© mason morfit 2002/Workbook Stock/Jupiterimages

13

INFERENCE ABOUT COMPARING TWO POPULATIONS

13.1 Inference about the Difference between Two Means: Independent Samples

13.2 Observational and Experimental Data

13.3 Inference about the Difference between Two Means: Matched Pairs Experiment

13.4 Inference about the Ratio of Two Variances

13.5 Inference about the Difference between Two Population Proportions

Appendix 13 Review of Chapters 12 and 13

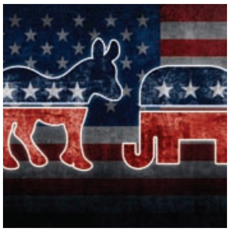
American National Election Survey

Comparing Democrats and Republicans: Who Is More Educated?

DATA ANES2008*

In the business of politics it is important to be able to determine what differences exist between supporters and opponents. In 2008, the American National Election Survey asked people who had indicated that they had completed 13 or more years of education the following question: What is the highest degree earned (DEGREE)?

0. No degree earned
1. Bachelor's degree
2. Master's degree
3. PhD, etc.



© KamiGami/Shutterstock

On page 505 we will provide our answer.

4. LLB, JD
5. MD, DDS, etc.
6. JDC, STD, THD
7. Associate's degree

The survey also asked, Do you think of yourself as Democrat, Republican, Independent, or what (PARTY)?

1 = Democrat, 2 = Republican, 3 = Independent, 4 = Other party, 5 = No preference

Do these data allow us to infer that people who identify themselves as Republican Party supporters are more educated than their Democratic counterparts?

INTRODUCTION

We can compare learning how to use statistical techniques to learning how to drive a car. We began by describing what you are going to do in this course (Chapter 1) and then presented the essential background material (Chapters 2–9). Learning the concepts of statistical inference and applying them the way we did in Chapters 10 and 11 is akin to driving a car in an empty parking lot. You're driving, but it's not a realistic experience. Learning Chapter 12 is like driving on a quiet side street with little traffic. The experience represents real driving, but many of the difficulties have been eliminated. In this chapter, you begin to drive for real, with many of the actual problems faced by licensed drivers, and the experience prepares you to tackle the next difficulty.

In this chapter, we present a variety of techniques used to compare two populations. In Sections 13.1 and 13.3, we deal with interval variables; the parameter of interest is the difference between two means. The difference between these two sections introduces yet another factor that determines the correct statistical method—the design of the experiment used to gather the data. In Section 13.1, the samples are independently drawn, whereas in Section 13.3, the samples are taken from a matched pairs experiment. In Section 13.2, we discuss the difference between observational and experimental data, a distinction that is critical to the way in which we interpret statistical results.

Section 13.4 presents the procedures employed to infer whether two population variances differ. The parameter is the ratio σ_1^2/σ_2^2. (When comparing two variances, we use the ratio rather than the difference because of the nature of the sampling distribution.)

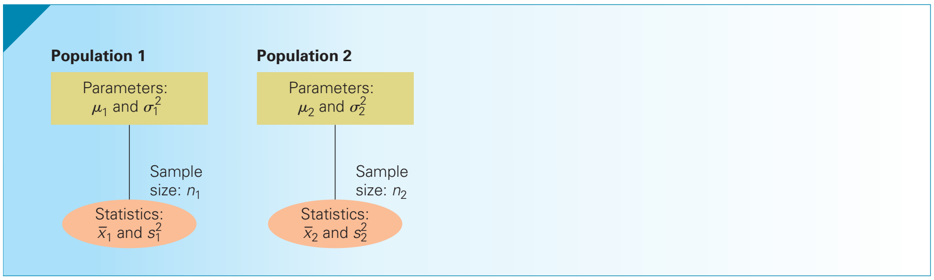
Section 13.5 addresses the problem of comparing two populations of nominal data. The parameter to be tested and estimated is the difference between two proportions.

13.1 INFERENCE ABOUT THE DIFFERENCE BETWEEN TWO MEANS: INDEPENDENT SAMPLES

In order to test and estimate the difference between two population means, the statistics practitioner draws random samples from each of two populations. In this section, we discuss independent samples. In Section 13.3, where we present the matched pairs experiment, the distinction between independent samples and matched pairs will be made clear. For now, we define independent samples as samples completely unrelated to one another.

Figure 13.1 depicts the sampling process. Observe that we draw a sample of size n_1 from population 1 and a sample of size n_2 from population 2. For each sample, we compute the sample means and sample variances.

FIGURE **13.1** Independent Samples from Two Populations



The best estimator of the difference between two population means, $\mu_1 - \mu_2$, is the difference between two sample means, $\bar{x}_1 - \bar{x}_2$. In Section 9.3 we presented the sampling distribution of $\bar{x}_1 - \bar{x}_2$.

Sampling Distribution of $\bar{x}_1 - \bar{x}_2$

1. $\bar{x}_1 - \bar{x}_2$ is normally distributed if the populations are normal and approximately normal if the populations are nonnormal and the sample sizes are large.
2. The expected value of $\bar{x}_1 - \bar{x}_2$ is
$$E(\bar{x}_1 - \bar{x}_2) = \mu_1 - \mu_2$$
3. The variance of $\bar{x}_1 - \bar{x}_2$ is
$$V(\bar{x}_1 - \bar{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$
The standard error of $\bar{x}_1 - \bar{x}_2$ is
$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Thus,
$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$
is a standard normal (or approximately normal) random variable. It follows that the test statistic is
$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$
The interval estimator is
$$(\bar{x}_1 - \bar{x}_2) \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

However, these formulas are rarely used because the population variances σ_1^2 and σ_2^2 are virtually always unknown. Consequently, it is necessary to estimate the standard error of the sampling distribution. The way to do this depends on whether the two unknown population variances are equal. When they are equal, the test statistic is defined in the following way.

Test Statistic for $\mu_1 - \mu_2$ when $\sigma_1^2 = \sigma_2^2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \nu = n_1 + n_2 - 2$$

where

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

The quantity s_p^2 is called the **pooled variance estimator**. It is the weighted average of the two sample variances with the number of degrees of freedom in each sample used as weights. The requirement that the population variances be equal makes this calculation feasible because we need only one estimate of the common value of σ_1^2 and σ_2^2. It makes sense for us to use the pooled variance estimator because, in combining both samples, we produce a better estimate.

The test statistic is Student t distributed with $n_1 + n_2 - 2$ degrees of freedom, provided that the two populations are normal. The confidence interval estimator is derived by mathematics that by now has become routine.

Confidence Interval Estimator of $\mu_1 - \mu_2$ When $\sigma_1^2 = \sigma_2^2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad \nu = n_1 + n_2 - 2$$

We will refer to these formulas as the **equal-variances test statistic** and **confidence interval estimator**, respectively.

When the population variances are unequal, we cannot use the pooled variance estimate. Instead, we estimate each population variance with its sample variance. Unfortunately, the sampling distribution of the resulting statistic

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

is neither normally nor Student t distributed. However, it can be approximated by a Student t distribution with degrees of freedom equal to

$$\nu = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$

(It is usually necessary to round this number to the nearest integer.) The test statistic and confidence interval estimator are easily derived from the sampling distribution.

Test Statistic for $\mu_1 - \mu_2$ When $\sigma_1^2 \neq \sigma_2^2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \qquad \nu = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$

Confidence Interval Estimator of $\mu_1 - \mu_2$ When $\sigma_1^2 \neq \sigma_2^2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)} \qquad \nu = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \frac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$

We will refer to these formulas as the **unequal-variances test statistic** and **confidence interval estimator**, respectively.

The question naturally arises, How do we know when the population variances are equal? The answer is that because σ_1^2 and σ_2^2 are unknown, we can't know for certain whether they're equal. However, we can perform a statistical test to determine whether there is evidence to infer that the population variances differ. We conduct the F-test of the ratio of two variances, which we briefly present here and save the details for Section 13.4.

Testing the Population Variances

The hypotheses to be tested are

$$H_0: \ \sigma_1^2/\sigma_2^2 = 1$$
$$H_1: \ \sigma_1^2/\sigma_2^2 \neq 1$$

The test statistic is the ratio of the sample variances s_1^2/s_2^2, which is F-distributed with degrees of freedom $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 2$. Recall that we introduced the F-distribution in Section 8.4. The required condition is the same as that for the t-test of $\mu_1 - \mu_2$, which is that both populations are normally distributed.

This is a two-tail test so that the rejection region is

$$F > F_{\alpha/2,\nu_1,\nu_2} \quad \text{or} \quad F < F_{1-\alpha/2,\nu_1,\nu_2}$$

Put simply, we will reject the null hypothesis that states that the population variances are equal when the ratio of the sample variances is large or if it is small. Table 6 in Appendix B, which lists the critical values of the F-distribution, defines "large" and "small."

Decision Rule: Equal-Variances or Unequal-Variances *t*-Tests and Estimators

Recall that we can never have enough statistical evidence to conclude that the null hypothesis is true. This means that we can only determine whether there is enough evidence to infer that the population variances *differ*. Accordingly, we adopt the following rule: We will use the equal-variances test statistic and confidence interval estimator unless there is evidence (based on the *F*-test of the population variances) to indicate that the population variances are unequal, in which case we will apply the unequal-variances test statistic and confidence interval estimator.

EXAMPLE 13.1*

Direct and Broker-Purchased Mutual Funds

DATA
Xm13-01

Millions of investors buy mutual funds (see page 181 for a description of mutual funds), choosing from thousands of possibilities. Some funds can be purchased directly from banks or other financial institutions whereas others must be purchased through brokers, who charge a fee for this service. This raises the question, Can investors do better by buying mutual funds directly than by purchasing mutual funds through brokers? To help answer this question, a group of researchers randomly sampled the annual returns from mutual funds that can be acquired directly and mutual funds that are bought through brokers and recorded the net annual returns, which are the returns on investment after deducting all relevant fees. These are listed next.

Direct					Broker				
9.33	4.68	4.23	14.69	10.29	3.24	3.71	16.4	4.36	9.43
6.94	3.09	10.28	−2.97	4.39	−6.76	13.15	6.39	−11.07	8.31
16.17	7.26	7.1	10.37	−2.06	12.8	11.05	−1.9	9.24	−3.99
16.97	2.05	−3.09	−0.63	7.66	11.1	−3.12	9.49	−2.67	−4.44
5.94	13.07	5.6	−0.15	10.83	2.73	8.94	6.7	8.97	8.63
12.61	0.59	5.27	0.27	14.48	−0.13	2.74	0.19	1.87	7.06
3.33	13.57	8.09	4.59	4.8	18.22	4.07	12.39	−1.53	1.57
16.13	0.35	15.05	6.38	13.12	−0.8	5.6	6.54	5.23	−8.44
11.2	2.69	13.21	−0.24	−6.54	−5.75	−0.85	10.92	6.87	−5.72
1.14	18.45	1.72	10.32	−1.06	2.59	−0.28	−2.15	−1.69	6.95

Can we conclude at the 5% significance level that directly purchased mutual funds outperform mutual funds bought through brokers?

SOLUTION

IDENTIFY

To answer the question, we need to compare the population of returns from direct and the returns from broker-bought mutual funds. The data are obviously interval (we've recorded real numbers). This problem objective–data type combination tells us that the parameter to be tested is the difference between two means, $\mu_1 - \mu_2$. The hypothesis to

*Source: D. Bergstresser, J. Chalmers, and P. Tufano, "Assessing the Costs and Benefits of Brokers in the Mutual Fund Industry."

be tested is that the mean net annual return from directly purchased mutual funds (μ_1) is larger than the mean of broker-purchased funds (μ_2). Hence, the alternative hypothesis is

$$H_1: \ (\mu_1 - \mu_2) > 0$$

As usual, the null hypothesis automatically follows:

$$H_0: \ (\mu_1 - \mu_2) = 0$$

To decide which of the t-tests of $\mu_1 - \mu_2$ to apply, we conduct the F-test of σ_1^2/σ_2^2.

$$H_0: \ \sigma_1^2/\sigma_2^2 = 1$$

$$H_1: \ \sigma_1^2/\sigma_2^2 \neq 1$$

COMPUTE

MANUALLY

From the data, we calculated the following statistics:

$$s_1^2 = 37.49 \quad \text{and} \quad s_2^2 = 43.34$$

$$\text{Test statistic: } F = s_1^2/s_2^2 = 37.49/43.34 = 0.86$$

$$\text{Rejection region: } F > F_{\alpha/2,\nu_1,\nu_2} = F_{.025,49,49} \approx F_{.025,50,50} = 1.75$$

or

$$F < F_{1-\alpha/2,\nu_1,\nu_2} = F_{.975,49,49} = 1/F_{.025,49,49} \approx 1/F_{.025,50,50} = 1/1.75 = .57$$

Because $F = .86$ is not greater than 1.75 or smaller than .57, we cannot reject the null hypothesis.

EXCEL

	A	B	C
1	F-Test: Two-Sample for Variances		
2			
3		*Direct*	*Broker*
4	Mean	6.63	3.72
5	Variance	37.49	43.34
6	Observations	50	50
7	df	49	49
8	F	0.8650	
9	P(F<=f) one-tail	0.3068	
10	F Critical one-tail	0.6222	

The value of the test statistic is $F = .8650$. Excel outputs the one-tail p-value. Because we're conducting a two-tail test, we double that value. Thus, the p-value of the test we're conducting is $2 \times .3068 = .6136$.

INSTRUCTIONS

1. Type or import the data into two columns. (Open Xm13-01.)
2. Click **Data, Data Analysis,** and ***F*-test Two-Sample for Variances.**
3. Specify the **Variable 1 Range** (A1:A51) and the **Variable 2 Range** (B1:B51). Type a value for α (.05).

MINITAB

Test for Equal Variances: Direct, Broker

F-Test (Normal Distribution)
Test statistic = 0.86, p-value = 0.614

INSTRUCTIONS

(*Note:* Some of the printout has been omitted.)

1. Type or import the data into two columns. (Open Xm13-01.)
2. Click **Stat, Basic Statistics,** and **2 Variances . . . .**
3. In the **Samples in different columns** box, select the **First** (Direct) and **Second** (Broker) variables.

INTERPRET

There is not enough evidence to infer that the population variances differ. It follows that we must apply the equal-variances t-test of $\mu_1 - \mu_2$.

The hypotheses are

$$H_0{:}(\mu_1 - \mu_2) = 0$$
$$H_1{:}(\mu_1 - \mu_2) > 0$$

COMPUTE

MANUALLY

From the data, we calculated the following statistics:

$$\bar{x}_1 = 6.63$$
$$\bar{x}_2 = 3.72$$
$$s_1^2 = 37.49$$
$$s_2^2 = 43.34$$

The pooled variance estimator is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$
$$= \frac{(50 - 1)37.49 + (50 - 1)43.34}{50 + 50 - 2}$$
$$= 40.42$$

The number of degrees of freedom of the test statistic is

$$\nu = n_1 + n_2 - 2 = 50 + 50 - 2 = 98$$

The rejection region is

$$t > t_{\alpha,\nu} = t_{.05,98} \approx t_{.05,100} = 1.660$$

We determine that the value of the test statistic is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(6.63 - 3.72) - 0}{\sqrt{40.42\left(\frac{1}{50} + \frac{1}{50}\right)}}$$

$$= 2.29$$

EXCEL

	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		*Direct*	*Broker*
4	Mean	6.63	3.72
5	Variance	37.49	43.34
6	Observations	50	50
7	Pooled Variance	40.41	
8	Hypothesized Mean Difference	0	
9	df	98	
10	t Stat	2.29	
11	P(T<=t) one-tail	0.0122	
12	t Critical one-tail	1.6606	
13	P(T<=t) two-tail	0.0243	
14	t Critical two-tail	1.9845	

INSTRUCTIONS

1. Type or import the data into two columns. (Open Xm13-01.)
2. Click **Data, Data Analysis,** and **t-Test: Two-Sample Assuming Equal Variances**.
3. Specify the **Variable 1 Range** (A1:A51) and the **Variable 2 Range** (B1:B51). Type the value of the **Hypothesized Mean Difference*** (0) and type a value for α(.05).

MINITAB

Two-Sample T-Test and CI: Direct, Broker

Two-sample T for Direct vs Broker

	N	Mean	StDev	SE Mean
Direct	50	6.63	6.12	0.87
Broker	50	3.72	6.58	0.93

Difference = mu (Direct) – mu (Broker)
Estimate for difference: 2.91
95% lower bond for difference: 0.80
T-Test of difference = 0 (vs >): T-Value = 2.29 P-Value = 0.012 DF = 98
Both use Pooled StDev = 6.3572

INSTRUCTIONS

1. Type or import the data into two columns. (Open Xm13-01.)
2. Click **Stat, Basic Statistics,** and **2-Sample t . . . .**

*This term is technically incorrect. Because we're testing $\mu_1 - \mu_2$, Excel should ask for and output the "Hypothesized Difference between Means."

3. If the data are stacked, use the **Samples in one column** box to specify the names of the variables. If the data are unstacked (as in Example 13.1), specify the **First** and **Second** variables in the **Samples in different columns** box (Direct, Broker). (See the discussion on Data Formats on page 465 for a discussion of stacked and unstacked data.) Click **Assume equal variances**. Click **Options . . . .**
4. In the **Test difference** box, type the value of the parameter under the null hypothesis (0) and select one of **less than, not equal,** or **greater than** for the **Alternative** hypothesis (greater than).

INTERPRET

The value of the test statistic is 2.29. The one-tail p-value is .0122. We observe that the p-value of the test is small (and the test statistic falls into the rejection region). As a result, we conclude that there is sufficient evidence to infer that on average directly purchased mutual funds outperform broker-purchased mutual funds

Estimating $\mu_1 - \mu_2$: Equal-Variances

In addition to testing a value of the difference between two population means, we can also estimate the difference between means. Next we compute the 95% confidence interval estimate of the difference between the mean return for direct and broker mutual funds.

COMPUTE

MANUALLY

The confidence interval estimator of the difference between two means with equal population variances is

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

The 95% confidence interval estimate of the difference between the return for directly purchased mutual funds and the mean return for broker-purchased mutual funds is

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = (6.63 - 3.72) \pm 1.984\sqrt{40.42\left(\frac{1}{50} + \frac{1}{50}\right)}$$
$$= 2.91 \pm 2.52$$

The lower and upper limits are .39 and 5.43.

EXCEL

	A	B	C	D	E	F
1	t-Estimate : Two Means (Equal Variances)					
2						
3				Direct	Broker	
4	Mean			6.63	3.72	
5	Variance			37.49	43.34	
6	Observations			50	50	
7						
8	Pooled Variance			40.41		
9	Degrees of Freedom			98		
10	Confidence Level			0.95		
11	Confidence Interval Estimate			2.91	±	2.52
12	LCL			0.38		
13	UCL			5.43		

(Continued)

INSTRUCTIONS

1. Type or import the data into two columns. (Open Xm13-01.)
2. Click **Add-Ins, Data Analysis Plus**, and **t-Estimate: Two Means**.
3. Specify the **Variable 1 Range** (A1:A51) and the **Variable 2 Range** (B1:B51). Click **Independent Samples with Equal Variances** and the value for α (.05).

MINITAB

Two-Sample T-Test and CI: Direct, Broker

Two-sample T for Direct vs Broker

	N	Mean	StDev	SE Mean
Direct	50	6.63	6.12	0.87
Broker	50	3.72	6.58	0.93

Difference = mu (Direct) – mu (Broker)
Estimate for difference: 2.91
95% CI for difference: (0.38, 5.43)
T-Test of difference = 0 (vs not =): T-Value = 2.29 P-Value = 0.024 DF = 98
Both use Pooled StDev = 6.3572

INSTRUCTIONS

To produce a confidence interval estimate, follow the instructions for the test, but specify **not equal** for the **Alternative**. Minitab will conduct a two-tail test and produce the confidence interval estimate.

INTERPRET

We estimate that the return on directly purchased mutual funds is on average between .38 and 5.43 percentage points larger than broker-purchased mutual funds.

EXAMPLE 13.2†

DATA
Xm13-02

Effect of New CEO in Family-Run Businesses

What happens to the family-run business when the boss's son or daughter takes over? Does the business do better after the change if the new boss is the offspring of the owner or does the business do better when an outsider is made chief executive officer (CEO)? In pursuit of an answer, researchers randomly selected 140 firms between 1994 and 2002, 30% of which passed ownership to an offspring and 70% of which appointed an outsider as CEO. For each company, the researchers calculated the operating income as a proportion of assets in the year before and the year after the new CEO took over. The change (operating income after – operating income before) in this variable

†*Source:* M. Bennedsen and K. Nielsen, Copenhagen Business School and D. Wolfenzon, New York University.

was recorded and is listed next. Do these data allow us to infer that the effect of making an offspring CEO is different from the effect of hiring an outsider as CEO?

Offspring			Outsider						
−1.95	0.91	−3.15	0.69	−1.05	1.58	−2.46	3.33	−1.32	−0.51
0	−2.16	3.27	−0.95	−4.23	−1.98	1.59	3.2	5.93	8.68
0.56	1.22	−0.67	−2.2	−0.16	4.41	−2.03	0.55	−0.45	1.43
1.44	0.67	2.61	2.65	2.77	4.62	−1.69	−1.4	−3.2	−0.37
1.5	−0.39	1.55	5.39	−0.96	4.5	0.55	2.79	5.08	−0.49
1.41	−1.43	−2.67	4.15	1.01	2.37	0.95	5.62	0.23	−0.08
−0.32	−0.48	−1.91	4.28	0.09	2.44	3.06	−2.69	−2.69	−1.16
−1.7	0.24	1.01	2.97	6.79	1.07	4.83	−2.59	3.76	1.04
−1.66	0.79	−1.62	4.11	1.72	−1.11	5.67	2.45	1.05	1.28
−1.87	−1.19	−5.25	2.66	6.64	0.44	−0.8	3.39	0.53	1.74
−1.38	1.89	0.14	6.31	4.75	1.36	1.37	5.89	3.2	−0.14
0.57	−3.7	2.12	−3.04	2.84	0.88	0.72	−0.71	−3.07	−0.82
3.05	−0.31	2.75	−0.42	−2.1	0.33	4.14	4.22	−4.34	0
2.98	−1.37	0.3	−0.89	2.07	−5.96	3.04	0.46	−1.16	2.68

SOLUTION

IDENTIFY

The objective is to compare two populations, and the data are interval. It follows that the parameter of interest is the difference between two population means $\mu_1 - \mu_2$, where μ_1 is the mean difference for companies where the owner's son or daughter became CEO and μ_2 is the mean difference for companies who appointed an outsider as CEO.

To determine whether to apply the equal or unequal variances t-test, we use the F-test of two variances.

$$H_0:\ \sigma_1^2/\sigma_2^2 = 1$$
$$H_1:\ \sigma_1^2/\sigma_2^2 \neq 1$$

COMPUTE

MANUALLY

From the data, we calculated the following statistics:

$s_1^2 = 3.79$ and $s_2^2 = 8.03$

Test statistic: $F = s_1^2/s_2^2 = 3.79/8.03 = 0.47$

The degrees of freedom are $\nu_1 = n_1 - 1 = 42 - 1 = 41$ and $\nu_2 = n_2 - 1 = 98 - 1 = 97$

Rejection region: $F > F_{\alpha/2,\nu_1,\nu_2} = F_{.025,41,97} \approx F_{.025,40,100} = 1.64$

or

$$F < F_{1-\alpha/2,\nu_1,\nu_2} = F_{.975,41,97} = 1/F_{.025,97,41} \approx 1/F_{.025,100,40} = 1/1.74 = .57$$

Because $F = .47$ is less than .57, we reject the null hypothesis.

EXCEL

	A	B	C
1	F-Test: Two-Sample for Variances		
2			
3		*Offspring*	*Outsider*
4	Mean	–0.10	1.24
5	Variance	3.79	8.03
6	Observations	42	98
7	df	41	97
8	F	0.47	
9	P(F<=f) one-tail	0.0040	
10	F Critical one-tail	0.6314	

The value of the test statistic is $F = .47$, and the p-value $= 2 \times .0040 = .0080$.

MINITAB

Test for Equal Variances: Offspring, Outsider

F-Test (Normal Distribution)
Test statistic = 0.47, p-value = 0.008

INTERPRET

There is enough evidence to infer that the population variances differ. The appropriate technique is the unequal-variances t-test of $\mu_1 - \mu_2$.

Because we want to determine whether there is a *difference* between means, the alternative hypothesis is

$$H_1: \quad (\mu_1 - \mu_2) \neq 0$$

and the null hypothesis is

$$H_0: \quad (\mu_1 - \mu_2) = 0$$

COMPUTE

MANUALLY

From the data, we calculated the following statistics:

$$\bar{x}_1 = -.10$$
$$\bar{x}_2 = 1.24$$
$$s_1^2 = 3.79$$
$$s_2^2 = 8.03$$

The number of degrees of freedom of the test statistic is

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}}$$

$$= \frac{(3.79/42 + 8.03/98)^2}{\dfrac{(3.79/42)^2}{42 - 1} + \dfrac{(8.03/98)^2}{98 - 1}}$$

$$= 110.69 \text{ rounded to } 111$$

The rejection region is

$$t < -t_{\alpha/2,\nu} = -t_{.025,111} \approx -t_{.025,110} = -1.982 \quad \text{or} \quad t > t_{\alpha/2,\nu} = t_{.025,111} \approx 1.982$$

The value of the test statistic is computed next:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)}}$$

$$= \frac{(-.10 - 1.24) - (0)}{\sqrt{\left(\dfrac{3.79}{42} + \dfrac{8.03}{98}\right)}} = -3.22$$

EXCEL

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		*Offspring*	*Outsider*
4	Mean	–0.10	1.24
5	Variance	3.79	8.03
6	Observations	42	98
7	Hypothesized Mean Difference	0	
8	df	111	
9	t Stat	–3.22	
10	P(T<=t) one-tail	0.0008	
11	t Critical one-tail	1.6587	
12	P(T<=t) two-tail	0.0017	
13	t Critical two-tail	1.9816	

INSTRUCTIONS

Follow the instructions for Example 13.1, except at step 2 click **Data, Data Analysis,** and **t-Test: Two-Sample Assuming Unequal Variances**.

MINITAB

Two-Sample T-Test and CI: Offspring, Outsider

Two-sample T for Offspring vs Outsider

	N	Mean	StDev	SE Mean
Offspring	42	–0.10	1.95	0.30
Outsider	98	1.24	2.83	0.29

Difference = mu (Offspring) – mu (Outsider)
Estimate for difference: –1.336
95% CI for difference: (–2.158, –0.514)
T-Test of difference = 0 (vs not =): T-Value = –3.22 P-Value = 0.002 DF = 110

INSTRUCTIONS

Follow the instructions for Example 13.1 except at step 3 do not click **Assume equal variances**.

INTERPRET

The t-statistic is -3.22, and its p-value is .0017. Accordingly, we conclude there is sufficient evidence to infer that the mean changes in operating income differ.

Estimating $\mu_1 - \mu_2$: Unequal-Variances

We can also draw inferences about the difference between the two population means by calculating the confidence interval estimator. We use the unequal-variances confidence interval estimator of $\mu_1 - \mu_2$ and a 95% confidence level.

