET X EX P	1	3.000	3.000	0.429	0.53106
Within	8	56.000	7.000		
Total	11	209.000			

Critical Thinking Challenges

Adult Children of Alcoholics

Shown here are the abstract and two tables from a research study entitled "Adult Children of Alcoholics: Are They at Greater Risk for Negative Health Behaviors?" by Arlene E. Hall. Based on the abstract and the tables, answer these questions.

1. What was the purpose of the study?
2. How many groups were used in the study?
3. By what means were the data collected?
4. What was the sample size?
5. What type of sampling method was used?
6. How might the population be defined?
7. What may have been the hypothesis for the ANOVA part of the study?
8. Why was the one-way ANOVA procedure used, as opposed to another test, such as the *t* test?
9. What part of the ANOVA table did the conclusion "ACOAs had significantly lower wellness scores (WS) than non-ACOAs" come from?
10. What level of significance was used?

11. In the following excerpts from the article, the researcher states that

. . . using the Tukey-HSD procedure revealed a significant difference between ACOAs and non-ACOAs, p = 0.05, but no significant difference was found between ACOAs and Unsures or between non-ACOAs and Unsures.

Using Tables 12–8 and 12–9 and the means, explain why the Tukey test would have enabled the researcher to draw this conclusion.

Abstract *The purpose of the study was to examine and compare the health behaviors of adult children of alcoholics (ACOAs) and their non-ACOA peers within a university population. Subjects were 980 undergraduate students from a major university in the East. Three groups (ACOA, non-ACOA, and Unsure) were identified from subjects' responses to three direct questions regarding parental drinking behaviors. A questionnaire was used to collect data for the study. Included were questions related to demographics, parental drinking behaviors, and the College Wellness Check (WS), a health risk appraisal designed especially for college students (Dewey & Cabral, 1986). Analysis of variance procedures revealed that ACOAs had significantly lower wellness scores (WS) than non-ACOAs. Chi-square analyses of the individual variables revealed that ACOAs and non-ACOAs were significantly different on 15 of the 50 variables of the WS. A discriminant analysis procedure revealed the similarities between Unsure subjects and ACOA subjects. The results provide valuable information regarding ACOAs in a nonclinical setting and contribute to our understanding of the influences related to their health risk behaviors.*

TABLE 12–8 Means and Standard Deviations for the Wellness Scores (WS) Group by (N = 945)

Group	N	$\bar{X}$	S.D.
ACOAs	143	69.0	13.6
Non-ACOAs	746	73.2	14.5
Unsure	56	70.1	14.0
Total	945	212.3	42.1

TABLE 12–9 ANOVA of Group Means for the Wellness Scores (WS)

Source	d.f.	SS	MS	F
Between groups	2	2,403.5	1,201.7	5.9*
Within groups	942	193,237.4	205.1	
Total	944	195,640.9		

*$p < 0.01$

Source: Arlene E. Hall, "Adult Children of Alcoholics: Are They at Greater Risk for Negative Health Behaviors?" *Journal of Health Education* 12, no. 4, pp. 232–238.

Data Projects

Use a significance level of 0.05 for all tests.

1. **Business and Finance** Select 10 stocks at random from the Dow Jones Industrials, the NASDAQ, and the S&P 500. For each, note the gain or loss in the last quarter. Use analysis of variance to test the claim that stocks from all three groups have had equal performance.
2. **Sports and Leisure** Use total earnings data for movies that were released in the previous year. Sort them by rating (G, PG, PG13, and R). Is the mean revenue for movies the same regardless of rating?
3. **Technology** Use the data collected in data project 3 of Chapter 2 regarding song lengths. Consider only three genres. For example, use rock, alternative, and hip hop/rap. Conduct an analysis of variance to determine if the mean song lengths for the genres are the same.
4. **Health and Wellness** Select 10 cereals from each of the following categories: cereal targeted at children, cereal targeted at dieters, and cereal that fits neither of the previous categories. For each cereal note its calories per cup (this may require some computation since serving sizes vary for cereals). Use analysis of variance to test the claim that the calorie content of these different types of cereals is the same.
5. **Politics and Economics** Conduct an anonymous survey and ask the participants to identify which of the following categories describes them best: registered Republican, Democrat, or Independent, or not registered to vote. Also ask them to give their age to obtain your data. Use an analysis of variance to determine whether there is a difference in mean age between the different political designations.
6. **Your Class** Split the class into four groups, those whose favorite type of music is rock, whose favorite is country, whose favorite is rap or hip hop, and those whose favorite is another type of music. Make a list of the ages of students for each of the four groups. Use analysis of variance to test the claim that the means for all four groups are equal.