COMPUTE

MANUALLY

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}$$

$$= (-.10 - 1.24) \pm 1.982\sqrt{\left(\frac{3.79}{42} + \frac{8.03}{98}\right)}$$

$$= -1.34 \pm .82$$

$$\text{LCL} = -2.16 \quad \text{and} \quad \text{UCL} = -.52$$

EXCEL

	A	B	C	D
1	**t-Estimate : Two Means (Unequal Variances)**			
2				
3		*Offspring*	*Outsider*	
4	Mean	–0.10	1.24	
5	Variance	3.79	8.03	
6	Observations	42	98	
7				
8	Degrees of Freedom	110.75		
9	Confidence Level	0.95		
10	Confidence Interval Estimate	–1.34	±	0.82
11	LCL	–2.16		
12	UCL	–0.51		

INSTRUCTIONS

1. Type or import the data into two columns. (Open Xm13-01.)
2. Click **Add-Ins, Data Analysis Plus**, and **t-Estimate: Two Means**.
3. Specify the **Variable 1 Range** (A1:A43) and the **Variable 2 Range** (B1:B99). Click **Independent Samples with Unequal Variances** and the value for α (.05).

MINITAB

Minitab prints the confidence interval estimate as part of the output of the test statistic. However, you must specify the **Alternative** hypothesis as **not equal** to produce a two-sided interval.

INTERPRET

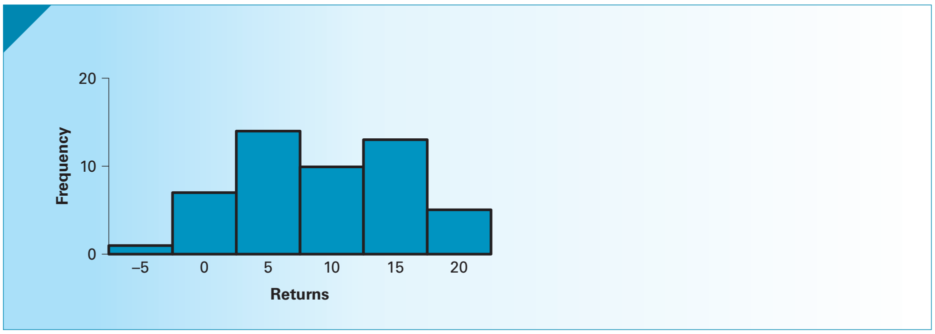
We estimate that the mean change in operating incomes for outsiders exceeds the mean change in the operating income for offspring lies between .51 and 2.16 percentage points.

Checking the Required Condition

Both the equal-variances and unequal-variances techniques require that the populations be normally distributed.* As before, we can check to see whether the requirement is satisfied by drawing the histograms of the data.

To illustrate, we used Excel (Minitab histograms are almost identical) to create the histograms for Example 13.1 (Figures 13.2 and 13.3) and Example 13.2 (Figures 13.4

FIGURE **13.2** Histogram of Rates of Return for Directly Purchased Mutual Funds in Example 13.1



*As we pointed out in Chapter 12 large sample sizes can overcome the effects of extreme nonnormality.

and 13.5). Although the histograms are not perfectly bell shaped, it appears that in both examples the data are at least approximately normal. Because this technique is robust, we can be confident in the validity of the results.

FIGURE **13.3** Histogram of Rates of Return for Broker-Purchased Mutual Funds in Example 13.1

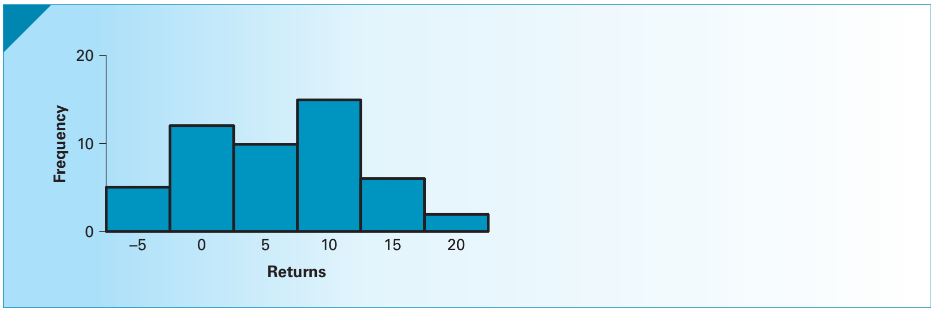
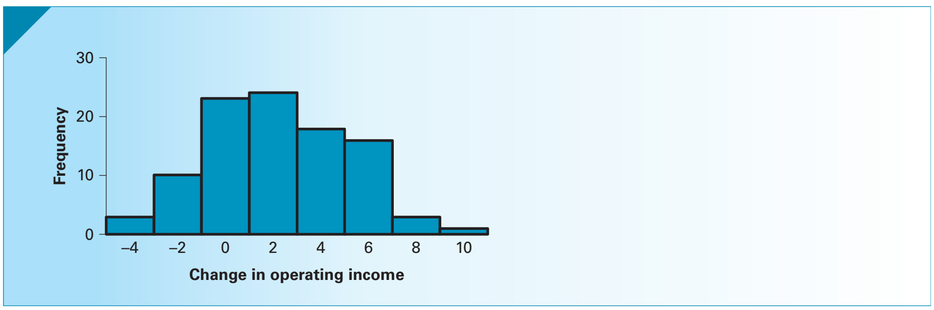


FIGURE **13.4** Histogram of Change in Operating Income for Offspring-Run Businesses in Example 13.2

FIGURE **13.5** Histogram of Change in Operating Income for Outsider-Run Businesses in Example 13.2



Violation of the Required Condition

When the normality requirement is unsatisfied, we can use a nonparametric technique: the Wilcoxon rank sum test (Chapter 19*) to replace the equal-variances test of $\mu_1 - \mu_2$. We have no alternative to the unequal-variances test of $\mu_1 - \mu_2$ when the populations are very nonnormal.

Data Formats

There are two formats for storing the data when drawing inferences about the difference between two means. The first, which you have seen demonstrated in both Examples 13.1 and 13.2, is called *unstacked*, wherein the observations from sample 1 are stored in one column and the observations from sample 2 are stored in a second column. We may also store the data in stacked format. In this format, all the observations are stored in one column. A second column contains the codes, usually 1 and 2, that indicate from which sample the corresponding observation was drawn. Here is an example of unstacked data.

Column 1 (Sample 1)	Column 2 (Sample 2)
12	18
19	23
13	25

Here are the same data in stacked form.

Column 1	Column 2
12	1
19	1
13	1
18	2
23	2
25	2

It should be understood that the data need not be in order. Hence, they could have been stored in this way:

Column 1	Column 2
18	2
25	2
13	1
12	1
23	2
19	1

If there are two populations to compare and only one variable, then it is probably better to record the data in unstacked form. However, it is frequently the case that we want to observe several variables and compare them. For example, suppose that we survey male and female MBAs and ask each to report his or her age, income, and number of years of experience. These data are usually stored in stacked form using the following format.

*Instructors who wish to teach the use of nonparametric techniques for testing the difference between two means when the normality requirement is not satisfied should use Keller's website Appendix Introduction to Nonparametric Techniques and Keller's website Appendix Wilcoxon Rank Sum Test and Wilcoxon Signed Rank Sum Test.

Column 1: Code identifying female (1) and male (2)

Column 2: Age

Column 3: Income

Column 4: Years of experience

To compare ages, we would use columns 1 and 2. Columns 1 and 3 are used to compare incomes, and columns 1 and 4 are used to compare experience levels.

Most statistical software requires one format or the other. Some but not all of Excel's techniques require unstacked data. Some of Minitab's procedures allow either format, whereas others specify only one. Fortunately, both of our software packages allow the statistics practitioner to alter the format. (See Keller's website Appendix Excel and Minitab Instructions for Stacking and Unstacking Data.) We say "fortunately" because this allowed us to store the data in either form on our website. In fact, we've used both forms to allow you to practice your ability to manipulate the data as necessary. You will need this ability to perform statistical techniques in this and other chapters in this book.

Developing an Understanding of Statistical Concepts 1

The formulas in this section are relatively complicated. However, conceptually both test statistics are based on the techniques we introduced in Chapter 11 and repeated in Chapter 12: The value of the test statistic is the difference between the statistic $\bar{x}_1 - \bar{x}_2$ and the hypothesized value of the parameter $\mu_1 - \mu_2$ measured in terms of the standard error.

Developing an Understanding of Statistical Concepts 2

The standard error must be estimated from the data for all inferential procedures introduced here. The method we use to compute the standard error of $\bar{x}_1 - \bar{x}_2$ depends on whether the population variances are equal. When they are equal we calculate and use the pooled variance estimator s_p^2. We are applying an important principle here, and we will do so again in Section 13.5 and in later chapters. The principle can be loosely stated as follows: Where possible, it is advantageous to pool sample data to estimate the standard error. In Example 13.1, we are able to pool because we assume that the two samples were drawn from populations with a common variance. Combining both samples increases the accuracy of the estimate. Thus, s_p^2 is a better estimator of the common variance than either s_1^2 or s_2^2 separately. When the two population variances are unequal, we cannot pool the data and produce a common estimator. We must compute s_1^2 and s_2^2 and use them to estimate σ_1^2 and σ_2^2, respectively.

Here is a summary of how we recognize the techniques presented in this section.

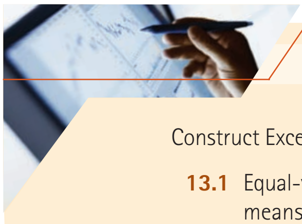
Factors That Identify the Equal-Variances *t*-Test and Estimator of $\mu_1 - \mu_2$

1. **Problem objective**: Compare two populations
2. **Data type:** Interval
3. **Descriptive measurement**: Central location
4. **Experimental design**: Independent samples
5. **Population variances**: Equal

Factors That Identify the Unequal-Variances *t*-Test and Estimator of $\mu_1 - \mu_2$

1. **Problem objective**: Compare two populations
2. **Data type**: Interval
3. **Descriptive measurement**: Central location
4. **Experimental design**: Independent samples
5. **Population variances**: Unequal

EXERCISES



DO-IT-YOURSELF **EXCEL**

Construct Excel spreadsheets for each of the following:

13.1 Equal-variance t-test of $\mu_1 - \mu_2$. Inputs: Sample means, sample standard deviations, sample sizes, hypothesized difference between means. Outputs: Test statistic, critical values, and one- and two-tail p-values. Tools: **TINV, TDIST**

13.2 Equal-variance t-estimator of $\mu_1 - \mu_2$. Inputs: Sample means, sample standard deviations, sample sizes, and confidence level. Outputs: Upper and lower confidence limits. Tools: **TINV**

13.3 Unequal-variance t-test of $\mu_1 - \mu_2$. Inputs: Sample means, sample standard deviations, sample sizes, hypothesized difference between means. Outputs: Test statistic, critical values, and one- and two-tail p-values. Tools: **TINV, TDIST**

13.4 Unequal-variance t-estimator of $\mu_1 - \mu_2$. Inputs: Sample means, sample standard deviations, sample sizes, and confidence level. Outputs: Upper and lower confidence limits. Tools: **TINV**

Developing an Understanding of Statistical Concepts

Exercises 13.5 to 13.10 are "what-if" analyses designed to determine what happens to the test statistics and interval estimates when elements of the statistical inference change. These problems can be solved manually, using the Excel spreadsheets you created or Minitab.

13.5 In random samples of 25 from each of two normal populations, we found the following statistics:

$\bar{x}_1 = 524 \qquad s_1 = 129$

$\bar{x}_2 = 469 \qquad s_2 = 141$

a. Estimate the difference between the two population means with 95% confidence.
b. Repeat part (a) increasing the standard deviations to $s_1 = 255$ and $s_2 = 260$.
c. Describe what happens when the sample standard deviations get larger.
d. Repeat part (a) with samples of size 100.
e. Discuss the effects of increasing the sample size.

13.6 In random samples of 12 from each of two normal populations, we found the following statistics:

$\bar{x}_1 = 74 \qquad s_1 = 18$

$\bar{x}_2 = 71 \qquad s_2 = 16$

a. Test with $\alpha = .05$ to determine whether we can infer that the population means differ.
b. Repeat part (a) increasing the standard deviations to $s_1 = 210$ and $s_2 = 198$.
c. Describe what happens when the sample standard deviations get larger.
d. Repeat part (a) with samples of size 150.
e. Discuss the effects of increasing the sample size.
f. Repeat part (a) changing the mean of sample 1 to $\bar{x}_1 = 76$.
g. Discuss the effect of increasing $\bar{x}_1$.

13.7 Random sampling from two normal populations produced the following results:

$\bar{x}_1 = 63 \qquad s_1 = 18 \qquad n_1 = 50$

$\bar{x}_2 = 60 \qquad s_2 = 7 \qquad n_2 = 45$

a. Estimate with 90% confidence the difference between the two population means.
b. Repeat part (a) changing the sample standard deviations to 41 and 15, respectively.
c. What happens when the sample standard deviations increase?
d. Repeat part (a) doubling the sample sizes.
e. Describe the effects of increasing the sample sizes.

13.8 Random sampling from two normal populations produced the following results:

$\bar{x}_1 = 412$	$s_1 = 128$	$n_1 = 150$
$\bar{x}_2 = 405$	$s_2 = 54$	$n_2 = 150$

a. Can we infer at the 5% significance level that μ_1 is greater than μ_2?
b. Repeat part (a) decreasing the standard deviations to $s_1 = 31$ and $s_2 = 16$.
c. Describe what happens when the sample standard deviations get smaller.
d. Repeat part (a) with samples of size 20.
e. Discuss the effects of decreasing the sample size.
f. Repeat part (a) changing the mean of sample 1 to $\bar{x}_1 = 409$
g. Discuss the effect of decreasing $\bar{x}_1$.

13.9 For each of the following, determine the number of degrees of freedom assuming equal population variances and unequal population variances.
a. $n_1 = 15, n_2 = 15, s_1^2 = 25, s_2^2 = 15$
b. $n_1 = 10, n_2 = 16, s_1^2 = 100, s_2^2 = 15$
c. $n_1 = 50, n_2 = 50, s_1^2 = 8, s_2^2 = 14$
d. $n_1 = 60, n_2 = 45, s_1^2 = 75, s_2^2 = 10$

13.10 Refer to Exercise 13.9.
a. Confirm that in each case the number of degrees of freedom for the equal-variances test statistic and confidence interval estimator is larger than that for the unequal-variances test statistic and confidence interval estimator.
b. Try various combinations of sample sizes and sample variances to illustrate that the number of degrees of freedom for the equal-variances test statistic and confidence interval estimator is larger than that for the unequal-variances test statistic and confidence interval estimator.

Applications

13.11 Xr13-11 Every month a clothing store conducts an inventory and calculates losses from theft. The store would like to reduce these losses and is considering two methods. The first is to hire a security guard, and the second is to install cameras. To help decide which method to choose, the manager hired a security guard for 6 months. During the next 6-month period, the store installed cameras. The monthly losses were recorded and are listed here. The manager decided that because the cameras were cheaper than the guard, he would install the cameras unless there was enough evidence to infer that the guard was better. What should the manager do?

Security guard	355	284	401	398	477	254
Cameras	486	303	270	386	411	435

13.12 Xr13-12 A men's softball league is experimenting with a yellow baseball that is easier to see during night games. One way to judge the effectiveness is to count the number of errors. In a preliminary experiment, the yellow baseball was used in 10 games and the traditional white baseball was used in another 10 games. The number of errors in each game was recorded and is listed here. Can we infer that there are fewer errors on average when the yellow ball is used?

Yellow	5	2	6	7	2	5	3	8	4	9
White	7	6	8	5	9	11	8	3	6	10

13.13 Xr13-13 A number of restaurants feature a device that allows credit card users to swipe their cards at the table. It allows the user to specify a percentage or a dollar amount to leave as a tip. In an experiment to see how it works, a random sample of credit card users was drawn. Some paid the usual way, and some used the new device. The percent left as a tip was recorded and listed below. Can we infer that users of the device leave larger tips?

Usual	10.3	15.2	13.0	9.9	12.1	13.4	12.2	14.9	13.2	12.0	
Device	13.6	15.7	12.9	13.2	12.9	13.4	12.1	13.9	15.7	15.4	17.4

13.14 Xr13-14 Who spends more on their vacations, golfers or skiers? To help answer this question, a travel agency surveyed 15 customers who regularly take their spouses on either a skiing or a golfing vacation. The amounts spent on vacations last year are shown here. Can we infer that golfers and skiers differ in their vacation expenses?

Golfer	2,450	3,860	4,528	1,944	3,166	3,275
	4,490	3,685	2,950			
Skier	3,805	3,725	2,990	4,357	5,550	4,130

13.15 Xr13-15 A growing concern among fans and owners is the amount of time to complete a major league baseball game. To assess the extent of the problem, a statistician recorded the amount of time (in minutes) to complete a random sample of games 5 years ago and this year. Can we conclude that games take longer to complete this year than 5 years ago?

5 Years Ago

169	160	174	161	187	172	177	187	153	169	161	194

This Year

153	182	162	190	163	189	171	197	159	180	197	178

13.16 Xr13-16 How do drivers react to sudden large increases in the price of gasoline? To help answer the question, a statistician recorded the speeds of cars as they passed a large service station. He recorded the speeds (mph) in the same location after the service station sign showed that the price of gasoline had risen by 15 cents. Can we conclude that the speeds differ?

Speeds Before Price Increase

43	36	31	30	28	36	27	36	35	30	32	36

Speeds After Price Increase

32	33	36	31	32	29	28	39	26	30	32	30

Exercises 13.17–13.44 require the use of a computer and software. Use a 5% significance level unless specified otherwise. The answers to Exercises 13.17–13.37 may be calculated manually using the sample statistics listed in Appendix A.

13.17 Xr13-17 The president of Tastee Inc., a baby-food producer, claims that her company's product is superior to that of her leading competitor because babies gain weight faster with her product. (This is a good thing for babies.) To test this claim, a survey was undertaken. Mothers of newborn babies were asked which baby food they intended to feed their babies. Those who responded Tastee or the leading competitor were asked to keep track of their babies' weight gains over the next 2 months. There were 15 mothers who indicated that they would feed their babies Tastee and 25 who responded that they would feed their babies the product of the leading competitor. Each baby's weight gain (in ounces) was recorded.

a. Can we conclude, using weight gain as our criterion, that Tastee baby food is indeed superior?
b. Estimate with 95% confidence the difference between the mean weight gains of the two products.
c. Check to ensure that the required condition(s) is satisfied.

13.18 Xr13-18 Is eating oat bran an effective way to reduce cholesterol? Early studies indicated that eating oat bran daily reduces cholesterol levels by 5% to 10%. Reports of this study resulted in the introduction of many new breakfast cereals with various percentages of oat bran as an ingredient. However, an experiment performed by medical researchers in Boston cast doubt on the effectiveness of oat bran. In that study, 120 volunteers ate oat bran for breakfast, and another 120 volunteers ate another grain cereal for breakfast. At the end of 6 weeks, the percentage of cholesterol reduction was computed for both groups. Can we infer that oat bran is different from other cereals in terms of cholesterol reduction?

13.19 Xr13-19* In assessing the value of radio advertisements, sponsors consider not only the total number of listeners but also their ages. The 18-to-34 age group is considered to spend the most money. To examine the issue, the manager of an FM station commissioned a survey. One objective was to measure the difference in listening habits between the 18-to-34 and 35-to-50 age groups. The survey asked 250 people in each age category how much time they spent listening to FM radio per day. The results (in minutes) were recorded and stored in stacked format (column 1 = Age group and column 2 = Listening times).

a. Can we conclude that a difference exists between the two groups?
b. Estimate with 95% confidence the difference in mean time listening to FM radio between the two age groups.
c. Are the required conditions satisfied for the techniques you used in parts (a) and (b)?

13.20 Xr13-20 The cruise ship business is rapidly increasing. Although cruises have long been associated with seniors, it now appears that younger people are choosing a cruise as their vacations. To determine whether this is true, an executive for a cruise line sampled passengers 2 years ago and this year and determined their ages.

a. Do these data allow the executive to infer that cruise ships are attracting younger customers?
b. Estimate with 99% confidence the difference in ages between this year and 2 years ago.

13.21 Xr13-21* Automobile insurance companies take many factors into consideration when setting rates. These factors include age, marital status, and miles driven per year. To determine the effect of gender, a random sample of young (under 25, with at least 2 years of driving experience) male and female drivers was surveyed. Each was asked how many miles he or she had driven in the past year. The distances (in thousands of miles) are stored in stacked format (column 1 = driving distances and column 2 identifies the gender where 1 = male and code 2 = female).

a. Can we conclude that male and female drivers differ in the numbers of miles driven per year?
b. Estimate with 95% confidence the difference in mean distance driven by male and female drivers.
c. Check to ensure that the required condition(s) of the techniques used in parts (a) and (b) is satisfied.

13.22 Xr13-22 The president of a company that manufactures automobile air conditioners is considering switching his supplier of condensers. Supplier A, the current producer of condensers for the manufacturer, prices its product 5% higher than supplier B. Because the president wants to maintain his company's reputation for quality, he wants to be sure that supplier B's condensers last at least as long as supplier A's. After a careful analysis, the president decided to retain supplier A if there is sufficient statistical evidence that supplier A's condensers last longer on average than supplier B's. In an experiment, 30 midsize cars were equipped with air conditioners using type A condensers while another 30 midsize cars were equipped with type B condensers. The number of miles (in thousands) driven by each car before the condenser broke down was recorded. Should the president retain supplier A?

13.23 Xr13-23 An important function of a firm's human resources manager is to track worker turnover. As a general rule, companies prefer to retain workers. New workers frequently need to be trained, and it often takes time for new workers to learn how to perform their jobs. To investigate nationwide results, a human resources manager organized a survey wherein a random sample of men and women was asked how long they had worked for their current employers. Can we infer that men and women have different job tenures? (Adapted from the *Statistical Abstract of the United States, 2000*, Table 664).

13.24 Xr13-24 A statistics professor is about to select a statistical software package for her course. One of the most important features, according to the professor, is the ease with which students learn to use the software. She has narrowed the selection to two possibilities: software A, a menu-driven statistical package with some high-powered techniques, and software B, a spreadsheet that has the capability of performing most techniques. To help make her decision, she asks 40 statistics students selected at random to choose one of the two packages. She gives each student a statistics problem to solve by computer and the appropriate manual. The amount of time (in minutes) each student needed to complete the assignment was recorded.

a. Can the professor conclude from these data that the two software packages differ in the amount of time needed to learn how to use them? (Use a 1% significance level.)
b. Estimate with 95% confidence the difference in the mean amount of time needed to learn to use the two packages.
c. What are the required conditions for the techniques used in parts (a) and (b)?
d. Check to see whether the required conditions are satisfied.

13.25 Xr13-25 One factor in low productivity is the amount of time wasted by workers. Wasted time includes time spent cleaning up mistakes, waiting for more material and equipment, and performing any other activity not related to production. In a project designed to examine the problem, an operations-management consultant took a survey of 200 workers in companies that were classified as successful (on the basis of their latest annual profits) and another 200 workers from unsuccessful companies. The amount of time (in hours) wasted during a standard 40-hour workweek was recorded for each worker.

a. Do these data provide enough evidence at the 1% significance level to infer that the amount of time wasted in unsuccessful firms exceeds that of successful ones?
b. Estimate with 95% confidence how much more time is wasted in unsuccessful firms than in successful ones.

13.26 Xr13-26 Recent studies seem to indicate that using a cell phone while driving is dangerous. One reason for this is that a driver's reaction time may slow while he or she is talking on the phone. Researchers at Miami (Ohio) University measured the reaction times of a sample of drivers who owned a cell phone. Half the sample was tested while on the phone and the other half was tested while not on the phone. Can we conclude that reaction times are slower for drivers using cell phones?

13.27 Xr13-27 Refer to Exercise 13.26. To determine whether the type of phone usage affects reaction times, another study was launched. A group of drivers was asked to participate in a discussion. Half the group engaged in simple chitchat, and the other half participated in a political discussion. Once again, reaction times were measured. Can we infer that the type of telephone discussion affects reaction times?

13.28 Xr13-28 Most consumers who require someone to perform various professional services undertake research before making their selection. A random sample of people who recently selected a financial planner and a random sample of individuals who chose a stockbroker were asked to report the amount of time they spent researching before deciding. Can we infer that people spend more time researching for a financial planner than they do for a stockbroker? (*Source*: Yankelovich Partners.)

13.29 Xr13-29 Xr13-23 A recent study by researchers at North Carolina State University found thousands of errors in 12 of the most widely used high school science texts. For example, the Statue of Liberty is left-handed; volume is equal to length multiplied by depth (*Time Magazine*, February 12, 2001). The books are so bad that Philip Sadler, director of science education at the Harvard-Smithsonian Center for Astrophysics, decided to conduct a study of their effects. He recorded the physics marks of college students who had used a textbook in high school and the marks of students who did not have a high school textbook. Do these data allow us to infer that students without high school textbooks in science outperform students who used textbooks?

13.30 Xr13-30 Between Wendy's and McDonald's, which fast-food drive-through window is faster? To answer the question, a random sample of service times for each restaurant was measured. Can we infer from these data that there are differences in service times between the two chains? (*Source*: 2000 QSR Drive-Thru Time Study.)

13.31 Xr13-31 Lack of sleep is a serious medical problem. It has been linked to heart attacks and automobile collisions. A Statistics Canada study asked a random sample of Canadian adults to report the amount of sleep they normally get. Can we conclude from the data that men and women differ in the amount of sleep?

13.32 Xr13-32 It is often useful for companies to know who their customers are and how they became customers. In a study of credit card use, random samples were drawn of cardholders who applied for the credit card and credit cardholders who were contacted by telemarketers or by mail. The total purchases made by each last month were recorded. Can we conclude from these data that differences exist on average between the two types of customers?

13.33 Xr13-33 Tire manufacturers are constantly researching ways to produce tires that last longer. New innovations are tested by professional drivers on racetracks. However, any promising inventions are also test-driven by ordinary drivers. The latter tests are closer to what the tire company's customers will actually experience. Suppose that to determine whether a new steel-belted radial tire lasts longer than the company's current model, two new-design tires were installed on the rear wheels of 20 randomly selected cars and two existing-design tires were installed on the rear wheels of another 20 cars. All drivers were told to drive in their usual way until the tires wore out. The number of miles driven by each driver was recorded. Can the company infer that the new tire will last longer on average than the existing tire?

13.34 Xr13-34 It is generally believed that salespeople who are paid on a commission basis outperform salespeople who are paid a fixed salary. Some management consultants argue, however, that in certain industries the fixed-salary salesperson may sell more because the consumer will feel less sales pressure and respond to the salesperson less as an antagonist. In an experiment to study this, a random sample of 180 salespeople from a retail clothing chain was selected. Of these, 90 salespeople were paid a fixed salary, and the remaining 90 were paid a commission on each sale. The total dollar amount of 1 month's sales for each was recorded. Can we conclude that the commission salesperson outperforms the fixed-salary salesperson?

13.35 Xr13-35 Credit scorecards were designed to be used to help financial institutions make decisions about loan applications (see page 63). However, some insurance companies have suggested that credit scores could also be used to determine insurance premiums, particularly car insurance. The Massachusetts Public Interest Research Group has come out against this proposal. To acquire more information, an executive for a car-insurance company gathered data about a random sample of the company's customers. She recorded whether the individual was involved in an accident in the last 3 years and determined the credit score. Can the executive infer that there is a difference in scores between those who did and those who did not have accidents in a 3-year period?

13.36 Xr13-36* Traditionally, wine has been sold in glass bottles with cork stoppers. The stoppers are supposed to keep air out of the bottle because oxygen is the enemy of wine, particularly red wine. Recent research appears to indicate that metal screw caps are more effective in keeping air out of the bottle. However, metal caps are perceived to be inferior and usually associated with cheaper brands of wine. To determine if this perception is wrong, a random sample of 130 people who drink at least one bottle per week on average was asked to participate in an experiment. All were given the same wine in two types of bottles. One group was given a corked bottle, and the other was given a bottle with a metal cap and asked to taste the wine and indicate what they think the retail price of the wine should be. Determine whether there is enough evidence to conclude that bottles of wine with metal caps are perceived to be cheaper.

13.37 Xr13-37 Studies have shown that tired children have trouble learning because neurons become incapable of forming new synaptic connections that are necessary to encode memory. The problem is that the school day starts too early. Awakened at dawn, teenage brains are still releasing melatonin, which makes them sleepy. Several years ago, Edina, Minnesota, changed its high school start from 7:25 A.M. to 8:30 A.M. The SAT scores for a random sample of students taken before the change and a random sample of SAT scores after the change were recorded. Can we infer from the data that SAT scores increased after the change in the school start time?

General Social Survey Exercises

13.38 GSS2008* Study after study indicates that men earn higher incomes than women. To determine the extent of the differential in 2008, estimate with 95% confidence the difference between male and female (SEX: 1 = Male, 2 = Female) annual incomes (INCOME).

13.39 GSS2006* Repeat Exercise 13.38 using data from the 2006 General Social Survey.

13.40 [Ch03:\\CPI-Annual] Use the CPI annual to allow a comparison of the results of Exercises 13.38 and 13.39. Is the income differential decreasing?

13.41 GSS2008* One of the major economic issues in 2010 was the growing size of federal, state, and municipal payrolls. One issue is that people who work for the government earn more than those who work in the private sector. Conduct a test using the 2008 General Social Survey to determine whether we can infer that government employees (WRKGOVT: 1 = Government, 2 = Private) earn more income (INCOME) than other workers?

American National Election Survey Exercises

13.42 ANES2008* The chapter-opening example compares Republicans and Democrats in terms of whether they had graduated from high school. Another way of judging is to measure the number of years of education (EDUC). Conduct a test to determine whether Republicans have more years of education than do Democrats (PARTY: 1 = Democrat and 2 = Republicans)?

13.43 GSS2008* Do the data from the American National Election Survey in 2008 allow us to infer than males have higher incomes than females (INCOME).

13.44 ANES04* Repeat Exercise 13.43 using the ANES data from 2004.

13.2 Observational and Experimental Data

As we've pointed out several times, the ability to properly interpret the results of a statistical technique is a crucial skill for students to develop. This ability is dependent on your understanding of Type I and Type II errors and the fundamental concepts that are part of statistical inference. However, there is another component that must be understood: the difference between **observational data** and **experimental data**. The difference results from the way the data are generated. The following example will demonstrate the difference between the two types.

EXAMPLE 13.3 Dietary Effects of High-Fiber Breakfast Cereals

DATA Xm13-03

Despite some controversy, scientists generally agree that high-fiber cereals reduce the likelihood of various forms of cancer. However, one scientist claims that people who eat high-fiber cereal for breakfast will consume, on average, fewer calories for lunch than people who don't eat high-fiber cereal for breakfast. If this is true, high-fiber cereal manufacturers will be able to claim another advantage of eating their product—potential weight reduction for dieters. As a preliminary test of the claim, 150 people were randomly selected and asked what they regularly eat for breakfast and lunch. Each person was identified as either a consumer or a nonconsumer of high-fiber cereal, and the number of calories consumed at lunch was measured and recorded. These data are listed here. Can the scientist conclude at the 5% significance level that his belief is correct?

Calories Consumed at Lunch by Consumers of High-Fiber Cereal

568	646	607	555	530	714	593	647	650
498	636	529	565	566	639	551	580	629
589	739	637	568	687	693	683	532	651
681	539	617	584	694	556	667	467	
540	596	633	607	566	473	649	622	

Calories Consumed at Lunch by Nonconsumers of High-Fiber Cereal

705	754	740	569	593	637	563	421	514	536
819	741	688	547	723	553	733	812	580	833
706	628	539	710	730	620	664	547	624	644
509	537	725	679	701	679	625	643	566	594
613	748	711	674	672	599	655	693	709	596
582	663	607	505	685	566	466	624	518	750
601	526	816	527	800	484	462	549	554	582
608	541	426	679	663	739	603	726	623	788
787	462	773	830	369	717	646	645	747	
573	719	480	602	596	642	588	794	583	
428	754	632	765	758	663	476	490	573	

SOLUTION

The appropriate technique is the unequal-variances t-test of $\mu_1 - \mu_2$, where μ_1 is the mean of the number of calories for lunch by consumers of high-fiber cereal for breakfast and μ_2 is the mean of the number of calories for lunch by nonconsumers of high-fiber cereal for breakfast. [The F-test of the ratio of two variances (not shown here) yielded $F = .3845$ and p-value $= .0008$.]

The hypotheses are

$$H_0: \quad (\mu_1 - \mu_2) = 0$$
$$H_1: \quad (\mu_1 - \mu_2) < 0$$

The Excel printout is shown next. The manually calculated and Minitab-produced results are identical.

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		*Consumers*	*Nonconsumers*
4	Mean	604.02	633.23
5	Variance	4103	10670
6	Observations	43	107
7	Hypothesized Mean Difference	0	
8	df	123	
9	t Stat	–2.09	
10	P(T<=t) one-tail	0.0193	
11	t Critical one-tail	1.6573	
12	P(T<=t) two-tail	0.0386	
13	t Critical two-tail	1.9794	

INTERPRET

The value of the test statistic is −2.09. The one-tail *p*-value is .0193. We observe that the *p*-value of the test is small (and the test statistic falls into the rejection region). As a result, we conclude that there is sufficient evidence to infer that consumers of high-fiber cereal do eat fewer calories at lunch than do nonconsumers. From this result, we're inclined to believe that eating a high-fiber cereal at breakfast may be a way to reduce weight. However, other interpretations are plausible. For example, people who eat fewer calories are probably more health conscious, and such people are more likely to eat high-fiber cereal as part of a healthy breakfast. In this interpretation, high-fiber cereals do not necessarily lead to fewer calories at lunch. Instead, another factor, general health consciousness, leads to both fewer calories at lunch and high-fiber cereal for breakfast. Notice that the conclusion of the statistical procedure is unchanged. On average, people who eat high-fiber cereal consume fewer calories at lunch. However, because of the way the data were gathered, we have more difficulty interpreting this result.

Suppose that we redo Example 13.3 using the experimental approach. We randomly select 150 people to participate in the experiment. We randomly assign 75 to eat high-fiber cereal for breakfast and the other 75 to eat something else. We then record the number of calories each person consumes at lunch. Ideally, in this experiment both groups will be similar in all other dimensions, including health consciousness. (Larger sample sizes increase the likelihood that the two groups will be similar.) If the statistical result is about the same as in Example 13.3, we may have some valid reason to believe that high-fiber cereal at breakfast leads to a decrease in caloric intake at lunch.

Experimental data are usually more expensive to obtain because of the planning required to set up the experiment; observational data usually require less work to gather. Furthermore, in many situations it is impossible to conduct a controlled experiment. For example, suppose that we want to determine whether an undergraduate degree in engineering better prepares students for an MBA than does an arts degree. In a controlled experiment, we would randomly assign some students to achieve a degree in engineering and other students to obtain an arts degree. We would then make them sign up for an MBA program where we would record their grades. Unfortunately for statistical despots (and fortunately for the rest of us), we live in a democratic society, which makes the coercion necessary to perform this controlled experiment impossible.

To answer our question about the relative performance of engineering and arts students, we have no choice but to obtain our data by observational methods. We would take a random sample of engineering students and arts students who have already entered MBA programs and record their grades. If we find that engineering students do better, we may tend to conclude that an engineering background better prepares students for an MBA program. However, it may be true that better students tend to choose engineering as their undergraduate major and that better students achieve higher grades in all programs, including the MBA program.

Although we've discussed observational and experimental data in the context of the test of the difference between two means, you should be aware that the issue of how the data are obtained is relevant to the interpretation of all the techniques that follow.

EXERCISES

13.45 Refer to Exercise 13.17. If the data are observational, describe another conclusion besides the one that infers that Tastee is better for babies.

13.46 Are the data in Exercise 13.18 observational or experimental? Explain. If the data are observational, describe a method of producing experimental data.

13.47 Refer to Exercise 13.24.
a. Are the data observational or experimental?
b. If the data are observational, describe a method of answering the question with experimental data?
c. If the data are observational produce another explanation for the statistical outcome.

13.48 Suppose that you wish to test to determine whether one method of teaching statistics is better than another.
a. Describe a data-gathering process that produces observational data.
b. Describe a data-gathering process that produces experimental data.

13.49 Put yourself in place of the director of research and development for a pharmaceutical company. When a new drug is developed, it undergoes a number of tests. One test is designed to determine whether the drug is safe and effective. Your company has just developed a drug that is designed to alleviate the symptoms of degenerative diseases such as multiple sclerosis. Design an experiment that tests the new drug.

13.50 You wish to determine whether MBA graduates who majored in finance attract higher starting salaries than MBA graduates who majored in marketing.
a. Describe a data-gathering process that produces observational data.
b. Describe a data-gathering process that produces experimental data.
c. If observational data indicate that finance majors attract higher salaries than do marketing majors, provide two explanations for this result.

13.51 Suppose that you are analyzing one of the hundreds of statistical studies that link smoking with lung cancer. The study analyzed thousands of randomly selected people, some of whom had lung cancer. The statistics indicate that those who have lung cancer smoked on average significantly more than those who did not have lung cancer.
a. Explain how you know that the data are observational.
b. Is there another interpretation of the statistics besides the obvious one that smoking causes lung cancer? If so, what is it? (Students who produce the best answers will be eligible for a job in the public relations department of a tobacco company.)
c. Is it possible to conduct a controlled experiment to produce data that address the question of the relationship between smoking and lung cancer? If so, describe the experiment.

13.3 INFERENCE ABOUT THE DIFFERENCE BETWEEN TWO MEANS: MATCHED PAIRS EXPERIMENT

We continue our presentation of statistical techniques that address the problem of comparing two populations of interval data. In Section 13.1, the parameter of interest was the difference between two population means, where the data were generated from independent samples. In this section, the data are gathered from a matched pairs experiment. To illustrate why matched pairs experiments are needed and how we deal with data produced in this way, consider the following example.

EXAMPLE 13.4 Comparing Salary Offers for Finance and Marketing MBA Majors, Part 1

DATA
Xm13-04

In the last few years, a number of web-based companies that offer job placement services have been created. The manager of one such company wanted to investigate the job offers recent MBAs were obtaining. In particular, she wanted to know whether finance majors were being offered higher salaries than marketing majors. In a preliminary study, she randomly sampled 50 recently graduated MBAs, half of whom majored in finance and half in marketing. From each she obtained the highest salary offer (including benefits). These data are listed here. Can we infer that finance majors obtain higher salary offers than do marketing majors among MBAs?

Highest salary offer made to finance majors

61,228	51,836	20,620	73,356	84,186	79,782	29,523	80,645	76,125
62,531	77,073	86,705	70,286	63,196	64,358	47,915	86,792	75,155
65,948	29,392	96,382	80,644	51,389	61,955	63,573		

Highest salary offer made to marketing majors

73,361	36,956	63,627	71,069	40,203	97,097	49,442	75,188	59,854
79,816	51,943	35,272	60,631	63,567	69,423	68,421	56,276	47,510
58,925	78,704	62,553	81,931	30,867	49,091	48,843		

SOLUTION

IDENTIFY

The objective is to compare two populations of interval data. The parameter is the difference between two means $\mu_1 - \mu_2$ (where μ_1 = mean highest salary offer to finance majors and μ_2 = mean highest salary offer to marketing majors). Because we want to determine whether finance majors are offered higher salaries, the alternative hypothesis will specify that μ_1 is greater than μ_2. The *F*-test for variances was conducted, and the results indicate that there is not enough evidence to infer that the population variances differ. Hence we use the equal-variances test statistic:

$$H_0: \quad (\mu_1 - \mu_2) = 0$$

$$H_1: \quad (\mu_1 - \mu_2) > 0$$

$$\text{Test statistic: } t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

COMPUTE

MANUALLY

From the data, we calculated the following statistics:

$$\bar{x}_1 = 65{,}624$$

$$\bar{x}_2 = 60{,}423$$

$$s_1^2 = 360{,}433{,}294$$

$$s_2^2 = 262{,}228{,}559$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$= \frac{(25 - 1)(360{,}433{,}294) + (25 - 1)(262{,}228{,}559)}{25 + 25 - 2}$$

$$= 311{,}330{,}926$$

The value of the test statistic is computed next:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= \frac{(65{,}624 - 60{,}423) - (0)}{\sqrt{311{,}330{,}926\left(\frac{1}{25} + \frac{1}{25}\right)}}$$

$$= 1.04$$

The number of degrees of freedom of the test statistic is

$$\nu = n_1 + n_2 - 2 = 25 + 25 - 2 = 48$$

The rejection region is

$$t > t_{\alpha,\nu} = t_{.05,48} \approx 1.676$$

EXCEL

	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		*Finance*	*Marketing*
4	Mean	65,624	60,423
5	Variance	360,433,294	262,228,559
6	Observations	25	25
7	Pooled Variance	311,330,926	
8	Hypothesized Mean Difference	0	
9	df	48	
10	t Stat	1.04	
11	P(T<=t) one-tail	0.1513	
12	t Critical one-tail	1.6772	
13	P(T<=t) two-tail	0.3026	
14	t Critical two-tail	2.0106	

MINITAB

Two-Sample T-Test and CI: Finance, Marketing

Two-sample T for Finance vs Marketing

	N	Mean	StDev	SE Mean
Finance	25	65624	18985	3797
Marketing	25	60423	16193	3239

Difference = mu (Finance) – mu (Marketing)
Estimate for difference: 5201.00
95% lower bound for difference: –3169.42
T-Test of difference = 0 (vs >): T-Value = 1.04 P-Value = 0.151 DF = 48
Both use Pooled StDev = 17644.5722

INTERPRET

The value of the test statistic ($t = 1.04$) and its p-value (.1513) indicate that there is very little evidence to support the hypothesis that finance majors receive higher salary offers than marketing majors.

Notice that we have some evidence to support the alternative hypothesis. The difference in sample means is

$$(\bar{x}_1 - \bar{x}_2) = (65{,}624 - 60{,}423) = 5{,}201$$

However, we judge the difference between sample means in relation to the standard error of $\bar{x}_1 - \bar{x}_2$. As we've already calculated,

$$s_p^2 = 311{,}330{,}926$$

and

$$\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 4{,}991$$

Consequently, the value of the test statistic is $t = 5{,}201/4{,}991 = 1.04$, a value that does not allow us to infer that finance majors attract higher salary offers. We can see that although the difference between the sample means was quite large, the variability of the data as measured by s_p^2 was also large, resulting in a small test statistic value.

EXAMPLE 13.5 Comparing Salary Offers for Finance and Marketing MBA Majors, Part 2

DATA Xm13-05

Suppose now that we redo the experiment in the following way. We examine the transcripts of finance and marketing MBA majors. We randomly select a finance and a marketing major whose grade point average (GPA) falls between 3.92 and 4 (based on a maximum of 4). We then randomly select a finance and a marketing major whose GPA is between 3.84 and 3.92. We continue this process until the 25th pair of finance and marketing majors is selected whose GPA fell between 2.0 and 2.08. (The minimum GPA required for graduation is 2.0.) As we did in Example 13.4, we recorded the highest salary offer. These data, together with the GPA group, are listed here. Can we conclude from these data that finance majors draw larger salary offers than do marketing majors?

Group	Finance	Marketing
1	95,171	89,329
2	88,009	92,705
3	98,089	99,205
4	106,322	99,003
5	74,566	74,825
6	87,089	77,038
7	88,664	78,272
8	71,200	59,462
9	69,367	51,555
10	82,618	81,591
11	69,131	68,110
12	58,187	54,970
13	64,718	68,675
14	67,716	54,110
15	49,296	46,467
16	56,625	53,559
17	63,728	46,793
18	55,425	39,984
19	37,898	30,137
20	56,244	61,965
21	51,071	47,438

Group	Finance	Marketing
22	31,235	29,662
23	32,477	33,710
24	35,274	31,989
25	45,835	38,788

SOLUTION

The experiment described in Example 13.4 is one in which the samples are independent. In other words, there is no relationship between the observations in one sample and the observations in the second sample. However, in this example the experiment was designed in such a way that each observation in one sample is matched with an observation in the other sample. The matching is conducted by selecting finance and marketing majors with similar GPAs. Thus, it is logical to compare the salary offers for finance and marketing majors in each group. This type of experiment is called **matched pairs**. We now describe how we conduct the test.

For each GPA group, we calculate the matched pair difference between the salary offers for finance and marketing majors.

Group	Finance	Marketing	Difference
1	95,171	89,329	5,842
2	88,009	92,705	−4,696
3	98,089	99,205	−1,116
4	106,322	99,003	7,319
5	74,566	74,825	−259
6	87,089	77,038	10,051
7	88,664	78,272	10,392
8	71,200	59,462	11,738
9	69,367	51,555	17,812
10	82,618	81,591	1,027
11	69,131	68,110	1,021
12	58,187	54,970	3,217
13	64,718	68,675	−3,957
14	67,716	54,110	13,606
15	49,296	46,467	2,829
16	56,625	53,559	3,066
17	63,728	46,793	16,935
18	55,425	39,984	15,441
19	37,898	30,137	7,761
20	56,244	61,965	−5,721
21	51,071	47,438	3,633
22	31,235	29,662	1,573
23	32,477	33,710	−1,233
24	35,274	31,989	3,285
25	45,835	38,788	7,047

In this experimental design, the parameter of interest is the **mean of the population of differences**, which we label μ_D. Note that μ_D does in fact equal $\mu_1 - \mu_2$, but we test μ_D because of the way the experiment was designed. Hence, the hypotheses to be tested are

$$H_0: \quad \mu_D = 0$$
$$H_1: \quad \mu_D > 0$$

We have already presented inferential techniques about a population mean. Recall that in Chapter 12 we introduced the t-test of μ. Thus, to test hypotheses about μ_D, we use the following test statistic.

Test Statistic for μ_D

$$t = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n_D}}$$

which is Student t distributed with $\nu = n_D - 1$ degrees of freedom, provided that the differences are normally distributed.

Aside from the subscript D, this test statistic is identical to the one presented in Chapter 12. We conduct the test in the usual way.

COMPUTE

MANUALLY

Using the differences computed above, we find the following statistics:

$$\bar{x}_D = 5{,}065$$

$$s_D = 6{,}647$$

from which we calculate the value of the test statistic:

$$t = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n_D}} = \frac{5{,}065 - 0}{6{,}647/\sqrt{25}} = 3.81$$

The rejection region is

$$t > t_{\alpha,\nu} = t_{.05,24} = 1.711$$

EXCEL

	A	B	C
1	t-Test: Paired Two Sample for Means		
2			
3		*Finance*	*Marketing*
4	Mean	65,438	60,374
5	Variance	444,981,810	469,441,785
6	Observations	25	25
7	Pearson Correlation	0.9520	
8	Hypothesized Mean Difference	0	
9	df	24	
10	t Stat	3.81	
11	P(T<=t) one-tail	0.0004	
12	t Critical one-tail	1.7109	
13	P(T<=t) two-tail	0.0009	
14	t Critical two-tail	2.0639	

Excel prints the sample means, variances, and sample sizes for each sample (as well as the coefficient of correlation), which implies that the procedure uses these statistics. It doesn't. The technique is based on computing the paired differences from which the mean, variance, and sample size are determined. Excel should have printed these statistics.

INSTRUCTIONS

1. Type or import the data into two columns*. (Open Xm13-05.)
2. Click **Data, Data Analysis,** and **t-Test: Paired Two-Sample for Means**.

*If one or both columns contain a blank (representing missing data) the row must be deleted.

3. Specify the **Variable 1 Range** (B1:B26) and the **Variable 2 Range** (C1:C26)**.** Type the value of **the Hypothesized Mean Difference** (0) and specify a value for α (.05).

Warning: If there are blank spaces (representing missing data) in any of the rows in either **Variable 1** or **Variable 2 Range**, Excel will produce the wrong answer. You must delete all rows that contain one or two blanks**.** See Keller's website appendix Deleting blank rows in Excel.

MINITAB

Paired T-Test and CI: Finance, Marketing

Paired T for Finance - Marketing

	N	Mean	StDev	SE Mean
Finance	25	65438.2	21094.6	4218.9
Marketing	25	60373.7	21666.6	4333.3
Difference	25	5064.52	6646.90	1329.38

95% lower bound for mean difference: 2790.11
T-Test of mean difference = 0 (vs > 0): T-Value = 3.81 P-Value = 0.000

INSTRUCTIONS

1. Type or import the data into two columns. (Open Xm13-05.)
2. **Click Stat, Basic Statistics,** and **Paired t . . . .**
3. Select the variable names of the **First sample** (Finance) and **Second sample** (Marketing). Click **Options . . . .**
4. In the **Test Mean** box, type the hypothesized mean of the paired difference (0), and specify the **Alternative** (greater than)**.**

INTERPRET

The value of the test statistic is $t = 3.81$ with a p-value of .0004. There is now overwhelming evidence to infer that finance majors obtain higher salary offers than marketing majors. By redoing the experiment as matched pairs, we were able to extract this information from the data.

Estimating the Mean Difference

We derive the confidence interval estimator of μ_D using the usual form for the confidence interval.

Confidence Interval Estimator of μ_D

$$\bar{x}_D \pm t_{\alpha/2}\frac{s_D}{\sqrt{n_D}}$$

EXAMPLE 13.6 Comparing Salary Offers for Finance and Marketing MBA Majors, Part 3

DATA
Xm13-05

Compute the 95% confidence interval estimate of the mean difference in salary offers between finance and marketing majors in Example 13.5.

SOLUTION

COMPUTE

MANUALLY

The 95% confidence interval estimate of the mean difference is

$$\bar{x}_D \pm t_{\alpha/2}\frac{s_D}{\sqrt{n_D}} = 5{,}065 \pm 2.064\frac{6{,}647}{\sqrt{25}} = 5{,}065 \pm 2{,}744$$

$$\text{LCL} = 2{,}321 \quad \text{and} \quad \text{UCL} = 7{,}809$$

EXCEL

	A	B	C	D
1	t-Estimate : Two Means (Matched Pairs)			
2				
3		*Difference*		
4	Mean	5065		
5	Variance	44181217		
6	Observations	25		
7	Degrees of Freedom	24		
8	Confidence Level	0.95		
9	Confidence Interval Estimate	5065	±	2744
10	LCL	2321		
11	UCL	7808		

INSTRUCTIONS

1. Type or import the data into two columns*. (Open Xm13-05.)
2. Click **Add-Ins, Data Analysis Plus**, and **t-Estimate: Two Means**.
3. Specify the **Variable 1 Range** (B1:B51) and the **Variable 2 Range** (C1:C51). Click **Matched Pairs** and the value for α (.05).

MINITAB

Paired T-Test and CI: Finance, Marketing

Paired T for Finance - Marketing

	N	Mean	StDev	SE Mean
Finance	25	65438.2	21094.6	4218.9
Marketing	25	60373.7	21666.6	4333.3
Difference	25	5064.52	6646.90	1329.38

95% CI for mean difference: (2320.82, 7808.22)
T-Test of mean difference = 0 (vs not = 0): T-Value = 3.81 P-Value = 0.001

*If one or both columns contain a blank (representing missing data) the row must be deleted.

INSTRUCTIONS

Follow the instructions to test the paired difference. However, you must specify **not equal** for the **Alternative** hypothesis to produce the two-sided confidence interval estimate of the mean difference.

INTERPRET

We estimate that the mean salary offer to finance majors exceeds the mean salary offer to marketing majors by an amount that lies between \$2,321 and \$7,808 (using the computer output).

Independent Samples or Matched Pairs: Which Experimental Design Is Better?

Examples 13.4 and 13.5 demonstrated that the experimental design is an important factor in statistical inference. However, these two examples raise several questions about experimental designs.

1. Why does the matched pairs experiment result in concluding that finance majors receive higher salary offers than do marketing majors, whereas the independent samples experiment could not?
2. Should we always use the matched pairs experiment? In particular, are there disadvantages to its use?
3. How do we recognize when a matched pairs experiment has been performed?

Here are our answers.

1. The matched pairs experiment worked in Example 13.5 by reducing the variation in the data. To understand this point, examine the statistics from both examples. In Example 13.4, we found $\bar{x}_1 - \bar{x}_2 = 5{,}201$. In Example 13.5, we computed $\bar{x}_D = 5{,}065$. Thus, the numerators of the two test statistics were quite similar. However, the test statistic in Example 13.5 was much larger than the test statistic in Example 13.4 because of the standard errors. In Example 13.4, we calculated

$$s_p^2 = 311{,}330{,}926 \quad \text{and} \quad \sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 4{,}991$$

Example 13.5 produced

$$s_D = 6{,}647 \quad \text{and} \quad \frac{s_D}{\sqrt{n_D}} = 1{,}329$$

As you can see, the difference in the test statistics was caused not by the numerator, but by the denominator. This raises another question: Why was the variation in the data of Example 13.4 so much greater than the variation in the data of Example 13.5? If you examine the data and statistics from Example 13.4, you will find that there was a great deal of variation *between* the salary offers in each sample. In other words, some MBA graduates received high salary offers and others relatively low ones. This high level of variation, as expressed by s_p^2, made the difference between the sample means appear to be small. As a result, we could not conclude that finance majors attract higher salary offers.

Looking at the data from Example 13.5, we see that there is very little variation between the observations of the paired differences. The variation caused by different GPAs has been decreased markedly. The smaller variation causes the value of the test statistic to be larger. Consequently, we conclude that finance majors obtain higher salary offers.

2. Will the matched pairs experiment always produce a larger test statistic than the independent samples experiment? The answer is, not necessarily. Suppose that in our example we found that companies did not consider grade point averages when making decisions about how much to offer the MBA graduates. In such circumstances, the matched pairs experiment would result in no significant decrease in variation when compared to independent samples. It is possible that the matched pairs experiment may be less likely to reject the null hypothesis than the independent samples experiment. The reason can be seen by calculating the degrees of freedom. In Example 13.4, the number of degrees of freedom was 48, whereas in Example 13.5, it was 24. Even though we had the same number of observations (25 in each sample), the matched pairs experiment had half the number of degrees of freedom as the equivalent independent samples experiment. For exactly the same value of the test statistic, a smaller number of degrees of freedom in a Student *t* distributed test statistic yields a larger *p*-value. What this means is that if there is little reduction in variation to be achieved by the matched pairs experiment, the statistics practitioner should choose instead to conduct the experiment with independent samples.

3. As you've seen, in this book we deal with questions arising from experiments that have already been conducted. Consequently, one of your tasks is to determine the appropriate test statistic. In the case of comparing two populations of interval data, you must decide whether the samples are independent (in which case the parameter is $\mu_1 - \mu_2$) or matched pairs (in which case the parameter is μ_D) to select the correct test statistic. To help you do so, we suggest you ask and answer the following question: Does some natural relationship exist between each pair of observations that provides a logical reason to compare the first observation of sample 1 with the first observation of sample 2, the second observation of sample 1 with the second observation of sample 2, and so on? If so, the experiment was conducted by matched pairs. If not, it was conducted using independent samples.