Answers to Applying the Concepts

Section 12–1 Colors That Make You Smarter

1. The ANOVA produces a test statistic of $F = 3.06$, with a P-value of 0.059. We would fail to reject the null hypothesis and find that there is not enough evidence to conclude at $\alpha = 0.05$ that the color of a person's clothing is related to people's perceptions of how intelligent the person looks.
2. The purpose of the study was to determine if the color of a person's clothing is related to people's perceptions of how intelligent the person looks.
3. We would have to perform three separate t tests, which would inflate the error rate.

Section 12–2 Colors That Make You Smarter

1. Tukey's pairwise comparisons show no significant difference in the three pairwise comparisons of the means.
2. This agrees with the nonsignificant results of the general ANOVA test conducted in Applying the Concepts 12–1.
3. The t tests should not be used since they would inflate the error rate.
4. We prefer the Tukey test over the Scheffé test when the samples are all the same size.

Section 12–3 Automobile Sales Techniques

There is no significant difference between levels 1 and 2 of experience. Level 3 and level 4 salespersons did significantly better than those at levels 1 and 2, with level 4 showing the best results, on average. If type of presentation is taken into consideration, the interaction plot shows a significant difference. The best combination seems to be level 4 experience with presentation style 1.

Hypothesis-Testing Summary 2*

7. Test of the significance of the correlation coefficient.

   Example: H_0: $\rho = 0$

   Use a t test:

   $$t = r\sqrt{\frac{n-2}{1-r^2}} \qquad \text{with d.f.} = n-2$$

8. Formula for the F test for the multiple correlation coefficient.

   Example: H_0: $\rho = 0$

   $$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

   $$\text{d.f.N.} = n-k \qquad \text{d.f.D.} = n-k-1$$

9. Comparison of a sample distribution with a specific population.

   Example: H_0: There is no difference between the two distributions.

   Use the chi-square goodness-of-fit test:

   $$\chi^2 = \sum\frac{(O-E)^2}{E}$$

   d.f. = no. of categories − 1

10. Comparison of the independence of two variables.

    Example: H_0: Variable A is independent of variable B.

    Use the chi-square independence test:

    $$\chi^2 = \sum\frac{(O-E)^2}{E}$$

    $$\text{d.f.} = (R-1)(C-1)$$

11. Test for homogeneity of proportions.

    Example: H_0: $p_1 = p_2 = p_3$

    Use the chi-square test:

    $$\chi^2 = \sum\frac{(O-E)^2}{E}$$

    $$\text{d.f.} = (R-1)(C-1)$$

12. Comparison of three or more sample means.

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

    Use the analysis of variance test:

    $$F = \frac{s_B^2}{s_W^2}$$

    where

    $$s_B^2 = \frac{\Sigma n_i(\overline{X}_i - \overline{X}_{GM})^2}{k-1}$$

    $$s_W^2 = \frac{\Sigma(n_i-1)s_i^2}{\Sigma(n_i-1)}$$

    $$\text{d.f.N.} = k-1 \qquad N = n_1 + n_2 + \cdots + n_k$$

    $$\text{d.f.D.} = N-k \qquad k = \text{number of groups}$$

*This summary is a continuation of Hypothesis-Testing Summary 1, at the end of Chapter 9.

13. Test when the F value for the ANOVA is significant. Use the Scheffé test to find what pairs of means are significantly different.

$$F_s = \frac{(\bar{X}_i - \bar{X}_j)^2}{s_W^2[(1/n_i) + (1/n_j)]}$$

$$F' = (k - 1)(\text{C.V.})$$

Use the Tukey test to find which pairs of means are significantly different.

$$q = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{s_W^2/n}} \qquad \begin{array}{l} \text{d.f.N.} = k \\ \text{d.f.D.} = \text{degrees of freedom for } s_W^2 \end{array}$$

14. Test for the two-way ANOVA.

Example:

H_0: There is no significant difference between the variables.

H_0: There is no interaction effect between the variables.

$$\text{MS}_A = \frac{\text{SS}_A}{a - 1}$$

$$\text{MS}_B = \frac{\text{SS}_B}{b - 1}$$

$$\text{MS}_{A\times B} = \frac{\text{SS}_{A\times B}}{(a - 1)(b - 1)}$$

$$\text{MS}_W = \frac{\text{SS}_W}{ab(n - 1)}$$

$$F_A = \frac{\text{MS}_A}{\text{MS}_W} \qquad \begin{array}{l} \text{d.f.N.} = a - 1 \\ \text{d.f.D.} = ab(n - 1) \end{array}$$

$$F_B = \frac{\text{MS}_B}{\text{MS}_W} \qquad \begin{array}{l} \text{d.f.N.} = (b - 1) \\ \text{d.f.D.} = ab(n - 1) \end{array}$$

$$F_{A\times B} = \frac{\text{MS}_{A\times B}}{\text{MS}_W} \qquad \begin{array}{l} \text{d.f.N.} = (a - 1)(b - 1) \\ \text{d.f.D.} = ab(n - 1) \end{array}$$