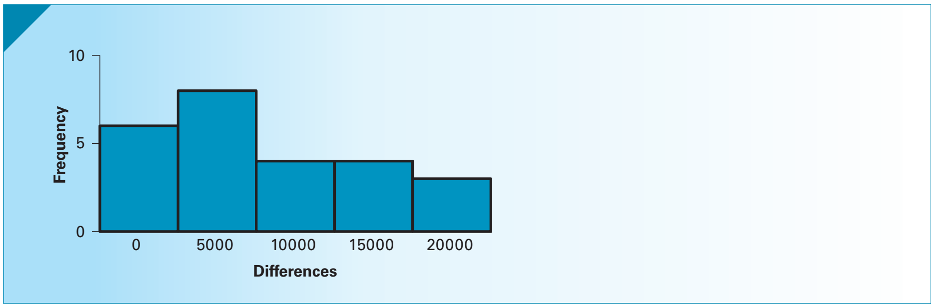
Observational and Experimental Data

The points we made in Section 13.2 are also valid in this section: We can design a matched pairs experiment where the data are gathered using a controlled experiment or by observation. The data in Examples 13.4 and 13.5 are observational. As a consequence, when the statistical result provided evidence that finance majors attracted higher salary offers, it did not necessarily mean that students educated in finance are more attractive to prospective employers. It may be, for example, that better students major in finance and better students achieve higher starting salaries.

Checking the Required Condition

The validity of the results of the *t*-test and estimator of μ_D depends on the normality of the differences (or large enough sample sizes). The histogram of the differences (Figure 13.6) is positively skewed but not enough so that the normality requirement is violated.

FIGURE 13.6 Histogram of Differences in Example 13.5



Violation of Required Condition

If the differences are very nonnormal, we cannot use the t-test of μ_D. We can, however, employ a nonparametric technique—the Wilcoxon signed rank sum test for matched pairs, which we present in Chapter 19.*

Developing an Understanding of Statistical Concepts 1

Two of the most important principles in statistics were applied in this section. The first is the concept of analyzing sources of variation. In Examples 13.4 and 13.5, we showed that by reducing the variation between salary offers in each sample we were able to detect a real difference between the two majors. This was an application of the more general procedure of analyzing data and attributing some fraction of the variation to several sources. In Example 13.5, the two sources of variation were the GPA and the MBA major. However, we were not interested in the variation between graduates with differing GPAs. Instead, we only wanted to eliminate that source of variation, making it easier to determine whether finance majors draw larger salary offers.

In Chapter 14, we will introduce a technique called the *analysis of variance* that does what its name suggests: It analyzes sources of variation in an attempt to detect real differences. In most applications of this procedure, we will be interested in each source of variation and not simply in reducing one source. We refer to the process as *explaining the variation*. The concept of explained variation is also applied in Chapters 16–18, where we introduce regression analysis.

Developing an Understanding of Statistical Concepts 2

The second principle demonstrated in this section is that statistics practitioners can design data-gathering procedures in such a way that they can analyze sources of variation. Before conducting the experiment in Example 13.5, the statistics practitioner suspected that there were large differences between graduates with different GPAs. Consequently, the experiment was organized so that the effects of those differences were mostly eliminated. It is also possible to design experiments that allow for easy detection

*Instructors who wish to teach the use of nonparametric techniques for testing the mean difference when the normality requirement is not satisfied should use Keller's website Appendix Introduction to Nonparametric Techniques and Keller's website Appendix Wilcoxon Rank Sum Test and Wilcoxon Signed Rank Sum Test.

of real differences and minimize the costs of data gathering. Unfortunately, we will not present this topic. However, you should understand that the entire subject of the design of experiments is an important one, because statistics practitioners often need to be able to analyze data to detect differences, and the cost is almost always a factor.

Here is a summary of how we determine when to use these techniques.

Factors That Identify the *t*-Test and Estimator of μ_D

1. **Problem objective**: Compare two populations
2. **Data type**: Interval
3. **Descriptive measurement**: Central location
4. **Experimental design**: Matched pairs

Exercises

Applications

Conduct all tests of hypotheses at the 5% significance level.

13.52 Xr13-52 Many people use scanners to read documents and store them in a Word (or some other software) file. To help determine which brand of scanner to buy, a student conducts an experiment in which eight documents are scanned by each of the two scanners he is interested in. He records the number of errors made by each. These data are listed here. Can he infer that brand A (the more expensive scanner) is better than brand B?

Document	1	2	3	4	5	6	7	8
Brand A	17	29	18	14	21	25	22	29
Brand B	21	38	15	19	22	30	31	37

13.53 Xr13-53 How effective is an antilock braking system (ABS), which pumps very rapidly rather than lock and thus avoid skids? As a test, a car buyer organized an experiment. He hit the brakes and, using a stopwatch, recorded the number of seconds it took to stop an ABS-equipped car and another identical car without ABS. The speeds when the brakes were applied and the number of seconds each took to stop on dry pavement are listed here. Can we infer that ABS is better?

Speeds	20	25	30	35	40	45	50	55
ABS	3.6	4.1	4.8	5.3	5.9	6.3	6.7	7.0
Non-ABS	3.4	4.0	5.1	5.5	6.4	6.5	6.9	7.3

13.54 Xr13-54 In a preliminary study to determine whether the installation of a camera designed to catch cars that go through red lights affects the number of violators, the number of red-light runners was recorded for each day of the week before and after the camera was installed. These data are listed here. Can we infer that the camera reduces the number of red-light runners?

Day	Sunday	Monday	Tuesday	Wednesday
Before	7	21	27	18
After	8	18	24	19

Day	Thursday	Friday	Saturday
Before	20	24	16
After	16	19	16

13.55 Xr13-55 In an effort to determine whether a new type of fertilizer is more effective than the type currently in use, researchers took 12 two-acre plots of land scattered throughout the county. Each plot was divided into two equal-sized subplots, one of which was treated with the current fertilizer and the other with the new fertilizer. Wheat was planted, and the crop yields were measured.

Plot	1	2	3	4	5	6	7	8	9	10	11	12
Current fertilizer	56	45	68	72	61	69	57	55	60	72	75	66
New fertilizer	60	49	66	73	59	67	61	60	58	75	72	68

a. Can we conclude at the 5% significance level that the new fertilizer is more effective than the current one?
b. Estimate with 95% confidence the difference in mean crop yields between the two fertilizers.
c. What is the required condition(s) for the validity of the results obtained in parts (a) and (b)?

d. Is the required condition(s) satisfied?
e. Are these data experimental or observational? Explain.
f. How should the experiment be conducted if the researchers believed that the land throughout the county was essentially the same?

13.56 Xr13-56 The president of a large company is in the process of deciding whether to adopt a lunchtime exercise program. The purpose of such programs is to improve the health of workers and thus reduce medical expenses. To get more information, he instituted an exercise program for the employees in one office. The president knows that during the winter months medical expenses are relatively high because of the incidence of colds and flu. Consequently, he decides to use a matched pairs design by recording medical expenses for the 12 months before the program and for 12 months after the program. The "before" and "after" expenses (in thousands of dollars) are compared on a month-to-month basis and shown here.
a. Do the data indicate that exercise programs reduce medical expenses? (Test with $\alpha = .05$.)
b. Estimate with 95% confidence the mean savings produced by exercise programs.
c. Was it appropriate to conduct a matched pairs experiment? Explain.

Month	Jan	Feb	Mar	Apr	May	Jun
Before program	68	44	30	58	35	33
After program	59	42	20	62	25	30

Month	Jul	Aug	Sep	Oct	Nov	Dec
Before program	52	69	23	69	48	30
After program	56	62	25	75	40	26

Exercises 13.57–13.72 require the use of a computer and software. Use a 5% significance level unless specified otherwise. The answers to Exercises 13.57 to 13.69 may be calculated manually. See Appendix A for the sample statistics.

13.57 Xr13-57 One measure of the state of the economy is the amount of money homeowners pay on their mortgage each month. To determine the extent of change between this year and 5 years ago, a random sample of 150 homeowners was drawn. The monthly mortgage payments for each homeowner for this year and for 5 years ago were recorded. (The amounts have been adjusted so that we're comparing constant dollars.) Can we infer that mortgage payments have risen over the past 5 years?

13.58 Xr13-58 Do waiters or waitresses earn larger tips? To answer this question, a restaurant consultant undertook a preliminary study. The study involved measuring the percentage of the total bill left as a tip for one randomly selected waiter and one randomly selected waitress in each of 50 restaurants during a 1-week period. What conclusions can be drawn from these data?

13.59 Xr13-59 To determine the effect of advertising in the Yellow Pages, Bell Telephone took a sample of 40 retail stores that did not advertise in the Yellow Pages last year but did so this year. The annual sales (in thousands of dollars) for each store in both years were recorded.
a. Estimate with 90% confidence the improvement in sales between the 2 years.
b. Can we infer that advertising in the Yellow Pages improves sales?
c. Check to ensure that the required condition(s) of the techniques used in parts (a) and (b) is satisfied.
d. Would it be advantageous to perform this experiment with independent samples? Explain why or why not.

13.60 Xr13-60 Because of the high cost of energy, homeowners in northern climates need to find ways to cut their heating costs. A building contractor wanted to investigate the effect on heating costs of increasing the insulation. As an experiment, he located a large subdevelopment built around 1970 with minimal insulation. His plan was to insulate some of the houses and compare the heating costs in the insulated homes with those that remained uninsulated. However, it was clear to him that the size of the house was a critical factor in determining heating costs. Consequently, he found 16 pairs of identical-sized houses ranging from about 1,200 to 2,800 square feet. He insulated one house in each pair (levels of R20 in the walls and R32 in the attic) and left the other house unchanged. The heating cost for the following winter season was recorded for each house.
a. Do these data allow the contractor to infer at the 10% significance level that the heating cost for insulated houses is less than that for the uninsulated houses?
b. Estimate with 95% confidence the mean savings due to insulating the house.
c. What is the required condition for the use of the techniques in parts (a) and (b)?

13.61 Xr13-61 The cost of health care is rising faster than most other items. To learn more about the problem, a survey was undertaken to determine whether differences in health-care expenditures exist between men and women. The survey randomly sampled men and women aged 21, 22, . . . , 65 and determined the total amount spent on health care. Do these data allow us to infer that men and women spend different amounts on health care? (*Source*: Bureau of Labor Statistics, Consumer Expenditure Survey.)

13.62 Xr13-62 The fluctuations in the stock market induce some investors to sell and move their money into more stable investments. To determine the degree to which recent fluctuations affected ownership, a random sample of 170 people who confirmed that they owned some stock was surveyed. The values of the holdings were recorded at the end of last year and at the end of the year before. Can we infer that the value of the stock holdings has decreased?

13.63 Xr13-63 Are Americans more deeply in debt this year compared to last year? To help answer this question, a statistics practitioner randomly sampled Americans this year and last year. The sampling was conducted so that the samples were matched by the age of the head of the household. For each, the ratio of debt payments to household income was recorded. Can we infer that the ratios are higher this year than last?

13.64 Xr13-64 Every April Americans and Canadians fill out their tax return forms. Many turn to tax preparation companies to do this tedious job. The question arises, Are there differences between companies? In an experiment, two of the largest companies were asked to prepare the tax returns of a sample of 55 taxpayers. The amounts of tax payable were recorded. Can we conclude that company 1's service results in higher tax payable?

13.65 Xr13-65 Refer to Exercise 13.33. Suppose now we redo the experiment in the following way. On 20 randomly selected cars, one of each type of tire is installed on the rear wheels and, as before, the cars are driven until the tires wear out. The number of miles until wear-out occurred was recorded. Can we conclude from these data that the new tire is superior?

13.66 Refer to Exercises 13.33 and 13.65. Explain why the matched pairs experiment produced significant results whereas the independent samples t-test did not.

13.67 Xr13-67 Refer to Examples 13.4 and 13.5. Suppose that another experiment is conducted. Finance and marketing MBA majors were matched according to their undergraduate GPA. As in the previous examples, the highest starting salary offers were recorded. Can we infer from these data that finance majors attract higher salary offers than marketing majors?

13.68 Discuss why the experiment in Example 13.5 produced a significant test result whereas the one in Exercise 13.67 did not.

13.69 Xr13-69 Refer to Example 13.2.The actual after and before operating incomes were recorded.

a. Test to determine whether there is enough evidence to infer that for companies where an offspring takes the helm there is a decrease in operating income.

b. Is there sufficient evidence to conclude that when an outsider becomes CEO the operating income increases?

General Social Survey Exercises

Warning: Some rows contain blanks representing missing data.

13.70 GSS2008* The general trend over the last century is that each generation is more educated that its predecessor. Has this trend continued? To answer this question, determine whether there is sufficient evidence that Americans are more educated than their fathers (EDUC and PAEDUC).

13.71 GSS2008* Is there sufficient evidence to infer that Americans are more educated than their mothers (EDUC and MAEDUC)?

13.72 GSS2008* If it is true that this generation is more educated than its parents, does it follow that its members have more prestigious occupations? To help answer this question, conduct a statistical procedure to determine whether adults today have more prestigious jobs than their fathers (PRESTG80 and PAPRES80).

American National Election Survey Exercise

Warning: Some rows contain blanks representing missing data.

13.73 ANES2008* Estimate with 95% confidence the average difference between the amount of time spent watching news on television (not including sports) and the amount of time spent reading news in a printed newspaper during a typical day (TIME2 and TIME3).

13.4 Inference about the Ratio of Two Variances

In Sections 13.1 and 13.3, we dealt with statistical inference concerning the difference between two population means. The problem objective in each case was to compare two populations of interval data, and our interest was in comparing measures of central location. This section discusses the statistical technique to use when the problem objective and the data type are the same as in Sections 13.1 and 13.3, but our interest is in comparing variability. Here we will study the ratio of two population variances. We make inferences about the ratio because the sampling distribution is based on ratios rather than differences.

We have already encountered this technique when we used the *F*-test of two variances to determine which *t*-test and estimator of the difference between two means to use. In this section, we apply the technique to other problems where our interest is in comparing the variability in two populations.

In the previous chapter, we presented the procedures used to draw inferences about a single population variance. We pointed out that variance can be used to address problems where we need to judge the consistency of a production process. We also use variance to measure the risk associated with a portfolio of investments. In this section, we compare two variances, enabling us to compare the consistency of two production processes. We can also compare the relative risks of two sets of investments.

We will proceed in a manner that is probably becoming quite familiar.

Parameter

As you will see shortly, we compare two population variances by determining the ratio. Consequently, the parameter is σ_1^2/σ_2^2.

Statistic and Sampling Distribution

We have previously noted that the sample variance (defined in Chapter 4) is an unbiased and consistent estimator of the population variance. Not surprisingly, the estimator of the parameter σ_1^2/σ_2^2 is the ratio of the two sample variances drawn from their respective populations s_1^2/s_2^2.

The sampling distribution of s_1^2/s_2^2 is said to be *F*-distributed provided that we have independently sampled from two normal populations. (The *F*-distribution was introduced in Section 8.4.)

Statisticians have shown that the ratio of two independent chi-squared variables divided by their degrees of freedom is *F*-distributed. The degrees of freedom of the *F*-distribution are identical to the degrees of freedom for the two chi-squared distributions. In Section 12.2, we pointed out that $(n - 1)\, s^2 / \sigma^2$ is chi-squared distributed, provided that the sampled population is normal. If we have independent samples drawn from two normal populations, then both $(n_1 - 1)s_1^2/\sigma_1^2$ and $(n_2 - 1)s_2^2/\sigma_2^2$ are chi-squared distributed. If we divide each by their respective number of degrees of freedom and take the ratio, we produce

$$\frac{\dfrac{(n_1 - 1)s_1^2/\sigma_1^2}{(n_1 - 1)}}{\dfrac{(n_2 - 1)s_2^2/\sigma_2^2}{(n_2 - 1)}}$$

which simplifies to

$$\frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

This statistic is F-distributed with $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$ degrees of freedom. Recall that ν_1 is called the **numerator degrees of freedom** and ν_2 is called the **denominator degrees of freedom**.

Testing and Estimating a Ratio of Two Variances

In this book, our null hypothesis will always specify that the two variances are equal. As a result, the ratio will equal 1. Thus, the null hypothesis will always be expressed as

$$H_0: \quad \sigma_1^2/\sigma_2^2 = 1$$

The alternative hypothesis can state that the ratio σ_1^2/σ_2^2 is either not equal to 1, greater than 1, or less than 1. Technically, the test statistic is

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

However, under the null hypothesis, which states that $\sigma_1^2/\sigma_2^2 = 1$, the test statistic becomes as follows.

Test Statistic for $\boldsymbol{\sigma_1^2/\sigma_2^2}$

The test statistic employed to test that σ_1^2/σ_2^2 is equal to 1 is

$$F = \frac{s_1^2}{s_2^2}$$

which is F-distributed with $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$ degrees of freedom provided that the populations are normal.

With the usual algebraic manipulation, we can derive the confidence interval estimator of the ratio of two population variances.

Confidence Interval Estimator of $\boldsymbol{\sigma_1^2/\sigma_2^2}$

$$\text{LCL} = \left(\frac{s_1^2}{s_2^2}\right)\frac{1}{F_{\alpha/2,\nu_1,\nu_2}}$$

$$\text{UCL} = \left(\frac{s_1^2}{s_2^2}\right)F_{\alpha/2,\nu_2,\nu_1}$$

where $\nu_1 = n_1 - 1$ and $\nu_2 = n_2 - 1$

EXAMPLE 13.7 Testing the Quality of Two-Bottle Filling Machines

DATA
Xm13-07

In Example 12.3, we applied the chi-squared test of a variance to determine whether there was sufficient evidence to conclude that the population variance was less than 1.0. Suppose that the statistics practitioner also collected data from another container-filling machine and recorded the fills of a randomly selected sample. Can we infer at the 5% significance level that the second machine is superior in its consistency?

SOLUTION

IDENTIFY

The problem objective is to compare two populations where the data are interval. Because we want information about the consistency of the two machines, the parameter we wish to test is σ_1^2/σ_2^2, where σ_1^2 is the variance of machine 1 and σ_2^2 is the variance for machine 2. We need to conduct the F-test of σ_1^2/σ_2^2 to determine whether the variance of population 2 is less than that of population 1. Expressed differently, we wish to determine whether there is enough evidence to infer that σ_1^2 is larger than σ_2^2. Hence, the hypotheses we test are

$$H_0: \quad \sigma_1^2/\sigma_2^2 = 1$$

$$H_1: \quad \sigma_1^2/\sigma_2^2 > 1$$

COMPUTE

MANUALLY

The sample variances are $s_1^2 = .6333$ and $s_2^2 = .4528$.

The value of the test statistic is

$$F = \frac{s_1^2}{s_2^2} = \frac{.6333}{.4528} = 1.40$$

The rejection region is

$$F > F_{\alpha,\nu_1,\nu_2} = F_{.05,24,24} = 1.98$$

Because the value of the test statistic is not greater than 1.98, we cannot reject the null hypothesis.

EXCEL

	A	B	C
1	F-Test Two-Sample for Variances		
2			
3		*Machine 1*	*Machine 2*
4	Mean	999.7	999.8
5	Variance	0.6333	0.4528
6	Observations	25	25
7	df	24	24
8	F	1.3988	
9	P(F<=f) one-tail	0.2085	
10	F Critical one-tail	1.9838	

The value of the test statistic is $F = 1.3988$. Excel outputs the one-tail p-value, which is .2085.

(Continued)

INSTRUCTIONS

1. Type or import the data into two columns. (Open Xm13-07.)
2. Click **Data, Data Analysis**, and ***F*-test Two-Sample for Variances**.
3. Specify the **Variable 1 Range** (A1:A26) and the **Variable 2 Range** (B1:B26). Type a value for α (.05).

MINITAB

```
Test for Equal Variances: Machine 1, Machine 2

F-Test (Normal Distribution)
Test statistic = 1.40, p-value = 0.417
```

Note that Minitab conducts a two-tail test only. Thus, the *p*-value = .417/2 = .2085.

INSTRUCTIONS

1. Type or import the data into two columns. (Open Xm13-07.)
2. Click **Stat, Basic Statistics,** and **2 Variances . . . .**
3. In the **Samples in different columns** box, select the **First** (Machine 1) and **Second** (Machine 2) variables.

INTERPRET

There is not enough evidence to infer that the variance of machine 2 is less than the variance of machine 1.

The histograms (not shown) appear to be sufficiently bell shaped to satisfy the normality requirement.

EXAMPLE 13.8 Estimating the Ratio of the Variances in Example 13.7

DATA Xm13-07

Determine the 95% confidence interval estimate of the ratio of the two population variances in Example 13.7.

SOLUTION

COMPUTE

MANUALLY

We find

$$F_{\alpha/2,\nu_1,\nu_2} = F_{.025,24,24} = 2.27$$

Thus,

$$\text{LCL} = \left(\frac{s_1^2}{s_2^2}\right)\frac{1}{F_{\alpha/2,\nu_1,\nu_2}} = \left(\frac{.6333}{.4528}\right)\frac{1}{2.27} = .616$$

$$\text{UCL} = \left(\frac{s_1^2}{s_2^2}\right)F_{\alpha/2,\nu_2,\nu_1} = \left(\frac{.6333}{.4528}\right)2.27 = 3.17$$

We estimate that σ_1^2/σ_2^2 lies between .616 and 3.17.

EXCEL

	A	B	C
1	F-Estimate : Two Variances		
2			
3		*Machine 1*	*Machine 2*
4	Mean	999.7	999.8
5	Variance	0.6333	0.4528
6	Observations	25	25
7	df	24	24
8	LCL	0.6164	
9	UCL	3.1743	

INSTRUCTIONS

1. Type or import the data into two columns. (Open Xm13-07.)
2. Click **Add-ins, Data Analysis Plus,** and **F Estimate 2 Variances**.
3. Specify the **Variable 1 Range** (A1:A26) and the **Variable 2 Range** (B1:B26). Type a value for α (.05).

MINITAB

Minitab does not compute the estimate of the ratio of two variances.

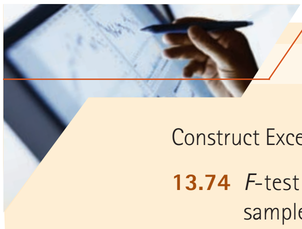
INTERPRET

As we pointed out in Chapter 11, we can often use a confidence interval estimator to test hypotheses. In this example, the interval estimate excludes the value of 1. Consequently, we can draw the same conclusion as we did in Example 13.7.

Factors That Identify the *F*-Test and Estimator of σ_1^2/σ_2^2

1. **Problem objective**: Compare two populations
2. **Data type**: Interval
3. **Descriptive measurement**: Variability

EXERCISES



DO-IT-YOURSELF **EXCEL**

Construct Excel spreadsheets for each of the following:

13.74 *F*-test of σ_1^2/σ_2^2. Inputs: Sample variances, sample sizes, and hypothesized ratio of population variances. Outputs: Test statistic, critical values, and one- and two-tail *p*-values. Tools: **FINV, FDIST**

13.74 *F*-estimator of σ_1^2/σ_2^2. Inputs: Sample variances, sample sizes, and confidence level. Outputs: Upper and lower confidence limits, Tools: **FINV**

Developing an Understanding of Statistical Concepts

Exercises 13.76 and 13.77 are "what-if" analyses designed to determine what happens to the test statistics and interval estimates when elements of the statistical inference change. These problems can be solved manually, using Do-It-Yourself Excel spreadsheets you just created, or Minitab.

13.76 Random samples from two normal populations produced the following statistics:

$$s_1^2 = 350 \quad n_1 = 30 \quad s_2^2 = 700 \quad n_2 = 30$$

a. Can we infer at the 10% significance level that the two population variances differ?
b. Repeat part (a) changing the sample sizes to $n_1 = 15$ and $n_2 = 15$.
c. Describe what happens to the test statistic and the conclusion when the sample sizes decrease.

13.77 Random samples from two normal populations produced the following statistics:

$$s_1^2 = 28 \quad n_1 = 10 \quad s_2^2 = 19 \quad n_2 = 10$$

a. Estimate with 95% confidence the ratio of the two population variances.
b. Repeat part (a) changing the sample sizes to $n_1 = 25$ and $n_2 = 25$.
c. Describe what happens to the width of the confidence interval estimate when the sample sizes increase.

Applications

Use a 5% significance level in all tests unless specified otherwise.

13.78 Xr13-78 The manager of a dairy is in the process of deciding which of two new carton-filling machines to use. The most important attribute is the consistency of the fills. In a preliminary study, she measured the fills in the 1-liter carton and listed them here. Can the manager infer that the two machines differ in their consistency of fills?

Machine 1	.998	.997	1.003	1.000	.999	
	1.000	.998	1.003	1.004	1.000	
Machine 2	1.003	1.004	.997	.996	.999	1.003
	1.000	1.005	1.002	1.004	.996	

13.79 Xr13-79 An operations manager who supervises an assembly line has been experiencing problems with the sequencing of jobs. The problem is that bottlenecks are occurring because of the inconsistency of sequential operations. He decides to conduct an experiment wherein two different methods are used to complete the same task. He measures the times (in seconds). The data are listed here. Can he infer that the second method is more consistent than the first method?

Method 1	8.8	9.6	8.4	9.0	8.3	9.2	9.0	8.7	8.5	9.4
Method 2	9.2	9.4	8.9	9.6	9.7	8.4	8.8	8.9	9.0	9.7

13.80 Xr13-80 A statistics professor hypothesized that not only would the means vary but also so would the variances if the business statistics course was taught in two different ways but had the same final exam. He organized an experiment wherein one section of the course was taught using detailed PowerPoint slides whereas the other required students to read the book and answer questions in class discussions. A sample of the marks was recorded and listed next. Can we infer that the variances of the marks differ between the two sections?

Class 1	64	85	80	64	48	62	75	77	50	81	90
Class 2	73	78	66	69	79	81	74	59	83	79	84

The following exercises require the use of a computer and software. The answers may be calculated manually. See Appendix A for the sample statistics.

13.81 Xr13-81 A new highway has just been completed, and the government must decide on speed limits. There are several possible choices. However, on advice from police who monitor traffic, the objective was to reduce the variation in speeds, which it is thought to contribute to the number of collisions. It has been acknowledged that speed contributes to the severity of collisions. It is decided to conduct an experiment to acquire more information. Signs are posted for 1 week indicating that the speed limit is 70 mph. A random sample of cars' speeds is measured. During the second week, signs are posted indicating that the maximum speed is 70 mph and that the minimum speed is 60 mph. Once again a random sample of speeds is measured. Can we infer that limiting the minimum and maximum speeds reduces the variation in speeds?

13.82 Xr13-82 In Exercise 12.66, we described the problem of whether to change all the lightbulbs at Yankee Stadium or change them one by one as they burn out. There are two brands of bulbs that can be used. Because both the mean and the variance of the lengths of life are important, it was decided to test the two brands. A random sample of both brands was drawn and left on until they burned out. The times were recorded. Can the Yankee Stadium management conclude that the variances differ?

13.83 Xr13-83 In deciding where to invest her retirement fund, an investor recorded the weekly returns of two portfolios for 1 year. Can we conclude that portfolio 2 is riskier than portfolio 1?

13.84 Xr13-84 An important statistical measurement in service facilities (such as restaurants and banks) is the variability in service times. As an experiment, two bank tellers were observed, and the service times for each of 100 customers were recorded. Do these data allow us to infer at the 10% significance level that the variance in service times differs between the two tellers?

13.5 Inference about the Difference between Two Population Proportions

In this section, we present the procedures for drawing inferences about the difference between populations whose data are nominal. The number of applications of these techniques is almost limitless. For example, pharmaceutical companies test new drugs by comparing the new and old or the new versus a placebo. Marketing managers compare market shares before and after advertising campaigns. Operations managers compare defective rates between two machines. Political pollsters measure the difference in popularity before and after an election.

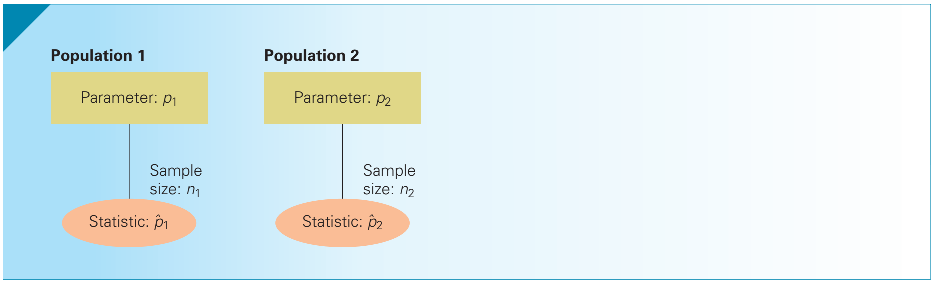
Parameter

When data are nominal, the only meaningful computation is to count the number of occurrences of each type of outcome and calculate proportions. Consequently, the parameter to be tested and estimated in this section is the difference between two population proportions $p_1 - p_2$.

Statistic and Sampling Distribution

To draw inferences about $p_1 - p_2$, we take a sample of size n_1 from population 1 and a sample of size n_2 from population 2 (Figure 13.7 depicts the sampling process).

FIGURE 13.7 Sampling From Two Populations of Nominal Data



For each sample, we count the number of successes (recall that we call anything we're looking for a success), which we label x_1 and x_2, respectively. The sample proportions are then computed:

$$\hat{p}_1 = \frac{x_1}{n_1} \quad \text{and} \quad \hat{p}_2 = \frac{x_2}{n_2}$$

Statisticians have proven that the statistic $\hat{p}_1 - \hat{p}_2$ is an unbiased consistent estimator of the parameter $p_1 - p_2$. Using the same mathematics as we did in Chapter 9 to derive the sampling distribution of the sample proportion $\hat{p}$, we determine the sampling distribution of the difference between two sample proportions.

Sampling Distribution of $\hat{p}_1 - \hat{p}_2$

1. The statistic $\hat{p}_1 - \hat{p}_2$ is approximately normally distributed provided that the sample sizes are large enough so that n_1p_1, $n_1(1 - p_1)$, n_2p_2, and $n_2(1 - p_2)$ are all greater than or equal to 5. [Because p_1 and p_2 are unknown, we express the sample size requirement as $n_1\hat{p}_1$, $n_1(1 - \hat{p}_1)$, $n_2\hat{p}_2$, and $n_2(1 - \hat{p}_2)$ are greater than or equal to 5.]
2. The mean of $\hat{p}_1 - \hat{p}_2$ is

$$E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$$

3. The variance of $\hat{p}_1 - \hat{p}_2$ is

$$V(\hat{p}_1 - \hat{p}_2) = \frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}$$

The standard error is

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}$$

Thus, the variable

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}}$$

is approximately standard normally distributed.

Testing and Estimating the Difference between Two Proportions

We would like to use the z-statistic just described as our test statistic; however, the standard error of $\hat{p}_1 - \hat{p}_2$, which is

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

is unknown because both p_1 and p_2 are unknown. As a result, the standard error of $\hat{p}_1 - \hat{p}_2$ must be estimated from the sample data. There are two different estimators of this quantity, and the determination of which one to use depends on the null hypothesis. If the null hypothesis states that $p_1 - p_2 = 0$, the hypothesized equality of the two population proportions allows us to pool the data from the two samples to produce an estimate of the common value of the two proportions p_1 and p_2. The **pooled proportion estimate** is defined as

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

Thus, the estimated standard error of $\hat{p}_1 - \hat{p}_2$ is

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n_1} + \frac{\hat{p}(1-\hat{p})}{n_2}} = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

The principle used in estimating the standard error of $\hat{p}_1 - \hat{p}_2$ is analogous to that applied in Section 13.1 to produce the pooled variance estimate s_p^2, which is used to test $\mu_1 - \mu_2$ with σ_1^2 and σ_2^2 unknown but equal. The principle roughly states that, where possible, pooling data from two samples produces a better estimate of the standard error. Here, pooling is made possible by hypothesizing (under the null hypothesis) that $p_1 = p_2$. (In Section 13.1, we used the pooled variance estimate because we assumed that $\sigma_1^2 = \sigma_2^2$.) We will call this application Case 1.

Test Statistic for $p_1 - p_2$: Case 1

If the null hypothesis specifies

$$H_0: \quad (p_1 - p_2) = 0$$

the test statistic is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Because we hypothesize that $p_1 - p_2 = 0$, we simplify the test statistic to

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

The second case applies when, under the null hypothesis, we state that $p_1 - p_2 = D$, where D is some value other than 0. Under such circumstances, we cannot pool the sample data to estimate the standard error of $\hat{p}_1 - \hat{p}_2$. The appropriate test statistic is described next as Case 2.

Test Statistic for $p_1 - p_2$: Case 2

If the null hypothesis specifies

$$H_0\colon \quad (p_1 - p_2) = D \quad (D \neq 0)$$

the test statistic is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}}$$

which can also be expressed as

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - D}{\sqrt{\dfrac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}}$$

Notice that this test statistic is determined by simply substituting the sample statistics $\hat{p}_1$ and $\hat{p}_2$ in the standard error of $\hat{p}_1 - \hat{p}_2$.

You will find that, in most practical applications (including the exercises in this book), Case 1 applies—in most problems, we want to know whether the two population proportions differ: that is,

$$H_1\colon \quad (p_1 - p_2) \neq 0$$

or if one proportion exceeds the other; that is,

$$H_1\colon \quad (p_1 - p_2) > 0 \qquad \text{or} \qquad H_1\colon \quad (p_1 - p_2) < 0$$

In some other problems, however, the objective is to determine whether one proportion exceeds the other by a specific nonzero quantity. In such situations, Case 2 applies.

We derive the interval estimator of $p_1 - p_2$ in the same manner we have been using since Chapter 10.

Confidence Interval Estimator of $p_1 - p_2$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

This formula is valid when $n_1\hat{p}_1$, $n_1(1 - \hat{p}_1)$, $n_2\hat{p}_2$, and $n_2(1 - \hat{p}_2)$ are greater than or equal to 5.

Notice that the standard error is estimated using the individual sample proportions rather than the pooled proportion. In this procedure we cannot assume that the population proportions are equal as we did in the Case 1 test statistic.

APPLICATIONS in **MARKETING**


© White Packert/The Image Bank/ Getty Images

Test Marketing

Marketing managers frequently make use of test marketing to assess consumer reaction to a change in a characteristic (such as price or packaging) of an existing product, or to assess consumers' preferences regarding a proposed new product. *Test marketing* involves experimenting with changes to the marketing mix in a small, limited test market and assessing consumers' reaction in the test market before undertaking costly changes in production and distribution for the entire market.

EXAMPLE 13.9

Test Marketing of Package Designs, Part 1

DATA
Xm13-09

The General Products Company produces and sells a variety of household products. Because of stiff competition, one of its products, a bath soap, is not selling well. Hoping to improve sales, General Products decided to introduce more attractive packaging. The company's advertising agency developed two new designs. The first design features several bright colors to distinguish it from other brands. The second design is light green in color with just the company's logo on it. As a test to determine which design is better, the marketing manager selected two supermarkets. In one supermarket, the soap was packaged in a box using the first design; in the second supermarket, the second design was used. The product scanner at each supermarket tracked every buyer of soap over a 1-week period. The supermarkets recorded the last four digits of the scanner code for each of the five brands of soap the supermarket sold. The code for the General Products brand of soap is 9077 (the other codes are 4255, 3745, 7118, and 8855). After the trial period, the scanner data were transferred to a computer file. Because the first design is more expensive, management has decided to use this design only if there is sufficient evidence to allow it to conclude that design is better. Should management switch to the brightly colored design or the simple green one?

SOLUTION

IDENTIFY

The problem objective is to compare two populations. The first is the population of soap sales in supermarket 1, and the second is the population of soap sales in supermarket 2. The data are nominal because the values are "buy General Products soap" and "buy other companies' soap." These two factors tell us that the parameter to be tested is the difference between two population proportions $p_1 - p_2$ (where p_1 and p_2 are the proportions of soap sales that are a General Products brand in supermarkets 1 and 2, respectively). Because we want to know whether there is enough evidence to adopt the brightly colored design, the alternative hypothesis is

$$H_1: \ (p_1 - p_2) > 0$$

The null hypothesis must be

$$H_0: \ (p_1 - p_2) = 0$$

which tells us that this is an application of Case 1. Thus, the test statistic is

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

COMPUTE

MANUALLY

To compute the test statistic manually requires the statistics practitioner to tally the number of successes in each sample, where success is represented by the code 9077. Reviewing all the sales reveals that

$$x_1 = 180 \qquad n_1 = 904 \qquad x_2 = 155 \qquad n_2 = 1{,}038$$

The sample proportions are

$$\hat{p}_1 = \frac{180}{904} = .1991$$

and

$$\hat{p}_2 = \frac{155}{1{,}038} = .1493$$

The pooled proportion is

$$\hat{p} = \frac{180 + 155}{904 + 1{,}038} = \frac{335}{1{,}942} = .1725$$

The value of the test statistic is

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{(.1991 - .1493)}{\sqrt{(.1725)(1 - .1725)\left(\frac{1}{904} + \frac{1}{1{,}038}\right)}} = 2.90$$

A 5% significance level seems to be appropriate. Thus, the rejection region is

$$z > z_\alpha = z_{.05} = 1.645$$

EXCEL

	A	B	C	D
1	**z-Test: Two Proportions**			
2				
3			*Supermarket 1*	*Supermarket 2*
4	Sample Proportions		0.1991	0.1493
5	Observations		904	1038
6	Hypothesized Difference		0	
7	z Stat		2.90	
8	P(Z<=z) one tail		0.0019	
9	z Critical one-tail		1.6449	
10	P(Z<=z) two-tail		0.0038	
11	z Critical two-tail		1.96	

INSTRUCTIONS

1. Type or import the data into two adjacent columns*. (Open Xm13-09.)
2. Click **Add-Ins, Data Analysis Plus,** and **Z-Test: 2 Proportions**.
3. Specify the **Variable 1 Range** (A1:A905) and the **Variable 2 Range** (B1:B1039). Type the **Code for Success** (9077), the **Hypothesized Difference** (0), and a value for α (.05).

*If one or both columns contain a blank (representing missing data) the row will have to be deleted.

MINITAB

Test and CI for Two Proportions: Supermarket 1, Supermarket 2

Event = 9077

Variable	X	N	Sample p
Supermarket 1	180	904	0.199115
Supermarket 2	155	1038	0.149326

Difference = p (Supermarket 1) – p (Supermarket 2)
Estimate for difference: 0.0497894
95% lower bound for difference: 0.0213577
Test for difference = 0 (vs > 0): Z = 2.90 P-Value = 0.002

INSTRUCTIONS

1. Type or import the data into two adjacent columns. (Open Xm13-09.) Recode the data if necessary. (Minitab requires that there be only two codes and the higher value is deemed to be a success. See Keller's website Appendix Excel and Minitab Instructions for Missing Data and Recoding data.)
2. Click **Stat, Basic Statistics,** and **2 Proportions . . . .**
3. In the **Samples in different columns** specify the **First** (Supermarket 1) and **Second** (Supermarket 2) samples. Click **Options . . . .**
4. Type the value of the **Test difference** (0), specify the **Alternative** hypothesis (greater than), and click **Use pooled estimate of *p* for test**.

Warning: If there are asterisks representing missing data, Minitab will be unable to conduct either the test or the estimate of the difference between two proportions. Click **Data** and **Sort**, which will eliminate the asterisks.

INTERPRET

The value of the test statistic is $z = 2.90$; its p-value is .0019. There is enough evidence to infer that the brightly colored design is more popular than the simple design. As a result, it is recommended that management switch to the first design.

EXAMPLE 13.10 Test Marketing of Package Designs, Part 2

DATA
Xm13-09

Suppose that in Example 13.9 the additional cost of the brightly colored design requires that it outsell the simple design by more than 3%. Should management switch to the brightly colored design?

SOLUTION

IDENTIFY

The alternative hypothesis is

$$H_1: \quad (p_1 - p_2) > .03$$

and the null hypothesis follows as

$$H_0:\ (p_1 - p_2) = .03$$

Because the null hypothesis specifies a nonzero difference, we would apply the Case 2 test statistic.

COMPUTE

MANUALLY

The value of the test statistic is

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}} = \frac{(.1991 - .1493) - (.03)}{\sqrt{\dfrac{.1991(1-.1991)}{904} + \dfrac{.1493(1-.1493)}{1{,}038}}} = 1.15$$

EXCEL

	A	B	C	D
1	**z-Test: Two Proportions**			
2				
3			*Supermarket 1*	*Supermarket 2*
4	Sample Proportions		0.1991	0.1493
5	Observations		904	1038
6	Hypothesized Difference		0.03	
7	z Stat		1.14	
8	P(Z<=z) one tail		0.1261	
9	z Critical one-tail		1.6449	
10	P(Z<=z) two-tail		0.2522	
11	z Critical two-tail		1.96	

INSTRUCTIONS

Use the same commands we used previously, except specify that the **Hypothesized Difference** is .03. Excel will apply the Case 2 test statistic when a nonzero value is typed.

MINITAB

Test and CI for Two Proportions: Supermarket 1, Supermarket 2

Event = 9077

Variable	X	N	Sample p
Supermarket 1	180	904	0.199115
Supermarket 2	155	1038	0.149326

Difference = p (Supermarket 1) – p (Supermarket 2)
Estimate for difference: 0.0497894
95% lower bound for difference: 0.0213577
Test for difference = 0.03 (vs > 0.03): Z = 1.14 P-Value = 0.126

INSTRUCTIONS

Use the same commands detailed previously except at step 4, specify that the **Test difference** is .03 and do not click **Use pooled estimate of *p* for test**.

INTERPRET

There is not enough evidence to infer that the proportion of soap customers who buy the product with the brightly colored design is more than 3% higher than the proportion of soap customers who buy the product with the simple design. In the absence of sufficient evidence, the analysis suggests that the product should be packaged using the simple design.

EXAMPLE 13.11 Test Marketing of Package Designs, Part 3

DATA
Xm13-09

To help estimate the difference in profitability, the marketing manager in Examples 13.9 and 13.10 would like to estimate the difference between the two proportions. A confidence level of 95% is suggested.

SOLUTION

IDENTIFY

The parameter is $p_1 - p_2$, which is estimated by the following confidence interval estimator:

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

COMPUTE

MANUALLY

The sample proportions have already been computed. They are

$$\hat{p}_1 = \frac{180}{904} = .1991$$

and

$$\hat{p}_2 = \frac{155}{1038} = .1493$$

The 95% confidence interval estimate of $p_1 - p_2$ is

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

$$= (.1991 - .1493) \pm 1.96\sqrt{\frac{.1991(1 - .1991)}{904} + \frac{.1493(1 - .1493)}{1{,}038}}$$

$$= .0498 \pm .0339$$

$$\text{LCL} = .0159 \qquad \text{and} \qquad \text{UCL} = .0837$$

EXCEL

	A	B	C	D
1	z-Estimate: Two Proportions			
2				
3			*Supermarket 1*	*Supermarket 2*
4	Sample Proportions		0.1991	0.1493
5	Observations		904	1038
6				
7	LCL		0.0159	
8	UCL		0.0837	

INSTRUCTIONS

1. Type or import the data into two adjacent columns*. (Open Xm13-09.)
2. Click **Add-Ins, Data Analysis Plus,** and **Z-Estimate: 2 Proportions**.
3. Specify the **Variable 1 Range** (A1:A905) and the **Variable 2 Range** (B1:B1039). Specify the **Code for Success** (9077) and a value for α (.05).

MINITAB

Test and CI for Two Proportions: Supermarket 1, Supermarket 2

Event = 9077

Variable	X	N	Sample p
Supermarket 1	180	904	0.199115
Supermarket 2	155	1038	0.149326

Difference = p (Supermarket 1) – p (Supermarket 2)
Estimate for difference: 0.0497894
95% CI for difference: (0.0159109, 0.0836679)
Test for difference = 0 (vs not = 0): Z = 2.88 P-Value = 0.004

INSTRUCTIONS

Follow the commands to test hypotheses about two proportions. Specify the alternative hypothesis as **not equal** and do not click **Use pooled estimate of *p* for test**.

INTERPRET

We estimate that the market share for the brightly colored design is between 1.59% and 8.37% larger than the market share for the simple design.

*If one or both columns contain a blank (representing missing data) the row must be deleted.

American National Election Survey

Comparing Democrats and Republicans: Who Is More Educated?

© KamiGami/Shutterstock

The problem objective is to compare two populations (Democrats and Republicans). The data are nominal. We've recoded the data so that all categories greater than 0 are represented by 2, which will be our definition of success. The parameter is $p_1 - p_2$, where p_1 = proportion of Democrats with at least a bachelor's degree and p_2 = proportion of Republicans with at least a bachelor's degree. The hypotheses are

$$H_0: \quad (p_1 - p_2) = 0$$
$$H_1: \quad (p_1 - p_2) < 0$$

The null hypothesis tells us that this is an application of Case 1. Thus, the test statistic is

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

EXCEL

	A	B	C	D
1	**z-Test: Two Proportions**			
2				
3			*Democrats*	*Republicans*
4	Sample Proportions		0.6246	0.7085
5	Observations		341	271
6	Hypothesized Difference		0	
7	z Stat		−2.18	
8	P(Z<=z) one tail		0.0147	
9	z Critical one-tail		1.6449	
10	P(Z<=z) two-tail		0.0294	
11	z Critical two-tail		1.96	

We copied the variables DEGREE and PARTY into a new spreadsheet and sorted the two columns by party. We collected the data for code 1 (Democrats) and code 2 (Republicans), recoded the data, and conducted the z-test of $p_1 - p_2$, using 2 as a success.

MINITAB

Test and CI for Two Proportions: Dem, Rep

Event = 2

Variable	X	N	Sample p
Dem	213	341	0.624633
Rep	192	271	0.708487

Difference = p (Dem) – p (Rep)
Estimate for difference: – 0.0838537
95% lower bound for difference: – 0.0212260
Test for difference = 0 (vs < 0): Z = –2.18 P-Value = 0.015

We copied the variables DEGREE and PARTY into new columns and sorted the two columns by party. We collected the data for code 1 (Democrats) and code 2 (Republicans). We then sorted each column to remove the asterisks. Finally, we coded the data so that codes 1 to 7 became 2 and 0 remained 0. We then conducted the z-test of $p_1 - p_2$.

INTERPRET

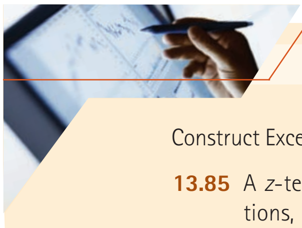
There is sufficient evidence to infer that the proportion of Republicans with at least a bachelor's degree is greater than the proportion of Democrats with at least a bachelor's degree. The popular perception (judging from the media, some politicians, and some comedians) that Democrats are more educated than Republicans is not supported by these data. At the end of this section, you will have the opportunity to test this perception again.

The factors that identify the inference about the difference between two proportions are listed below.

Factors That Identify the z-Test and Estimator of $p_1 - p_2$

1. **Problem objective**: Compare two populations
2. **Data type**: Nominal

Exercises



DO-IT-YOURSELF EXCEL

Construct Excel spreadsheets for each of the following:

13.85 A z-test of $p_1 - p_2$. Inputs: Sample proportions, sample sizes, and hypothesized difference between two populations. Outputs: Test statistic, critical values, and one- and two-tail p-values. Tools: **NORMSINV, NORMSDIST**

13.85 A z-estimate of $p_1 - p_2$. Inputs: Sample proportions, sample sizes, and confidence level. Outputs: Test statistic, one- and two-tail p-values. Tools: **NORMSINV**

Developing an Understanding of Statistical Concepts

Exercises 13.87 to 13.89 are "what-if" analyses designed to determine what happens to the test statistics and interval estimates when elements of the statistical inference change. These problems can be solved manually, using Do-It-Yourself Excel spreadsheets you created, or using Minitab.

13.87 Random samples from two binomial populations yielded the following statistics:

$$\hat{p}_1 = .45 \quad n_1 = 100 \quad \hat{p}_2 = .40 \quad n_2 = 100$$

a. Calculate the p-value of a test to determine whether we can infer that the population proportions differ.

b. Repeat part (a) increasing the sample sizes to 400.

c. Describe what happens to the p-value when the sample sizes increase.

13.88 These statistics were calculated from two random samples:

$$\hat{p}_1 = .60 \quad n_1 = 225 \quad \hat{p}_2 = .55 \quad n_2 = 225$$

a. Calculate the p-value of a test to determine whether there is evidence to infer that the population proportions differ.

b. Repeat part (a) with $\hat{p}_1 = .95$ and $\hat{p}_2 = .90$.

c. Describe the effect on the p-value of increasing the sample proportions.

d. Repeat part (a) with $\hat{p}_1 = .10$ and $\hat{p}_2 = .05$.
e. Describe the effect on the *p*-value of decreasing the sample proportions.

13.89 After sampling from two binomial populations we found the following.

$\hat{p}_1 = .18 \quad n_1 = 100 \quad \hat{p}_2 = .22 \quad n_2 = 100$

a. Estimate with 90% confidence the difference in population proportions.
b. Repeat part (a) increasing the sample proportions to .48 and .52, respectively.
c. Describe the effects of increasing the sample proportions.

Applications

13.90 Many stores sell extended warranties for products they sell. These are very lucrative for store owners. To learn more about who buys these warranties, a random sample was drawn of a store's customers who recently purchased a product for which an extended warranty was available. Among other variables, each respondent reported whether he or she paid the regular price or a sale price and whether he or she purchased an extended warranty.

	Regular Price	Sale Price
Sample size	229	178
Number who bought extended warranty	47	25

Can we conclude at the 10% significance level that those who paid the regular price are more likely to buy an extended warranty?

13.91 A firm has classified its customers in two ways: (1) according to whether the account is overdue and (2) whether the account is new (less than 12 months) or old. To acquire information about which customers are paying on time and which are overdue, a random sample of 292 customer accounts was drawn. Each was categorized as either a new account or an old account, and whether the customer has paid or is overdue. The results are summarized next.

	New Account	Old Account
Sample size	83	209
Overdue account	12	49

Is there enough evidence at the 5% significance level to infer that new and old accounts are different with respect to overdue accounts?

13.92 Credit scorecards are used by financial institutions to help decide to whom loans should be granted (see the Applications in Banking: Credit Scorecards summary on page 63). An analysis of the records of a random sample of loans at one bank produced the following results:

	Score Below 600	Score 600 or More
Sample size	562	804
Number defaulted	11	7

Do these results allow us to conclude that those who score below 600 are more likely to default than those who score 600 or more? Use a 10% significance level.

13.93 Surveys have been widely used by politicians around the world as a way of monitoring the opinions of the electorate. Six months ago, a survey was undertaken to determine the degree of support for a national party leader. Of a sample of 1,100, 56% indicated that they would vote for this politician. This month, another survey of 800 voters revealed that 46% now support the leader.
a. At the 5% significance level, can we infer that the national leader's popularity has decreased?
b. At the 5% significance level, can we infer that the national leader's popularity has decreased by more than 5%?
c. Estimate with 95% confidence the decrease in percentage support between now and 6 months ago.

13.94 The process that is used to produce a complex component used in medical instruments typically results in defective rates in the 40% range. Recently, two innovative processes have been developed to replace the existing process. Process 1 appears to be more promising, but it is considerably more expensive to purchase and operate than process 2. After a thorough analysis of the costs, management decides that it will adopt process 1 only if the proportion of defective components it produces is more than 8% smaller than that produced by process 2. In a test to guide the decision, both processes were used to produce 300 components. Of the 300 components produced by process 1, 33 were found to be defective, whereas 84 out of the 300 produced by process 2 were defective. Conduct a test using a significance level of 1% to help management make a decision.

APPLICATIONS in OPERATIONS MANAGEMENT

© Ryan McVay/PhotoDisc/ Getty Images

Pharmaceutical and Medical Experiments

When new products are developed, they are tested in several ways. First, does the new product work? Second, is it better than the existing product? Third, will customers buy it at a price that is profitable? Performing a customer survey or some other experiment that yields the information needed often tests the last question. This experiment is usually the domain of the marketing manager.

The other two questions are dealt with by the developers of the new product, which usually means the research department or the operations manager. When the product is a new drug, there are particular ways in which the data are gathered. The sample is divided into two groups. One group is assigned the new drug and the other is assigned a placebo, a pill that contains no medication. The experiment is often called "double-blind" because neither the subjects who take the drug nor the physician or scientist who provides the drug knows whether any individual is taking the drug or the placebo. At the end of the experiment, the data that are compiled allow statistics practitioners to do their work. Exercises 13.95–13.99 are examples of this type of statistical application. Exercise 13.100 describes a health-related problem where the use of a placebo is not possible.

13.95 Cold and allergy medicines have been available for a number of years. One serious side effect of these medications is that they cause drowsiness, which makes them dangerous for industrial workers. In recent years, a nondrowsy cold and allergy medicine has been developed. One such product, Hismanal, is claimed by its manufacturer to be the first once-a-day nondrowsy allergy medicine. The nondrowsy part of the claim is based on a clinical experiment in which 1,604 patients were given Hismanal and 1,109 patients were given a placebo. Of the first group, 7.1% reported drowsiness; of the second group, 6.4% reported drowsiness. Do these results allow us to infer at the 5% significance level that Hismanal's claim is false?

13.96 Plavix is a drug that is given to angioplasty patients to help prevent blood clots. A researcher at McMaster University organized a study that involved 12,562 patients in 482 hospitals in 28 countries. All the patients had acute coronary syndrome, which produces mild heart attacks or unstable angina, chest pain that may precede a heart attack. The patients were divided into two equal groups. Group 1 received daily Plavix pills; group 2 received a placebo. After 1 year, 9.3% of patients on Plavix suffered a stroke or new heart attack or had died of cardiovascular disease, compared with 11.5% of those who took the placebo.

a. Can we infer that Plavix is effective?
b. Describe your statistical analysis in a report to the marketing manager of the pharmaceutical company.

13.97 In a study that was highly publicized, doctors discovered that aspirin seems to help prevent heart attacks. The research project, which was scheduled to last for 5 years, involved 22,000 American physicians (all male). Half took an aspirin tablet three times per week, and the other half took a placebo on the same schedule. The researchers tracked all of the volunteers and updated the records regularly. Among the physicians who took aspirin, 104 suffered heart attacks; 189 physicians who took the placebo had heart attacks.

a. Determine whether these results indicate that aspirin is effective in reducing the incidence of heart attacks.
b. Write a report that describes the results of this experiment.

13.98 Exercise 13.97 described the experiment that determined that taking aspirin daily reduces one's probability of suffering a heart attack. The study was conducted in 1982; at that time, the mean age of the physicians was 50. In the years following the experiment, the physicians were monitored for other medical conditions. One of these was the incidence of cataracts. There were 1,084 cataracts in the aspirin group and 997 in the placebo group. Do these statistics allow researchers to conclude that aspirin leads to more cataracts?

13.99 According to the Canadian Cancer Society, more than 21,000 women will be diagnosed with breast cancer every year and more than 5,000 will die. (U.S. figures are more than 10 times those in Canada.) Surgery is generally considered the first method of treatment. However, many women suffer recurrences of cancer. For this reason, many women are treated with tamoxifen. But after 5 years, tumors develop a resistance to tamoxifen. A new drug called *letrozole* was developed by Novartis Pharmaceuticals to replace tamoxifen. To determine its effectiveness, a study involving 5,187 breast cancer survivors from Canada, the United States, and Europe was undertaken. Half the sample received letrozole and the other half a placebo. The study was to run for 5 years. However, after only 2.5 years, it was determined that 132 women receiving the placebo and 75 taking the drug had recurrences of their cancers. (The study was published in the *New England Journal of Medicine*.)

a. Do these results provide sufficient evidence to infer that letrozole works?
b. Prepare a presentation to the board of directors of Novartis describing your analysis.

13.100 A study described in the *British Medical Journal* (January 2004) sought to determine whether exercise would help extend the lives of patients with heart failure. A sample of 801 patients with heart failure was recruited; 395 received exercise training and 406 did not. There were 88 deaths among the exercise group and 105 among those who did not exercise. Can researchers infer that exercise training reduces mortality?

Exercises 13.101–13.125 require the use of a computer and software. Use a 5% significance level unless specified otherwise. The answers to Exercises 13.101 to 13.112 may be calculated manually. See Appendix A for the sample statistics.

13.101 Xr13-101 Automobile magazines often compare models and rate them in various ways. One question that is often asked of car owners, Would you buy the same model again? Suppose that a researcher for one magazine asked a random sample of Lexus owners and a random sample of Acura owners whether they plan to buy another Lexus or Acura the next time they shop for a new car. The responses (1 = no and 2 = yes) were recorded. Do these data allow the researcher to infer that the two populations of car owners differ in their satisfaction levels?

13.102 Xr13-102 An insurance company is thinking about offering discounts on its life-insurance policies to nonsmokers. As part of its analysis, the company randomly selects 200 men who are 60 years old and asks them whether they smoke at least one pack of cigarettes per day and if they have ever suffered from heart disease (2 = suffer from heart disease, and 1 = do not suffer from heart disease).

a. Can the company conclude at the 10% significance level that smokers have a higher incidence of heart disease than nonsmokers?
b. Estimate with 90% confidence the difference in the proportions of men suffering from heart disease between smokers and nonsmokers.

13.103 Xr13-103 Has the illicit use of drugs decreased over the past 10 years? Government agencies have undertaken surveys of Americans 12 years of age and older. Each was asked whether he or she used drugs at least once in the previous month. The results of this year's survey and the results of the survey completed 10 years ago were recorded as 1 = no and 2 = yes. Can we infer that the use of illicit drugs in the United States has increased in the past decade? (Adapted from the U.S. Substance Abuse and Mental Health Services Administration, National Household Survey on Drug Abuse.)

13.104 Xr13-104 It has been estimated that the oil sands in Alberta, Canada, contain 2 trillion barrels of oil. However, recovering the oil damages the environment. A survey of Canadians and Americans was asked, What is more important to you with regards to the oil sands: (1) environmental concerns or (2) the potential of a secure nonforeign supply of oil to North America? Do these data allow you to conclude that Canadians and Americans differ in their responses to this question? (Source: Flieshman-Hillard Oilsands Survey.)

13.105 Xr13-105 An operations manager of a computer chip maker is in the process of selecting a new machine to replace several older ones. Although technological innovations have improved the production process, it is quite common for the machines to produce defective chips. The operations manager must choose between two machines. The cost of machine A is several thousand dollars greater than the cost of machine B. After an analysis of the costs, it was determined that machine A is warranted, provided that its defective rate is more than 2% less than that of machine B. To help decide, both machines are used to produce 200 chips each. Each chip was examined, and whether it was defective (code = 2) or not (code = 1) was recorded. Should the operations manager select machine A?

13.106 Xr13-106 Parents often urge their children to get more education, not only for the increased income but also to perhaps work less hard. A survey asked a random sample of Canadians whether they work 11 or more hours a day (1 = no, 2 = yes) and whether they completed high school only or completed post-secondary education. Can we infer that those with more education are less likely to work 11 hours or more per day? (Source: Harris/Decima survey.)

13.107 Xr13-107 Are Americans becoming more unhappy at work? A survey of Americans in 2008 and again this year asked whether they were satisfied with their jobs (1 = no, 2 = yes). Can we infer that more Americans are unhappy compared to 2008?

Public Opinion about Global Warming and Climate Change

In Chapters 3 and 4, we described the issue of global warming and pointed out that Earth has not warmed since 1998, explaining why the media now refer to the problem as *climate change*, and that a weak linear relationship exists between temperature anomalies and CO_2 levels. In the last few years, news stories have appeared that seem to cast doubt on the entire theory. To measure the effect on public opinion, several surveys have been conducted. Below we describe two surveys in the United States, Canada, and Britain. For each exercise and each country, determine whether there is sufficient evidence that the belief that global warming is real has fallen.

13.108 Xr13-108 The following question was asked in the three countries in November 2009 and December 2009.

Which of the following statements comes closest to your view of global warming (or climate change)?

1. Global warming is a fact and is mostly caused by emissions from vehicles and industrial facilities.
2. Global warming is a fact and is mostly caused by natural changes.
3. Global warming is a theory and has not yet been proven.
4. Not sure

13.109 Xr13-109 Do you agree (1 = Yes, 2 = No) that climate change and how we respond to it are among the biggest issues that you worry about today? The question was asked in the three countries in November 2008 and November 2009.

APPLICATIONS in MARKETING


© Eyewire Collection/Getty Images

Market Segmentation

In Section 12.4, we introduced market segmentation and described how the size of market segments can be estimated. Once the segments have been defined we can use statistical techniques to determine whether members of the segments differ in their purchases of a firm's products.

13.110 Xr13-110* The market for breakfast cereals has been divided into several segments related to health. One company identified a segment as health-conscious adults. The marketing manager would like to know whether this segment is more likely to purchase its Special X cereal, which is pitched toward the health-conscious segment. A survey of adults was undertaken. On the basis of several probing questions, each was classified as either a member of the health-conscious group (code = 1) or not (code = 2). Each respondent was also asked whether he or she buys Special X (1 = no, 2 = yes). The data were recorded in stacked format. Can we infer from these data that health-conscious adults are more likely to buy Special X?

13.111 Xr13-111* Quik Lube is a company that offers an oil-change service while the customer waits. Its market has been broken down into the following segments:

1: Working men and women too busy to wait at a dealer or service center
2: Spouses who work in the home
3: Retired persons
4: Other

A random sample of car owners was drawn. All owners classified their market segment and also reported whether they usually use such services as Quik Lube (1 = yes, and 2 = no). These data are stored in stacked format.

a. Determine whether members of segment 1 are more likely than members of segment 4 to respond that they usually use the service?
b. Can we infer that retired persons and spouses who work in the home differ in their use of services such as Quik Lube?

13.112 Xr13-112 Telemarketers obtain names and telephone numbers from several sources. To determine whether one particular source is better than a second, a random sample of names and numbers from the two different sources was obtained. For each potential customer, a statistics practitioner recorded whether that individual made a purchase (code = 2) or not (code = 1). Can we infer that differences exist between the two sources?

GENERAL SOCIAL SURVEY EXERCISES

13.113 GSS2008* A generation ago, men were more likely to attend a university and acquire a graduate degree than women. However, women now appear to be attending universities in greater numbers than men. To gauge the extent of the difference, test to determine whether men and women (SEX: 1 = Male and 2 = Female) differ in completing a graduate degree (DEGREE: 4 = Graduate).

13.114 GSS2008* The deep recession of 2008–2010 may have changed patterns of employment. Because of the large number of layoffs an increasing number of individuals have chosen to work for themselves. The question arises, Do men and women (SEX: 1 = Male and 2 = Female) differ in their decision to work for themselves? (WRKSLF: 1 = self-employed, 2 = someone else.) Conduct a test to answer the question.

For each of the following four exercises, determine whether men and women are likely to differ in answering each question correctly.

13.115 GSS2008* A doctor tells a couple that there is one chance in four that their child will have an inherited disease. Does this mean that if the first child has the illness, the next three will not (ODDS1)? 1 = Yes, 2 = No. Correct answer: No.

13.116 GSS2008* A doctor tells a couple that there is one chance in four that their child will have an inherited disease. Does this mean that each of the couple's children will have the same risk of suffering the illness (ODDS2)? 1 = Yes, 2 = No. Correct answer: Yes.

13.117 GSS2008* True or false—Earth's center is very hot. 1 = True, 2 = False. Correct answer: True.

13.118 GSS2008* Does Earth go around the Sun or does the Sun go around Earth? 1 = Earth around Sun, 2 = Sun around Earth. Correct answer: Earth around Sun.

For each of the following variables, conduct a test to determine whether there is a difference between 2008 and 2006.

13.119 GSS2008* GSS2006* WRKGOVT: Are (were) you employed by the federal, state, or local government or by a private employer (including not-for-profit organizations)? 1 = Government, 2 = Private.

13.120 GSS2008* GSS2006* CAPPUN: Do you favor capital punishment for murder? 1 = Favor, 2 = Oppose.

13.121 GSS2008* GSS2006* GUNLAW: Do you favor requiring a police permit to buy a gun? 1 = Favor, 2 = Oppose.

13.122 GSS2002* GSS2004* GSS2006* GSS2008* Test to determine whether Democrats and Republicans (PARTYID: 0 and 1 = Democrat and 5 and 6 = Republicans) differ in each of the years 2002, 2004, 2006, and 2008 in completing a graduate degree (DEGREE: 4 = Graduate).

American National Election Survey Exercises

For each of the following variables, conduct a test to determine whether Democrats and Republicans (PARTY: 1 = Democrat and 2 = Republicans) differ.

13.123 ANES2008* Likely to be employed (EMPLOY: 1 = Working now, 2–8 = Other categories).

13.124 ANES2008* Have health insurance (HEALTH 1 = Yes, 5 = No).

13.125 ANES2008* Always vote (OFTEN: 1 = Always, 2, 3, 4 = Other categories).

Chapter Summary

In this chapter, we presented a variety of techniques that allow statistics practitioners to compare two populations. When the data are interval and we are interested in measures of central location, we encountered two more factors that must be considered when choosing the appropriate technique. When the samples are independent, we can use either the **equal-variances** or **unequal-variances formulas**. When the samples are **matched pairs**, we have only one set of formulas. We introduced the ***F*-statistic**, which is used to make inferences about two population variances. When the data are nominal, the parameter of interest is the difference between two proportions. For this parameter, we had two test statistics and one interval estimator. Finally, we discussed **observational** and **experimental data**, important concepts in attempting to interpret statistical findings.

Important Terms

Pooled variance estimator 451
Equal-variances test statistic 451
Equal-variances confidence interval estimator 451
Unequal-variances test statistic 452
Unequal-variances confidence interval estimator 452
Observational data 472
Experimental data 472
Matched pairs experiment 479
Mean of the population of differences 479
Numerator degrees of freedom 490
Denominator degrees of freedom 490
Pooled proportion estimate 497

SYMBOLS

Symbol	Pronounced	Represents
s_p^2	s sub p squared	Pooled variance estimator
μ_D	mu sub D or mu D	Mean of the paired differences
$\bar{x}_D$	x bar sub D or x bar D	Sample mean of the paired differences
s_D	s sub D or s D	Sample standard deviation of the paired differences
n_D	n sub D or n D	Sample size of the paired differences
$\hat{p}$	p hat	Pooled proportion

FORMULAS

Equal-variances t-test of $\mu_1 - \mu_2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad \nu = n_1 + n_2 - 2$$

Equal-variances interval estimator of $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \qquad \nu = n_1 + n_2 - 2$$

Unequal-variances t-test of $\mu_1 - \mu_2$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}} \qquad \nu = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

Unequal-variances interval estimator of $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \qquad \nu = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1}}$$

t-test of μ_D

$$t = \frac{\bar{x}_D - \mu_D}{s_D/\sqrt{n_D}} \qquad \nu = n_D - 1$$

t-estimator of μ_D

$$\bar{x}_D \pm t_{\alpha/2}\frac{s_D}{\sqrt{n_D}} \qquad \nu = n_D - 1$$

F-test of σ_1^2/σ_2^2

$$F = \frac{s_1^2}{s_2^2} \qquad \nu_1 = n_1 - 1 \text{ and } \nu_2 = n_2 - 1$$

F-estimator of σ_1^2/σ_2^2

$$\text{LCL} = \left(\frac{s_1^2}{s_2^2}\right)\frac{1}{F_{\alpha/2,\nu_1,\nu_2}}$$

$$\text{UCL} = \left(\frac{s_1^2}{s_2^2}\right)F_{\alpha/2,\nu_2,\nu_1}$$

z-test and estimator of $p_1 - p_2$

$$\text{Case 1:} \quad z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$\text{Case 2:} \quad z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}}$$

z-estimator of $p_1 - p_2$

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

COMPUTER OUTPUT AND INSTRUCTIONS

Technique	Excel	Minitab
Unequal-variances t-test of $\mu_1 - \mu_2$	461	462
Unequal-variances estimator of $\mu_1 - \mu_2$	463	463
Equal-variances t-test of $\mu_1 - \mu_2$	456	456
Equal-variances estimator of $\mu_1 - \mu_2$	457	458
t-test of μ_D	480	481
t-estimator of μ_D	482	482
F-test of σ_1^2/σ_2^2	491	492
F-estimator of σ_1^2/σ_2^2	493	493
z-test of $p_1 - p_2$ (Case 1)	500	501
z-test of $p_1 - p_2$ (Case 2)	502	502
z-estimator of $p_1 - p_2$	504	504

CHAPTER EXERCISES

The following exercises require the use of a computer and software. Use a 5% significance level unless specified otherwise.

13.126 Xr13-126 Obesity among children has quickly become an epidemic across North America. Television and video games are part of the problem. To gauge to what extent nonparticipation in organized sports contributes to the crisis, surveys of children 5 to 14 years old were conducted this year and 10 years ago. The gender of the child and whether he or she participated in organized sports (1 = No, 2 = Yes) were recorded.

a. Can we conclude that there has been a decrease in participation among boys over the past 10 years?
b. Repeat part (a) for girls.
c. Can we infer that girls are less likely to participate than boys this year?

13.127 Xr13-127 A restaurant located in an office building decides to adopt a new strategy for attracting customers to the restaurant. Every week it advertises in the city newspaper. To assess how well the advertising is working, the restaurant owner recorded the weekly gross sales for the 15 weeks after the campaign began and the weekly gross sales for the 24 weeks immediately prior to the campaign. Can the restaurateur conclude that the advertising campaign is successful?

13.128 Refer to Exercise 13.127. Assume that the profit is 20% of the gross. If the ads cost $50 per week, can the restaurateur conclude that the ads are profitable?

13.129 Xr13-129 How important to your health are regular vacations? In a study, a random sample of men and women were asked how frequently they take vacations. The men and women were divided into two groups each. The members of group 1 had suffered a heart attack; the members of group 2 had not. The number of days of vacation last year was recorded for each person. Can we infer that men and women who suffer heart attacks vacation less than those who did not suffer a heart attack?

13.130 Xr13-130 Research scientists at a pharmaceutical company have recently developed a new nonprescription sleeping pill. They decide to test its effectiveness by measuring the time it takes for people to fall asleep after taking the pill. Preliminary analysis indicates that the time to fall asleep varies considerably from one person to another. Consequently, the researchers organize the experiment in the following way. A random sample of 100 volunteers who regularly suffer from insomnia is chosen. Each person is given one pill containing the newly developed drug and one placebo. (They do not know whether the pill they are taking is the placebo or the real thing, and the order of use is random.) Each participant is fitted with a device that measures the time until sleep occurs. Can we conclude that the new drug is effective?

13.131 Xr13-131 The city of Toronto boasts four daily newspapers. Not surprisingly, competition is keen. To help learn more about newspaper readers, an advertiser selected a random sample of people who bought their newspapers from a street vendor and people who had the newspaper delivered to their homes. All were asked how many minutes they spent reading their newspapers. Can we infer that the amount of time reading differs between the two groups?

13.132 Xr13-132 In recent years, a number of state governments have passed mandatory seat-belt laws. Although the use of seat belts is known to save lives and reduce serious injuries, compliance with seat-belt laws is not universal. In an effort to increase the use of seat belts, a government agency sponsored a 2-year study. Among its objectives was to determine whether there was enough evidence to infer that seat-belt usage increased between last year and this year. To test this belief, random samples of drivers last year and this year were asked whether they always use their seat belts (2 = wear seat belt, 1 = do not wear seat belt). Can we infer that seat-belt usage has increased over the last year?

13.133 Xr13-133 An important component of the cost of living is the amount of money spent on housing. Housing costs include rent (for tenants), mortgage payments and property tax (for home owners), heating, electricity, and water. An economist undertook a 5-year study to determine how housing costs have changed. Five years ago, he took a random sample of 200 households and recorded the percentage of total income spent on housing. This year, he took another sample of 200 households.

a. Conduct a test (with $\alpha = .10$) to determine whether the economist can infer that housing cost as a percentage of total income has increased over the last 5 years.
b. Use whatever statistical method you deem appropriate to check the required condition(s) of the test used in part (a).

13.134 Xr13-134 In designing advertising campaigns to sell magazines, it is important to know how much time each of several demographic groups spends reading magazines. In a preliminary study, 40 people were randomly selected. Each was asked how much time per week he or she spends reading magazines; in addition, each was categorized by both gender and income level (high or low). The data are stored in the following way: column 1 = time spent reading magazines per week in minutes for all respondents; column 2 = gender (1 = male, 2 = female); column 3 = income level (1 = low, 2 = high).

a. Is there sufficient evidence at the 10% significance level to conclude that men and women differ in the amount of time spent reading magazines?
b. Is there sufficient evidence at the 10% significance level to conclude that high-income individuals devote more time to reading magazines than low-income people?

13.135 Xr13-135 In a study to determine whether gender affects salary offers for graduating MBA students, 25 pairs of students were selected. Each pair consisted of a female and a male student who were matched according to their grade point averages, courses taken, ages, and previous work experience. The highest salary offered (in thousands of dollars) to each graduate was recorded.

a. Is there enough evidence at the 10% significance level to infer that gender is a factor in salary offers?
b. Discuss why the experiment was organized in the way it was.
c. Is the required condition for the test in part (a) satisfied?

13.136 Xr13-136 Have North Americans grown to distrust television and newspaper journalists? A study was conducted this year to compare what Americans currently think of the news media versus what they said 3 years ago. The survey asked respondents whether they agreed that the news media tends to favor one side when reporting on political and social issues. A random sample of people was asked to participate in this year's survey. The results of a survey of another random sample taken 3 years ago are also available. The responses are 2 = agree and 1 = disagree. Can we conclude at the 10% significance level that Americans have become more distrustful of television and newspaper reporting this year than they were 3 years ago?

13.137 Xr13-137 Before deciding which of two types of stamping machines should be purchased, the plant manager of an automotive parts manufacturer wants to determine the number of units that each produces. The two machines differ in cost, reliability, and productivity. The firm's accountant has calculated that machine A must produce 25 more nondefective units per hour than machine B to warrant buying machine A. To help decide, both machines were operated for 24 hours. The total number of units and the number of defective units produced by each machine per hour were recorded. These data are stored in the following way: column 1 = total number of units produced by machine A, column 2 = number of defectives produced by machine A, column 3 = total number of units produced by machine B, and column 4 = number of defectives produced by machine B. Determine which machine should be purchased.

13.138 Refer to Exercise 13.137. Can we conclude that the defective rate differs between the two machines?

13.139 Xr13-139 The growing use of bicycles to commute to work has caused many cities to create exclusive bicycle lanes. These lanes are usually created by disallowing parking on streets that formerly allowed curbside parking. Merchants on such streets complain that the removal of parking will cause their businesses to suffer. To examine this problem, the mayor of a large city decided to launch an experiment on one busy street that had 1-hour parking meters. The meters were removed, and a bicycle lane was created. The mayor asked the three businesses (a dry cleaner, a doughnut shop, and a convenience store) in one block to record daily sales for two complete weeks (Sunday to Saturday) before the change and two complete weeks after the change. The data are stored as follows: column 1 = day of the week, column 2 = sales before change for dry cleaner, column 3 = sales after change for dry cleaner, column 4 = sales before change for doughnut shop, column 5 = sales after change for doughnut shop, column 6 = sales before change for convenience store, and column 7 = sales after change for convenience store. What conclusions can you draw from these data?

13.140 Xr13-140 Researchers at the University of Ohio surveyed 219 students and found that 148 had Facebook accounts. All students were asked for their current grade point average. Do the data allow us to infer that Facebook users have lower GPAs?

13.141 Xr13-141 Clinical depression is linked to several other diseases. Scientists at Johns Hopkins University undertook a study to determine whether heart disease is one of these. A group of 1,190 male medical students was tracked over a 40-year period. Of these, 132 had suffered clinically diagnosed depression. For each student, the scientists recorded whether the student died of a heart attack (code = 2) or did not (code = 1).

a. Can we infer at the 1% significance level that men who are clinically depressed are more likely to die from heart disease?
b. If the answer to part (a) is "yes," can you interpret this to mean that depression causes heart disease? Explain.

13.142 Xr13-142 High blood pressure (hypertension) is a leading cause of strokes. Medical researchers are constantly seeking ways to treat patients suffering from this condition. A specialist in hypertension claims that regular aerobic exercise can reduce high blood pressure just as successfully as drugs, with none of the adverse side effects. To test the claim, 50 patients who suffer from high blood pressure were chosen to participate in an experiment. For 60 days, half the sample exercised three times per week for 1 hour and did not take medication; the other half took the standard medication. The percentage reduction in blood pressure was recorded for each individual.

a. Can we conclude at the 1% significance level that exercise is more effective than medication in reducing hypertension?
b. Estimate with 95% confidence the difference in mean percentage reduction in blood pressure between drugs and exercise programs.
c. Check to ensure that the required condition(s) of the techniques used in parts (a) and (b) is satisfied.

13.143 Xr13-143 Most people exercise in order to lose weight. To determine better ways to lose weight, a random sample of male and female exercisers was divided into groups. The first group exercised vigorously twice a week. The second group exercised moderately four times per week. The weight loss for each individual was recorded. Can we infer that people who exercise moderately more frequently lose more weight than people who exercise vigorously?

13.144 Xr13-144 After observing the results of the test in Exercise 13.143, a statistics practitioner organized another experiment. People were matched according to gender, height, and weight. One member of each matched pair then exercised vigorously twice a week, and the other member exercised moderately four times per week. The weight losses were recorded. Can we infer that people who exercise moderately lose more weight?

13.145 Xr13-145 “Pass the Lotion,” a long-running television commercial for Special K cereal, features a flabby sunbather who asks his wife to smear sun lotion on his back. A random sample of Special K customers and a random sample of people who do not buy Special K were asked to indicate whether they liked (code = 1) or disliked (code = 2) the ad. Can we infer that Special K buyers like the ad more than nonbuyers?

13.146 Xr13-146 Refer to Exercise 13.145. The respondents were also asked whether they thought the ad would be effective in selling the product. The responses (1 = Yes and 2 = No) were recorded. Can we infer that Special K buyers are more likely to respond yes than nonbuyers?

13.147 Xr13-147 Most English professors complain that students don’t write very well. In particular, they point out that students often confuse quality and quantity. A study at the University of Texas examined this claim. In the study, undergraduate students were asked to compare the cost benefits of Japanese and American cars. All wrote their analyses on computers. Unbeknownst to the students, the computers were rigged so that some students would have to type twice as many words to fill a single page. The number of words used by each student was recorded. Can we conclude that students write in such a way as to fill the allotted space?

13.148 Xr13-148 Approximately 20 million Americans work for themselves. Most run single-person businesses out of their homes. One-quarter of these individuals use personal computers in their businesses. A market research firm, Computer Intelligence InfoCorp, wanted to know whether single-person businesses that use personal computers are more successful than those with no computer. They surveyed 150 single-person firms and recorded their annual incomes. Can we infer at the 10% significance level that single-person businesses that use a personal computer earn more than those that do not?

13.149 Xr13-149 Many small retailers advertise in their neighborhoods by sending out flyers. People deliver these to homes and are paid according to the number of flyers delivered. Each deliverer is given several streets whose homes become their responsibility. One of the ways retailers use to check the performance of deliverers is to randomly sample some of the homes and ask the home owner whether he or she received the flyer. Recently, university students started a new delivery service. They have promised better service at a competitive price. A retailer wanted to know whether the new company’s delivery rate is better than that of the existing firm. She had both companies deliver her flyers. Random samples of homes were drawn, and each was asked whether he or she received the flyer (2 = yes and 1 = no). Can the retailer conclude that the new company is better? (Test with $\alpha = .10$.)

13.150 Xr13-150 Medical experts advocate the use of vitamin and mineral supplements to help fight infections. A study undertaken by researchers at Memorial University (reported in the British journal *Lancet*, November 1992) recruited 96 men and women age 65 and older. One-half of them received daily supplements of vitamins and minerals, whereas the other half received placebos. The supplements contained the daily recommended amounts of 18 vitamins and minerals, including vitamins B-6, B-12, C, and D, as well as thiamine, riboflavin, niacin, calcium, copper, iodine, iron, selenium, magnesium, and zinc. The doses of vitamins A and E were slightly less than the daily requirements. The supplements included four times the amount of beta-carotene than the average person ingests daily. The number of days of illness from infections (ranging from colds to pneumonia) was recorded for each person. Can we infer that taking vitamin and mineral supplements daily increases the body's immune system?

13.151 Xr13-151 An inspector for the Atlantic City Gaming Commission suspects that a particular blackjack dealer may be cheating (in favor of the casino) when he deals at expensive tables. To test her belief, she observed 500 hands each at the $100-limit table and the $3,000-limit table. For each hand, she recorded whether the dealer won (code = 2) or lost (code = 1). When a tie occurs, there is no winner or loser. Can the inspector conclude at the 10% significance level that the dealer is cheating at the more expensive table?

13.152 Xr13-152 In 2005 Larry Summers, then president of Harvard University, received an avalanche of criticism for his attempt to explain why there are more male professors than female professors in mathematics. He suggested that there were innate differences that might permanently thwart the search for a more perfect gender balance. In an attempt to refute Dr. Summers's hypothesis, several researchers conducted large-scale mathematics tests of male and female students. Suppose the results were recorded. Conduct whatever tests you deem necessary to draw conclusions from these data. (*Note:* The data are simulated but represent actual results.)

Exercises 13.153 and 13.154 require access to the data files introduced in previous exercises.

13.153 Xr12-31* Exercise 12.31 dealt with the amount of time high school students spend per week at part-time jobs. In addition to the hours of part-time work, the school guidance counselor recorded the gender of the student surveyed (1 = female and 2 = male). Can we conclude that female and male high school students differ in the amount of time spent at part-time jobs?

13.154 Xm12-01* The company that organized the survey to determine the amount of discarded newspaper (Example 12.1) kept track of the type of neighborhood (1 = city and 2 = suburbs). Do these data allow the company management to infer that city households discard more newspaper than do suburban households?

APPLICATIONS IN **MARKETING**

© Picturenet/Blend Images/ Jupiterimages

Market Segmentation

In Section 12.4, we introduced market segmentation. The following exercises address the problem of determining whether two market segments differ in their pattern of purchases of a particular product or service.

13.155 Xr13-155 Movie studios segment their markets by age. Two segments that are particularly important to this industry are teenagers and 20-to-30-year-olds. To assess markets and guide the making of movies, a random sample of teenagers and 20-to-30-year-olds was drawn. All were asked to report the number of movies they saw in theaters last year. Do these data allow us to infer that teenagers see more movies than 20-to-30-year-olds?

The following exercises employ data files associated with examples and exercises seen previously in this book.

13.156 Xr12.125* In addition to asking about educational attainment, the survey conducted in Exercise 12.125 also asked whether the respondent had plans in the next 2 years to take a course (1 = no and 2 = yes). Can we conclude that Californians who did not complete high school are less likely to take a course in the university's evening program?

13.157 Xm12-06* The objective in the survey conducted in Example 12.6 was to estimate the size of the market segment of adults who are concerned about eating healthy foods. As part of the survey, each respondent was asked how much they

(Continued)

spend on breakfast cereal in an average month. The marketing manager of a company that produces several breakfast cereals would like to know whether on average the market segment concerned about eating health foods outspends the other market segments. Write a brief report detailing your findings.

13.158 Xr12-35* In Exercise 12.35, we described how the office equipment chain OfficeMax offers rebates on some products. The goal in that exercise was to estimate the total amount spent by customers who bought the package of 100 CD-ROMs. In addition to tracking these amounts, an executive also determined the amounts spent in the store by another sample of customers who purchased a fax machine/copier (regular price $89.99 minus $40 manufacturer's rebate and $10 OfficeMax mail-in rebate). Can OfficeMax conclude that those who buy the fax/copier outspend those who buy the package of CD-ROMs? Write a brief memo to the executives of OfficeMax describing your findings and any possible recommendations.

13.159 Xr12-91* In addition to recording whether faculty members who are between 55 and 64 plan to retire before they reach 65 in Exercise 12.91, the consultant asked each to report his or her annual salary. Can the president infer that professors aged 55 to 64 who plan to retire early have higher salaries than those who don't plan to retire early?

13.160 Xr12-96* In Exercise 12.96, the statistics practitioner also recorded the gender of the respondents where 1 = female and 2 = male. Can we infer that men and women differ in their choices of Christmas trees?

APPENDIX 13 / REVIEW OF CHAPTERS 12 AND 13

As you may have already discovered, the ability to identify the correct statistical technique is critical; any calculation performed without it is useless. When you solved problems at the end of each section in the preceding chapters (you *have* been solving problems at the end of each section covered, haven't you?), you probably had no great difficulty identifying the correct technique to use. You used the statistical technique introduced in that section. Although those exercises provided practice in setting up hypotheses, producing computer output of tests of hypothesis and confidence interval estimators, and interpreting the results, you did not address a fundamental question faced by statistics practitioners: Which technique should I use? If you still do not appreciate the dimension of this problem, examine Table A13.1, which lists all the inferential methods covered thus far.

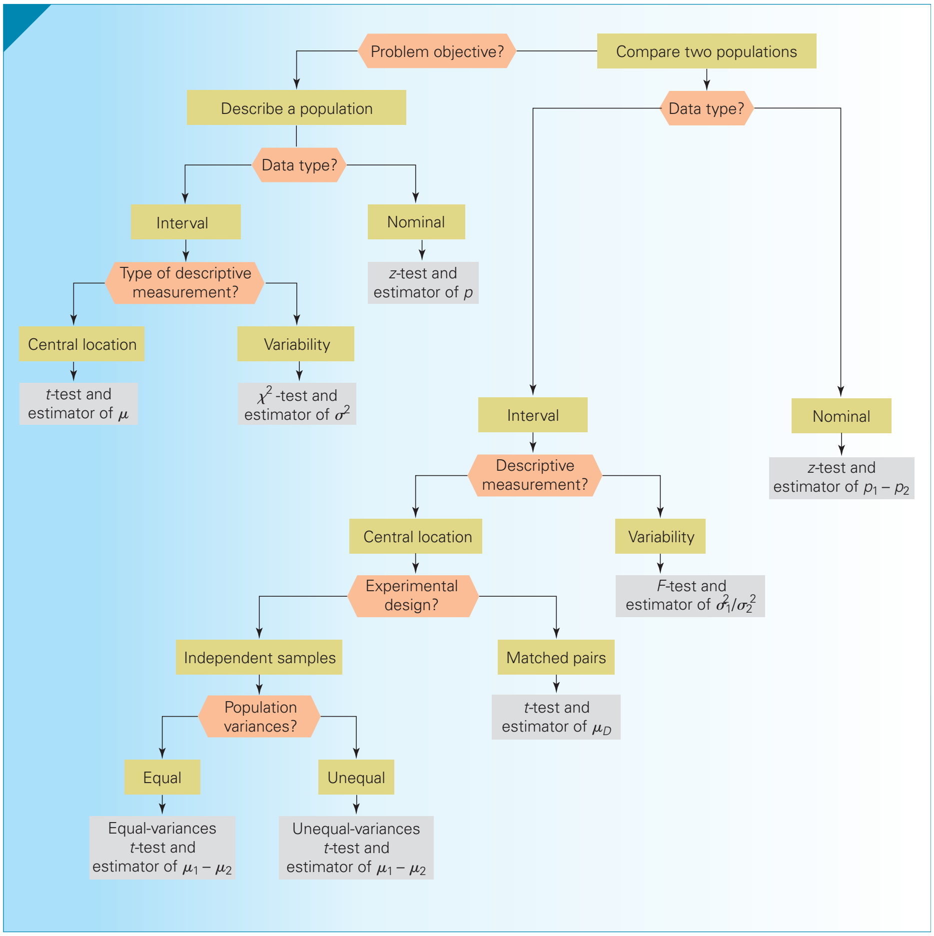
TABLE **A13.1** Summary of Statistical Techniques in Chapters 12 and 13

t-test of μ
Estimator of μ (including estimator of $N\mu$)
z-test of p
Estimator of p (including estimator of Np)
χ^2-test of σ^2
Estimator of σ^2
Equal-variances t-test of $\mu_1 - \mu_2$
Equal-variances estimator of $\mu_1 - \mu_2$
Unequal-variances t-test of $\mu_1 - \mu_2$
Unequal-variances estimator of $\mu_1 - \mu_2$
t-test of μ_D
Estimator of μ_D
F-test of σ_1^2/σ_2^2
Estimator of σ_1^2/σ_2^2
z-test of $p_1 - p_2$ (Case 1)
z-test of $p_1 - p_2$ (Case 2)
Estimator of $p_1 - p_2$

Counting tests and confidence interval estimators of a parameter as two different techniques, a total of 17 statistical procedures have been presented thus far, and there is much left to be done. Faced with statistical problems that require the use of some of these techniques (such as in real-world applications or on a quiz or midterm test), most students need some assistance in identifying the appropriate method. In this appendix and the appendixes of five more chapters, you will have the opportunity to practice your decision skills; we've provided exercises and cases that require all the inferential techniques introduced in Chapters 12 and 13. Solving these problems will require you to do what statistics practitioners must do: analyze the problem, identify the technique or techniques, employ statistical software and a computer to yield the required statistics, and interpret the results.

The flowchart in Figure A13.1 represents the logical process that leads to the identification of the appropriate method. Of course, it only shows the techniques covered to this point. Chapters 14, 15, 16, 17, and 19 will include appendixes that review all the techniques introduced up to that chapter. The list and the flowchart will be expanded in each appendix, and all appendixes will contain review exercises. (Some will contain cases.)

FIGURE **A13.1** **Flowchart of Techniques in Chapters 12 and 13**



As we pointed out in Chapter 11, the two most important factors in determining the correct statistical technique are the problem objective and the data type. In some situations, once these have been recognized, the technique automatically follows. In other cases, however, several additional factors must be identified before you can proceed. For example, when the problem objective is to compare two populations and the data are interval, three other significant issues must be addressed: the descriptive measurement (central location or variability), whether the samples are independently drawn, and, if so, whether the unknown population variances are equal.

EXERCISES

The purpose of the exercises that follow is twofold. First, the exercises provide you with practice in the critical skill of identifying the correct technique. Second, they allow you to improve your ability to determine the statistics needed to answer the question and interpret the results. We believe that the first skill is underdeveloped because up to now you have had little practice. The exercises you've worked on have appeared at the end of sections and chapters where the correct techniques have just been presented. Determining the correct technique should not have been difficult. Because the exercises that follow were selected from the types that you have already encountered in Chapters 12 and 13, they will help you develop your technique-identification skills.

You will note that in the exercises that require a test of hypothesis, we do not specify a significance level. We have left this decision to you. After analyzing the issues raised in the exercise, use your own judgment to determine whether the p-value is small enough to reject the null hypothesis.

A13.1 XrA13-01 Shopping malls are more than places where we buy things. We go to malls to watch movies; buy breakfast, lunch, and dinner; exercise; meet friends; and, in general, to socialize. To study the trends, a sociologist took a random sample of 100 mall shoppers and asked a variety of questions. This survey was first conducted 3 years ago with another sample of 100 shoppers. In both surveys, respondents were asked to report the number of hours they spend in malls during an average week. Can we conclude that the amount of time spent at malls has decreased over the past 3 years?

A13.2 XrA13-02 It is often useful for retailers to determine why their potential customers choose to visit their store. Possible reasons include advertising, advice from a friend, or previous experience. To determine the effect of full-page advertisements in the local newspaper, the owner of an electronic-equipment store asked 200 randomly selected people who visited the store whether they had seen the ad. He also determined whether the customers had bought anything, and, if so, how much they spent. There were 113 respondents who saw the ad. Of these, 49 made a purchase. Of the 87 respondents who did not see the ad, 21 made a purchase. The amounts spent were recorded.

a. Can the owner conclude that customers who see the ad are more likely to make a purchase than those who do not see the ad?
b. Can the owner conclude that customers who see the ad spend more than those who do not see the ad (among those who make a purchase)?
c. Estimate with 95% confidence the proportion of all customers who see the ad who then make a purchase.
d. Estimate with 95% confidence the mean amount spent by customers who see the ad and make a purchase.

A13.3 XrA13-03 In an attempt to reduce the number of person-hours lost as a result of industrial accidents, a large multiplant corporation installed new safety equipment in all departments and all plants. To test the effectiveness of the equipment, a random sample of 25 plants was drawn. The number of person-hours lost in the month before installation of the safety equipment and in the month after installation was recorded. Can we conclude that the equipment is effective?

A13.4 XrA13-04 Is the antilock braking system (ABS) now available as a standard feature on many cars really effective? The ABS works by automatically pumping brakes extremely quickly on slippery surfaces so the brakes do not lock and thus avoiding an uncontrollable skid. If ABS is effective, we would expect that cars equipped with ABS would have fewer accidents, and the costs of repairs for the accidents that do occur would be smaller. To investigate the effectiveness of ABS, the Highway Loss Data Institute gathered data on a random sample of 500 General Motors cars that did not have ABS and 500 GM cars that were equipped with ABS. For each year, the institute recorded whether the car was involved in an accident and, if so, the cost of making repairs. Forty-two cars without ABS and 38 ABS-equipped cars were involved in accidents. The costs of repairs were recorded. Using frequency of accidents and cost of repairs as measures of effectiveness, can we conclude that ABS is effective? If so, estimate how much better are cars equipped with ABS compared to cars without ABS.

A13.5 XrA13-05 The electric company is considering an incentive plan to encourage its customers to pay their bills promptly. The plan is to discount the bills 1% if the customer pays within 5 days as opposed to the usual 25 days. As an experiment, 50 customers are offered the discount on their September bill. The amount of time each takes to pay his or her bill is recorded. The amount of time a random sample of 50 customers not offered the discount take to pay their bills is also recorded. Do these data allow us to infer that the discount plan works?

A13.6 XrA13-06 Traffic experts are always looking for ways to control automobile speeds. Some communities have experimented with "traffic-calming" techniques. These include speed bumps and various obstructions that force cars to slow down to drive around them. Critics point out that the techniques are counterproductive because they cause drivers to speed on other parts of these roads. In an analysis of the effectiveness of speed bumps, a statistics practitioner organized a study over a 1-mile stretch of city road that had 10 stop signs. He then took a random sample of 100 cars and recorded their average speed (the speed limit was 30 mph) and the number of proper stops at the stop signs. He repeated the observations for another sample of 100 cars after speed bumps were placed on the road. Do these data allow the statistics practitioner to conclude that the speed bumps are effective?

A13.7 XrA13-07 The proliferation of self-serve pumps at gas stations has generally resulted in poorer automobile maintenance. One feature of poor maintenance is low tire pressure, which results in shorter tire life and higher gasoline consumption. To examine this problem, an automotive expert took a random sample of cars across the country and measured the tire pressure. The difference between the recommended tire pressure and the observed tire pressure was recorded. [A recording of 8 means that the pressure of the tire is 8 pounds per square inch (psi) less than the amount recommended by the tire manufacturer.] Suppose that for each psi below recommendation, tire life decreases by 100 miles and gasoline consumption increases by 0.1 gallon per mile. Estimate with 95% confidence the effect on tire life and gasoline consumption.

A13.8 XrA13-08 Many North American cities encourage the use of bicycles as a way to reduce pollution and traffic congestion. So many people now regularly use bicycles to get to work and for exercise that some jurisdictions have enacted bicycle helmet laws that specify that all bicycle riders must wear helmets to protect against head injuries. Critics of these laws complain that it is a violation of individual freedom and that helmet laws tend to discourage bicycle usage. To examine this issue, a researcher randomly sampled 50 bicycle users and asked each to record the number of miles he or she rode weekly. Several weeks later, the helmet law was enacted. The number of miles each of the 50 bicycle riders rode weekly was recorded for the week after the law was passed. Can we infer from these data that the law discourages bicycle usage?

A13.9 XrA13-09 Cardizem CD is a prescription drug that is used to treat high blood pressure and angina. One common side effect of such drugs is the occurrence of headaches and dizziness. To determine whether its drug has the same side effects, the drug's manufacturer, Marion Merrell Dow, Inc., undertook a study. A random sample of 908 high-blood-pressure sufferers was recruited; 607 took Cardizem CD and 301 took a placebo. Each reported whether they suffered from headaches or dizziness (2 = yes, 1 = no). Can the pharmaceutical company scientist infer that Cardizem CD users are more likely to suffer headache and dizziness side effects than nonusers?

A13.10 XrA13-10 A fast-food franchiser is considering building a restaurant at a downtown location. Based on a financial analysis, a site is acceptable only if the number of pedestrians passing the location during the work day averages more than 200 per hour. To help decide whether to build on the site, a statistics practitioner observes the number of pedestrians who pass the site each hour over a 40-hour workweek. Should the franchiser build on this site?

A13.11 XrA13-11 Most people who quit smoking cigarettes do so for health reasons. However, some quitters find that they gain weight after quitting, and scientists estimate that the health risks of smoking two packs of cigarettes per day or carrying 65 extra pounds of weight are about equivalent. In an attempt to learn more about the effects of quitting smoking, the U.S. Centers for Disease Control conducted a study (reported in *Time*, March 25, 1991). A sample of 1,885 smokers was taken. During the course of the experiment, some of the smokers quit their habit. The amount of weight gained by all the subjects was recorded. Do these data allow us to conclude that quitting smoking results in weight gains?

A13.12 XrA13-12 Golf-equipment manufacturers compete against one another by offering a bewildering array of new products and innovations. Oversized clubs, square grooves, and graphite shafts are examples of such innovations. The effect of these new products on the average golfer is, however, much in doubt. One product, a perimeter-weighted iron, was designed to increase the consistency of distance and accuracy. The most important aspect of irons is consistency, which means that ideally there should be no variation in distance from shot to shot. To examine the relative merits of two brands of perimeter-weighted irons, an average golfer used the 7-iron, hitting 100 shots using each of two brands. The distance in yards was recorded. Can the golfer conclude that brand B is superior to brand A?

A13.13 XrA13-13 Managers are frequently called on to negotiate in a variety of settings. This calls for an ability to think logically, which requires an ability to concentrate and ignore distractions. In a study of the effect of distractions, a random sample of 208 students was drawn by psychologists at McMaster University (reported in the *National Post*, December 11, 2003). The male students were shown pictures of women of varying attractiveness. The female students were shown pictures of men of varying attractiveness. All students were then offered a choice of an immediate reward of $15 or a wait of 8 months for a reward of $75. The choices of the male and of the female students (1 = immediate reward, 2 = larger reward 8 months later) were recorded. The results are stored in the following way:

Column 1: Choices of males shown most attractive women

Column 2: Choices of males shown less attractive women

Column 3: Choices of females shown most attractive men

Column 4: Choices of females shown less attractive men

a. Can we infer that men's choices are affected by the attractiveness of women's pictures?
b. Can we infer that women's choices are affected by the attractiveness of men's pictures?

A13.14 XrA13-14 Throughout the day, many exercise shows appear on television. These usually feature attractive and fit men and women performing various exercises and urging viewers to duplicate the activity at home. Some viewers are exercisers. However, some people like to watch the shows without exercising (which explains why attractive people are used as demonstrators). Various companies sponsor the shows, and there are commercial breaks. One sponsor wanted to determine whether there are differences between exercisers and nonexercisers in terms of how well they remember the sponsor's name. A random sample of viewers was selected and called after the exercise show was over. Each was asked to report whether he or she exercised or only watched. They were also asked to name the sponsor's brand name (2 = yes, they could; 1 = no, they couldn't). Can the sponsor conclude that exercisers are more likely to remember the sponsor's brand name than those who only watch?

A13.15 XrA13-15 According to the latest census, the number of households in a large metropolitan area is 425,000. The home-delivery department of the local newspaper reports that 104,320 households receive daily home delivery. To increase home-delivery sales, the marketing department launches an expensive advertising campaign. A financial analyst tells the publisher that for the campaign to be successful, home-delivery sales must increase to more than 110,000 households. Anxious to see whether the campaign is working, the publisher authorizes a telephone survey of 400 households within 1 week of the beginning of the campaign and asks each household head whether he or she has the newspaper delivered. The responses were recorded where 2 = yes and 1 = no.

a. Do these data indicate that the campaign will increase home-delivery sales?
b. Do these data allow the publisher to conclude that the campaign will be successful?

A13.16 XrA13-16 The Scholastic Aptitude Test (SAT), which is organized by the Educational Testing Service (ETS), is important to high school students seeking admission to colleges and universities throughout the United States. A number of companies offer courses to prepare students for the SAT. The Stanley H. Kaplan Educational Center claims that its students gain, on average, more than 110 points by taking its course. ETS, however, insists that preparatory courses can improve a score by no more than 40 points. (The minimum and maximum scores of the SAT are 400 and 1,600, respectively.) Suppose a random sample of 40 students wrote the exam, then took the Kaplan preparatory course, and then took the exam again.

a. Do these data provide sufficient evidence to refute the ETS claim?
b. Do these data provide sufficient evidence to refute Kaplan's claim?

A13.17 XrA13-17 A potato chip manufacturer has contracted for the delivery of 15,000,000 kilograms of potatoes. The supplier agrees to deliver the potatoes in 15,000 equal truckloads. The manufacturer suspects that the supplier will attempt to cheat him. He has the weight of the first 50 truckloads recorded.

a. Can the manufacturer conclude from these data that the supplier is cheating him?
b. Estimate with 95% confidence the total weight of potatoes for all 15,000 truckloads.

General Social Survey Exercises

A13.18 GSS2008* Is there sufficient evidence to conclude that people who work for the government (WRKGOVT: 1 = Government, 2 = Private) work fewer hours (HRS)?

For each of the following variables, conduct a test to determine whether Democrats and Republicans (PARTY 1 = Democrat, 3 = Republican) differ in their correct answers to the following questions.

A13.19 GSS2008* Correct answers to ODDS1: A doctor tells a couple that there is one chance in four that their child will have an inherited disease. Does this mean that if the first child has the illness, the next three will not? 1 = Yes, 2 = No. Correct answer: No.

A13.20 GSS2008* Correct answers to ODDS2: A doctor tells a couple that there is one chance in four that their child will have an inherited disease. Does this mean that each of the couple's children will have the same risk of suffering the illness? 1 = Yes, 2 = No. Correct answer: Yes.

A13.21 GSS2008* Correct answers to HOTCORE: The center of the earth is very hot. 1 = True, 2 = False. Correct answer: True.

A13.22 GSS2008* Correct answers to EARTHSUN: Does Earth go around the Sun or does the Sun go around Earth? 1 = Earth around Sun, 2 = Sun around Earth. Correct answer: Earth around Sun.

A13.23 GSS2008* Estimate with 95% confidence Americans mean position on the following question: Should government reduce income differences between rich and poor (EQWLTH: 1 = government should reduce differences, 2, 3, 4, 5, 6, 7 = No government action)?.

A13.24 Estimate with 95% confidence the mean number of years with current employer (CUREMPYR).

A13.25 Estimate with 90% confidence the proportion of Americans whose income is at least $75,000 (INCOME06).

A13.26 GSS2006* GSS2008* Can we infer from the data that the proportion of Americans earning at least $75,000 is greater in 2008 than in 2006 (INCOME06)?

American National Election Survey Exercises

A13.27 ANES2008* Conduct a test to determine whether Democrats and Republicans (PARTY: 1 = Democrat and 2 = Republican) differ in their intention to vote (DEFINITE: 1 = Definitely will not vote, 2, 3, 4, 5, 6, 7, 8 9, 10 = Definitely will vote).

A13.28 ANES2008* Estimate with 99% confidence the mean amount of time in a typical day spent by American adults watching news on television, not including sports (TIME2).

A13.29 ANES2008* Conduct a test to determine whether Democrats and Republicans (PARTY: 1 = Democrat and 2 = Republican) differ in how much they thought about the upcoming election for president (THOUGHT: 1 = Quite a lot, 5 = Only a little).

A13.30 ANES2008* Estimate with 95% confidence the proportion of Americans earning at least $100,000.

14

© Kzenon/Shutterstock

ANALYSIS OF VARIANCE

14.1 *One-Way Analysis of Variance*

14.2 *Multiple Comparisons*

14.3 *Analysis of Variance Experimental Designs*

14.4 *Randomized Block (Two-Way) Analysis of Variance*

14.5 *Two-Factor Analysis of Variance*

14.6 *(Optional) Applications in Operations Management: Finding and Reducing Variation*

Appendix 14 *Review of Chapters 12 to 14*

General Social Survey: Liberal–Conservative Spectrum and Income

DATA
GSS2008*

Are Americans' political views affected by their incomes, or perhaps vice versa? If so, we would expect that incomes would differ between groups who define themselves somewhere on the following scale (POLVIEW).

1 = Extremely liberal
2 = Liberal
3 = Slightly liberal
4 = Moderate
5 = Slightly conservative
6 = Conservative
7 = Extremely conservative

© AP Photo/Chris Carlson

The question to be answered (on page 537) is, Are there differences in income between the seven groups of political views?

INTRODUCTION

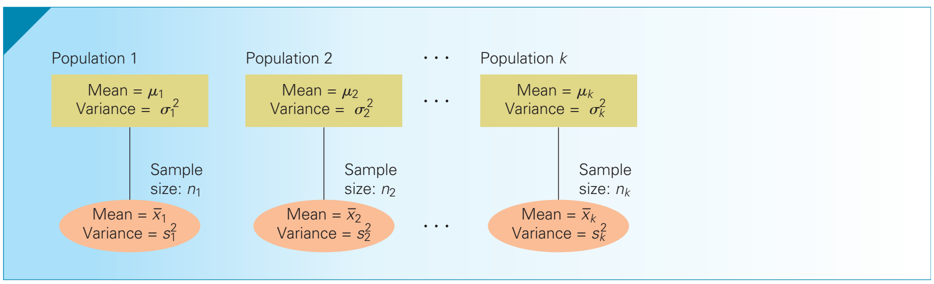
The technique presented in this chapter allows statistics practitioners to compare two or more populations of interval data. The technique is called the **analysis of variance**, and it is an extremely powerful and commonly used procedure. The analysis of variance technique determines whether differences exist between population means. Ironically, the procedure works by analyzing the sample variance, hence the name. We will examine several different forms of the technique.

One of the first applications of the analysis of variance was conducted in the 1920s to determine whether different treatments of fertilizer produced different crop yields. The terminology of that original experiment is still used. No matter what the experiment, the procedure is designed to determine whether there are significant differences between the **treatment means**.

14.1 ONE-WAY ANALYSIS OF VARIANCE

The analysis of variance is a procedure that tests to determine whether differences exist between two or more population means. The name of the technique derives from the way in which the calculations are performed; that is, the technique analyzes the variance of the data to determine whether we can infer that the population means differ. As in Chapter 13, the experimental design is a determinant in identifying the proper method to use. In this section, we describe the procedure to apply when the samples are independently drawn. The technique is called the **one-way analysis of variance**. Figure 14.1 depicts the sampling process for drawing independent samples. The mean and variance of population j ($j = 1, 2, \ldots, k$) are labeled μ_j and σ_j^2, respectively. Both parameters are unknown. For each population, we draw independent random samples. For each sample, we can compute the mean $\bar{x}_j$ and the variance s_j^2.

FIGURE **14.1** Sampling Scheme for Independent Samples



EXAMPLE 14.1*

DATA
Xm14-01

Proportion of Total Assets Invested in Stocks

In the last decade, stockbrokers have drastically changed the way they do business. Internet trading has become quite common, and online trades can cost as little as $7. It is now easier and cheaper to invest in the stock market than ever before. What are the effects of these changes? To help answer this question, a financial analyst randomly sampled 366 American households and asked each to report the age category of the head of

*Adapted from U.S. Census Bureau, "Asset Ownership of Households, May 2003," *Statistical Abstract of the United States, 2006*, Table 700.

the household and the proportion of its financial assets that are invested in the stock market. The age categories are

Young (less than 35)

Early middle age (35 to 49)

Late middle age (50 to 65)

Senior (older than 65)

The analyst was particularly interested in determining whether the ownership of stocks varied by age. Some of the data are listed next. Do these data allow the analyst to determine that there are differences in stock ownership between the four age groups?

Young	Early Middle Age	Late Middle Age	Senior
24.8	28.9	81.5	66.8
35.5	7.3	0.0	77.4
68.7	61.8	61.3	32.9
42.2	53.6	0.0	74.0
⋮	⋮	⋮	⋮

SOLUTION

You should confirm that the data are interval (percentage of total assets invested in the stock market) and that the problem objective is to compare four populations (age categories). The parameters are the four population means: μ_1, μ_2, μ_3, and μ_4. The null hypothesis will state that there are no differences between the population means. Hence,

$$H_0\text{:} \quad \mu_1 = \mu_2 = \mu_3 = \mu_4$$

The analysis of variance determines whether there is enough statistical evidence to show that the null hypothesis is false. Consequently, the alternative hypothesis will always specify the following:

$$H_1\text{:} \quad \text{At least two means differ}$$

The next step is to determine the test statistic, which is somewhat more involved than the test statistics we have introduced thus far. The process of performing the analysis of variance is facilitated by the notation in Table 14.1.

TABLE 14.1 Notation for the One-Way Analysis of Variance

	TREATMENT					
	1	2		j		k
	x_{11}	x_{12}	$\cdots$	x_{1j}	$\cdots$	x_{1k}
	x_{21}	x_{22}	$\cdots$	x_{2j}	$\cdots$	x_{2k}
	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	$x_{n_1 1}$	$x_{n_2 2}$		$x_{n_j j}$		$x_{n_k k}$
Sample size	n_1	n_2		n_j		n_k
Sample mean	$\bar{x}_1$	$\bar{x}_2$		$\bar{x}_j$		$\bar{x}_k$

x_{ij} = ith observation of the jth sample

n_j = number of observations in the sample taken from the jth population

$\bar{x}_j$ = mean of the jth sample = $\dfrac{\sum_{i=1}^{n_j} x_{ij}}{n_j}$

$\bar{\bar{x}}$ = grand mean of all the observations = $\dfrac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} x_{ij}}{n}$ where $n = n_1 + n_2 + \cdots + n_k$, and k is the number of populations

The variable X is called the **response variable**, and its values are called **responses**. The unit that we measure is called an **experimental unit**. In this example, the response variable is the percentage of assets invested in stocks, and the experimental units are the heads of households sampled. The criterion by which we classify the populations is called a **factor**. Each population is called a factor **level**. The factor in Example 14.1 is the age category of the head of the household and there are four levels. Later in this chapter, we'll discuss an experiment where the populations are classified using two factors. In this section, we deal with single-factor experiments only.

Test Statistic

The test statistic is computed in accordance with the following rationale. If the null hypothesis is true, the population means would all be equal. We would then expect that the sample means would be close to one another. If the alternative hypothesis is true, however, there would be large differences between some of the sample means. The statistic that measures the proximity of the sample means to each other is called the **between-treatments variation**; it is denoted **SST**, which stands for **sum of squares for treatments**.

Sum of Squares for Treatments

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

As you can deduce from this formula, if the sample means are close to each other, all of the sample means would be close to the grand mean; as a result, SST would be small. In fact, SST achieves its smallest value (zero) when all the sample means are equal. In other words, if

$$\bar{x}_1 = \bar{x}_2 = \cdots = \bar{x}_k$$

then

$$\text{SST} = 0$$

It follows that a small value of SST supports the null hypothesis. In this example, we compute the sample means and the grand mean as

$$\bar{x}_1 = 44.40$$
$$\bar{x}_2 = 52.47$$
$$\bar{x}_3 = 51.14$$
$$\bar{x}_4 = 51.84$$
$$\bar{\bar{x}} = 50.18$$

The sample sizes are

$$n_1 = 84$$
$$n_2 = 131$$
$$n_3 = 93$$
$$n_4 = 58$$
$$n = n_1 + n_2 + n_3 + n_4 = 84 + 131 + 93 + 58 = 366$$

Then

$$\text{SST} = \sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2$$
$$= 84(44.40 - 50.18)^2 + 131(52.47 - 50.18)^2$$
$$+ 93(51.14 - 50.18)^2 + 58(51.84 - 50.18)^2$$
$$= 3{,}738.8$$

If large differences exist between the sample means, at least some sample means differ considerably from the grand mean, producing a large value of SST. It is then reasonable to reject the null hypothesis in favor of the alternative hypothesis. The key question to be answered in this test (as in all other statistical tests) is, How large does the statistic have to be for us to justify rejecting the null hypothesis? In our example, SST = 3,738.8. Is this value large enough to indicate that the population means differ? To answer this question, we need to know how much variation exists in the percentage of assets, which is measured by the **within-treatments variation**, which is denoted by **SSE** (**sum of squares for error**). The within-treatments variation provides a measure of the amount of variation in the response variable that is not caused by the treatments. In this example, we are trying to determine whether the percentages of total assets invested in stocks vary by the age of the head of the household. However, other variables also affect the responses variable. We would expect that variables such as household income, occupation, and the size of the family would play a role in determining how much money families invest in stocks. All of these (as well as others we may not even be able to identify) are sources of variation, which we would group together and call the error. This source of variation is measured by the sum of squares for error.

Sum of Squares for Error

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

When SSE is partially expanded, we get

$$\text{SSE} = \sum_{i=1}^{n_1}(x_{i1} - \bar{x}_1)^2 + \sum_{i=1}^{n_2}(x_{i2} - \bar{x}_2)^2 + \cdots + \sum_{i=1}^{n_k}(x_{ik} - \bar{x}_k)^2$$

If you examine each of the k components of SSE, you'll see that each is a measure of the variability of that sample. If we divide each component by $n_j - 1$, we obtain the sample variances. We can express this by rewriting SSE as

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

where s_j^2 is the sample variance of sample j. SSE is thus the combined or pooled variation of the k samples. This is an extension of a calculation we made in Section 13.1, where we tested and estimated the difference between two means using the pooled estimate of the common population variance (denoted s_p^2). One of the required conditions for that statistical technique is that the population variances are equal. That same condition is now necessary for us to use SSE; that is, we require that

$$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

Returning to our example, we calculate the sample variances as follows:

$$s_1^2 = 386.55$$
$$s_2^2 = 469.44$$
$$s_3^2 = 471.82$$
$$s_4^2 = 444.79$$

Thus,

$$\begin{aligned}\text{SSE} &= (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2 + (n_4 - 1)s_4^2 \\ &= (84 - 1)(386.55) + (131 - 1)(469.44) \\ &\quad + (93 - 1)(471.82) + (58 - 1)(444.79) \\ &= 161{,}871.3\end{aligned}$$

The next step is to compute quantities called the **mean squares**. The **mean square for treatments** is computed by dividing SST by the number of treatments minus 1.

Mean Square for Treatments

$$\text{MST} = \frac{\text{SST}}{k - 1}$$

The **mean square for error** is determined by dividing SSE by the total sample size (labeled n) minus the number of treatments.

Mean Square for Error

$$\text{MSE} = \frac{\text{SSE}}{n - k}$$

Finally, the test statistic is defined as the ratio of the two mean squares.

Test Statistic

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

Sampling Distribution of the Test Statistic

The test statistic is F-distributed with $k - 1$ and $n - k$ degrees of freedom, provided that the response variable is normally distributed. In Section 8.4, we introduced the F-distribution, and in Section 13.4 we used it to test and estimate the ratio of two population variances. The test statistic in that application was the ratio of two sample variances s_1^2 and s_2^2. If you examine the definitions of SST and SSE, you will see that both measure variation similar to the numerator in the formula used to calculate the sample variance s^2 used throughout this book. When we divide SST by $k - 1$ and SSE by $n - k$ to calculate MST and MSE, respectively, we're actually computing unbiased estimators of the common population variance, assuming (as we do) that the null hypothesis is true. Thus, the ratio F = MST/MSE is the ratio of two sample variances. The degrees of freedom for this application are the denominators in the mean squares; that is, $\nu_1 = k - 1$ and $\nu_2 = n - k$. For Example 14.1, the degrees of freedom are

$$\nu_1 = k - 1 = 4 - 1 = 3$$
$$\nu_2 = n - k = 366 - 4 = 362$$

In our example, we found

$$\text{MST} = \frac{\text{SST}}{k-1} = \frac{3{,}738.8}{3} = 1{,}246.27$$
$$\text{MSE} = \frac{\text{SSE}}{n-k} = \frac{161{,}871.3}{362} = 447.16$$
$$F = \frac{\text{MST}}{\text{MSE}} = \frac{1{,}246.27}{447.16} = 2.79$$

Rejection Region and *p*-Value

The purpose of calculating the ***F*-statistic** is to determine whether the value of SST is large enough to reject the null hypothesis. As you can see, if SST is large, F will be large. Hence, we reject the null hypothesis only if

$$F > F_{\alpha,k-1,n-k}$$

If we let $\alpha = .05$, the rejection region for Example 14.1 is

$$F > F_{\alpha,k-1,n-k} = F_{.05,3,362} \approx F_{.05,3,\infty} = 2.61$$

We found the value of the test statistic to be $F = 2.79$. Thus, there is enough evidence to infer that the mean percentage of total assets invested in the stock market differs between the four age groups.

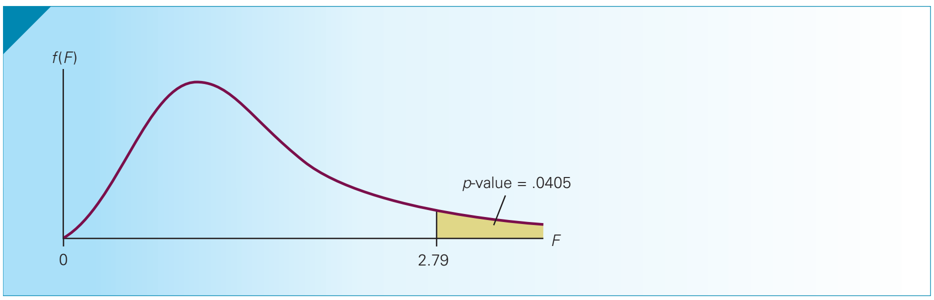
The p-value of this test is

$$P(F > 2.79)$$

A computer is required to calculate this value, which is .0405.

Figure 14.2 depicts the sampling distribution for Example 14.1.

FIGURE **14.2** Sampling Distribution for Example 14.1



The results of the analysis of variance are usually reported in an **analysis of variance (ANOVA) table**. Table 14.2 shows the general organization of the ANOVA table, and Table 14.3 shows the ANOVA table for Example 14.1.

TABLE **14.2** ANOVA Table for the One-Way Analysis of Variance

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-STATISTIC
Treatments	$k - 1$	SST	MST = SST/(k − 1)	F = MST/MSE
Error	$n - k$	SSE	MSE = SSE/(n − k)	
Total	$n - 1$	SS(Total)		

TABLE **14.3** ANOVA Table for Example 14.1

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-STATISTIC
Treatments	3	3,738.8	1,246.27	2.79
Error	362	161,871.3	447.16	
Total	365	165,610.1		

The terminology used in the ANOVA table (and for that matter, in the test itself) is based on the partitioning of the sum of squares. Such partitioning is derived from the following equation (whose validity can be demonstrated by using the rules of summation):

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

The term on the left represents the **total variation** of all the data. This expression is denoted **SS(Total)**. If we divide SS(Total) by the total sample size minus 1 (that is, by $n - 1$), we would obtain the sample variance (assuming that the null hypothesis is true). The first term on the right of the equal sign is SST, and the second term is SSE. As you can see, the total variation SS(Total) is partitioned into two sources of variation. The sum of squares for treatments (SST) is the variation attributed to the differences between the treatment means, whereas the sum of squares for error (SSE) measures the variation within the samples. The preceding equation can be restated as

$$\text{SS(Total)} = \text{SST} + \text{SSE}$$

The test is then based on the comparison of the mean squares of SST and SSE.

Recall that in discussing the advantages and disadvantages of the matched pairs experiment in Section 13.3, we pointed out that statistics practitioners frequently seek ways to reduce or explain the variation in a random variable. In the analysis of variance introduced in this section, the sum of squares for treatments explains the variation attributed to the treatments (age categories). The sum of squares for error measures the amount of variation that is unexplained by the different treatments. If SST explains a significant portion of the total variation, we conclude that the population means differ. In Sections 14.4 and 14.5, we will introduce other experimental designs of the analysis of variance—designs that attempt to reduce or explain even more of the variation.

If you've felt some appreciation of the computer and statistical software sparing you the need to manually perform the statistical techniques in earlier chapters, your appreciation should now grow, because the computer will allow you to avoid the incredibly time-consuming and boring task of performing the analysis of variance by hand. As usual, we've solved Example 14.1 using Excel and Minitab, whose outputs are shown here.

COMPUTE

EXCEL

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	Young	84	3729.5	44.40	386.55		
6	Early Middle Age	131	6873.9	52.47	469.44		
7	Late Middle Age	93	4755.9	51.14	471.82		
8	Senior	58	3006.6	51.84	444.79		
9							
10							
11	ANOVA						
12	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
13	Between Groups	3741.4	3	1247.12	2.79	0.0405	2.6296
14	Within Groups	161871.0	362	447.16			
15							
16	Total	165612.3	365				

INSTRUCTIONS

1. Type or import the data into adjacent columns. (Open Xm14-01.)
2. Click **Data, Data Analysis,** and **Anova: Single Factor.**
3. Specify the **Input Range** (A1:D132) and a value for α (.05).

MINITAB

One-way ANOVA: Young, Early Middle Age, Late Middle Age, Senior

```
Source   DF      SS    MS     F      P
Factor    3    3741  1247  2.79  0.041
Error   362  161871   447
Total   365  165612

S = 21.15   R-Sq = 2.26%   R-Sq(adj) = 1.45%

Level              N   Mean  StDev
Young             84  44.40  19.66
Early Middle Age 131  52.47  21.67
Late Middle Age   93  51.14  21.72
Senior            58  51.84  21.09

                      Individual 95% CIs For Mean Based on Pooled StDev
Level                 +-------------+-------------+-------------+-------------
Young                 (------------*------------)
Early Middle Age                             (---------*---------)
Late Middle Age                         (-----------*-----------)
Senior                                 (--------------*--------------)
                      +-------------+-------------+-------------+-------------
                    40.0          45.0          50.0          55.0

Pooled StDev = 21.15
```

(Continued)

INSTRUCTIONS

If the data are unstacked:

1. Type or import the data. (Open Xm14-01.)
2. Click **Stat**, **ANOVA**, and **Oneway (Unstacked) . . . .**
3. In the **Responses (in separate columns)** box, type or select the variable names of the treatments (Young, Early Middle Age, Late Middle Age, Senior).

If the data are stacked:

1. Type or import the data in two columns.
2. Click **Stat**, **ANOVA**, and **Oneway . . . .**
3. Type the variable name of the response variable and the name of the factor variable.

INTERPRET

The value of the test statistic is $F = 2.79$, and its p-value is .0405, which means there is evidence to infer that the percentage of total assets invested in stocks are different in at least two of the age categories.

Note that in this example the data are observational. We cannot conduct a controlled experiment. To do so would require the financial analyst to randomly assign households to each of the four age groups, which is impossible.

Incidentally, when the data are obtained through a controlled experiment in the one-way analysis of variance, we call the experimental design the **completely randomized design** of the analysis of variance.

Checking the Required Conditions

The F-test of the analysis of variance requires that the random variable be normally distributed with equal variances. The normality requirement is easily checked graphically by producing the histograms for each sample. From the Excel histograms in Figure 14.3, we can see that there is no reason to believe that the requirement is not satisfied.

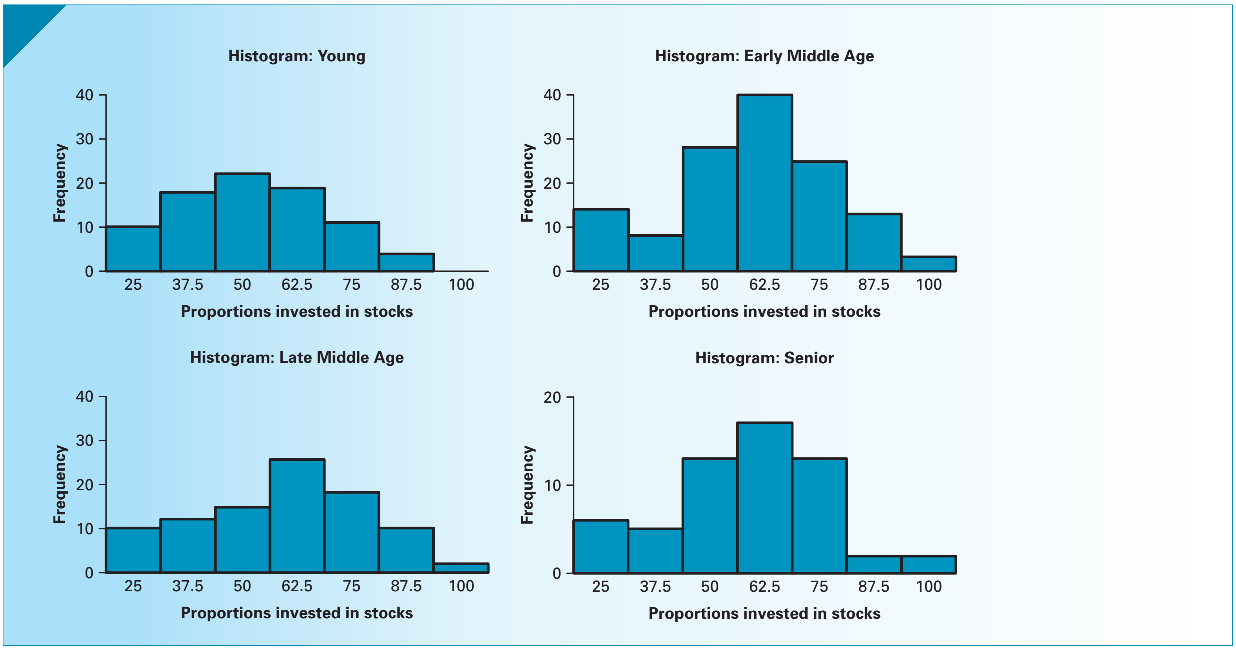
The equality of variances is examined by printing the sample standard deviations or variances. Excel output includes the variances, and Minitab calculates the standard deviations. The similarity of sample variances allows us to assume that the population variances are equal. In Keller's website Appendix Bartlett's Test, we present a statistical procedure designed to test for the equality of variances.

Violation of the Required Conditions

If the data are not normally distributed, we can replace the one-way analysis of variance with its nonparametric counterpart, which is the Kruskal–Wallis Test. (See Section 19.3.[†]) If the population variances are unequal, we can use several methods to correct the problem. However, these corrective measures are beyond the level of this book.

[†]Instructors who wish to teach the use of nonparametric techniques for testing the difference between two or more means when the normality requirement is not satisfied should use Keller's website Appendix Kruskal–Wallis Test and Friedman Test.

FIGURE **14.3** Histograms for Example 14.1



Can We Use the *t*-Test of the Difference between Two Means Instead of the Analysis of Variance?

The analysis of variance tests to determine whether there is evidence of differences between two or more population means. The *t*-test of $\mu_1 - \mu_2$ determines whether there is evidence of a difference between two population means. The question arises, Can we use *t*-tests instead of the analysis of variance? In other words, instead of testing all the means in one test as in the analysis of variance, why not test each pair of means? In Example 14.1, we would test $(\mu_1 - \mu_2)$, $(\mu_1 - \mu_3)$, $(\mu_1 - \mu_4)$, $(\mu_2 - \mu_3)$, $(\mu_2 - \mu_4)$, and $(\mu_3 - \mu_4)$. If we find no evidence of a difference in each test, we would conclude that none of the means differ. If there was evidence of a difference in at least one test, we would conclude that some of the means differ.

There are two reasons why we don't use multiple *t*-tests instead of one *F*-test. First, we would have to perform many more calculations. Even with a computer, this extra work is tedious. Second, and more important, conducting multiple tests increases the probability of making Type I errors. To understand why, consider a problem where we want to compare six populations, all of which are identical. If we conduct an analysis of variance where we set the significance level at 5%, there is a 5% chance that we would reject the true null hypothesis; that is, there is a 5% chance that we would conclude that differences exist when, in fact, they don't.

To replace the *F*-test, we would perform 15 *t*-tests. [This number is derived from the number of combinations of pairs of means to test, which is $C_2^6 = (6 \times 5)/2 = 15$.] Each test would have a 5% probability of erroneously rejecting the null hypothesis. The probability of committing one or more Type I errors is about 54%.[‡]

[‡]The probability of committing at least one Type I error is computed from a binomial distribution with $n = 15$ and $p = .05$. Thus, $P(X \geq 1) = 1 - P(X = 0) = 1 - .463 = .537$.

One remedy for this problem is to decrease the significance level. In this illustration, we would perform the t-tests with $\alpha = .05/15$, which is equal to .0033. (We will use this procedure in Section 14.2 when we discuss multiple comparisons.) Unfortunately, this would increase the probability of a Type II error. Regardless of the significance level, performing multiple t-tests increases the likelihood of making mistakes. Consequently, when we want to compare more than two populations of interval data, we use the analysis of variance.

Now that we've argued that the t-tests cannot replace the analysis of variance, we need to argue that the analysis of variance cannot replace the t-test.

Can We Use the Analysis of Variance Instead of the t-Test of $\mu_1 - \mu_2$?

The analysis of variance is the first of several techniques that allow us to compare two or more populations. Most of the examples and exercises deal with more than two populations. However, it should be noted that, like all other techniques whose objective is to compare two or more populations, the analysis of variance can be used to compare only two populations. If that's the case, then why do we need techniques to compare exactly two populations? Specifically, why do we need the t-test of $\mu_1 - \mu_2$ when the analysis of variance can be used to test two population means?

To understand why, we still need the t-test to make inferences about $\mu_1 - \mu_2$. Suppose that we plan to use the analysis of variance to test two population means. The null and alternative hypotheses are

H_0: $\mu_1 = \mu_2$

H_1: At least two means differ

Of course, the alternative hypothesis specifies that $\mu_1 \neq \mu_2$. However, if we want to determine whether μ_1 is greater than μ_2 (or vice versa), we cannot use the analysis of variance because this technique allows us to test for a difference only. Thus, if we want to test to determine whether one population mean exceeds the other, we must use the t-test of $\mu_1 - \mu_2$ (with $\sigma_1^2 = \sigma_2^2$). Moreover, the analysis of variance requires that the population variances are equal. If they are not, we must use the unequal variances test statistic.

Relationship between the F-Statistic and the t-Statistic

It is probably useful for you to understand the relationship between the t-statistic and the F-statistic. The test statistic for testing hypotheses about $\mu_1 - \mu_2$ with equal variances is

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

If we square this quantity, the result is the F-statistic: $F = t^2$. To illustrate this point, we'll redo the calculation of the test statistic in Example 13.1 using the analysis of variance. Recall that because we were able to assume that the population variances were equal, the test statistic was as follows:

$$t = \frac{(6.63 - 3.72) - 0}{\sqrt{40.42\left(\frac{1}{50} + \frac{1}{50}\right)}} = 2.29$$

Using the analysis of variance (the Excel output is shown here; Minitab's is similar), we find that the value of the test statistic is $F = 5.23$, which is $(2.29)^2$. Notice though that the analysis of variance p-value is .0243, which is twice the t-test p-value, which is .0122. The reason: The analysis of variance is conducting a test to determine whether the population means *differ*. If Example 13.1 had asked to determine whether the means differ, we would have conducted a two-tail test and the p-value would be .0243, the same as the analysis of variance p-value.

Excel Analysis of Variance Output for Example 13.1

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	Direct	50	331.6	6.63	37.49		
6	Broker	50	186.2	3.72	43.34		
7							
8							
9	ANOVA						
10	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
11	Between Groups	211.4	1	211.41	5.23	0.0243	3.9381
12	Within Groups	3960.5	98	40.41			
13							
14	Total	4172.0	99				

Developing an Understanding of Statistical Concepts

Conceptually and mathematically, the F-test of the independent samples' single-factor analysis of variance is an extension of the t-test of $\mu_1 - \mu_2$. Moreover, if we simply want to determine whether a difference between two means exists, we can use the analysis of variance. The advantage of using the analysis of variance is that we can partition the total sum of squares, which enables us to measure how much variation is attributable to differences between populations and how much variation is attributable to differences within populations. As we pointed out in Section 13.3, explaining the variation is an extremely important topic, one that we will see again in other experimental designs of the analysis of variance and in regression analysis (Chapters 16, 17, and 18).

General Social Survey: Liberal–Conservative Spectrum And Income


© AP Photo/Chris Carlson

IDENTIFY

The variable is income (INCOME) of American adults, which is interval. The problem objective is to compare seven populations (the political views) and the experimental design is independent samples. Thus, we apply the one-way analysis of variance.

COMPUTE

EXCEL

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	E Liberal	42	1,857,750	44,232	1,420,296,929		
6	Liberal	154	6,681,000	43,383	1,644,698,667		
7	S Liberal	152	5,613,500	36,931	1,003,760,097		
8	Moderate	442	16,946,750	38,341	1,290,463,769		
9	S Conservative	156	7,920,750	50,774	1,943,227,241		
10	Conservative	183	8,179,750	44,698	1,726,326,961		
11	E Conservative	32	1,947,750	60,867	2,886,806,389		
12							
13							
14	ANOVA						
15	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
16	Between Groups	34,931,882,213	6	5,821,980,369	3.87	0.0008	2.1064
17	Within Groups	1,735,416,094,316	1154	1,503,826,772			
18							
19	Total	1,770,347,976,529	1160				

MINITAB

One-way ANOVA: Income versus POLVIEWS

```
Source      DF           SS          MS     F      P
Polviews     6  34931882213  5821980369  3.87  0.001
Error     1154  1.73542E+12  1503826772
Total     1160  1.77035E+12

S = 38779  R-Sq = 1.97%    R-Sq(adj) = 1.46%


                           Individual 95% CIs For Mean Based on
                           Pooled StDev
Level      N   Mean   StDev  ---------+--------------+--------------+--------------+-----
1         42  44232   37687                      (-----------*-----------)
2        154  43383   40555                 (-----------*-----------)
3        152  36931   31682                (-----------*-----------)
4        442  38341   35923           (-----------*-----------)
5        156  50774   44082                  (-----------*-----------)
6        183  44698   41549                  (-----------*-----------)
7         32  60867   53729            (-----------*-----------)
                             ---------+--------------+--------------+--------------+-----
                                  36000          48000          60000          72000

Pooled StDev = 38779
```

INTERPRET

The p-value is .0008. There is sufficient evidence to infer that the incomes differ between the seven political views. It appears that conservatives have higher incomes than liberals.

Let's review how we recognize the need to use the techniques introduced in this section.

Factors That Identify the One-Way Analysis of Variance

1. **Problem objective**: Compare two or more populations
2. **Data type**: Interval
3. **Experimental design**: Independent samples

EXERCISES

Developing an Understanding of Statistical Concepts

Exercises 14.1–14.3 are "what-if" analyses designed to determine what happens to the test statistic when the means, variances, and sample sizes change. These problems can be solved manually or by creating an Excel worksheet.

14.1 A statistics practitioner calculated the following statistics:

	Treatment		
Statistic	1	2	3
n	5	5	5
$\bar{x}$	10	15	20
s^2	50	50	50

a. Complete the ANOVA table.
b. Repeat part (a) changing the sample sizes to 10 each.
c. Describe what happens to the F-statistic when the sample sizes increase.

14.2 You are given the following statistics:

	Treatment		
Statistic	1	2	3
n	4	4	4
$\bar{x}$	20	22	25
s^2	10	10	10

a. Complete the ANOVA table.
b. Repeat part (a) changing the variances to 25 each.
c. Describe the effect on the F-statistic of increasing the sample variances.

14.3 The following statistics were calculated:

	Treatment			
Statistic	1	2	3	4
n	10	14	11	18
$\bar{x}$	30	35	33	40
s^2	10	10	10	10

a. Complete the ANOVA table.
b. Repeat part (a) changing the sample means to 130, 135, 133, and 140.
c. Describe the effect on the F-statistic of increasing the sample means by 100.

Applications

14.4 Xr14-04 How does an MBA major affect the number of job offers received? An MBA student randomly sampled four recent graduates, one each in finance, marketing, and management, and asked them to report the number of job offers. Can we conclude at the 5% significance level that there are differences in the number of job offers between the three MBA majors?

Finance	Marketing	Management
3	1	8
1	5	5
4	3	4
1	4	6

14.5 Xr14-05 A consumer organization was concerned about the differences between the advertised sizes of containers and the actual amount of product. In a preliminary study, six packages of three different brands of margarine that are supposed to contain 500 ml were measured. The differences from 500 ml are listed here. Do these data provide sufficient evidence to conclude that differences exist between the three brands? Use $\alpha = .01$.

Brand 1	Brand 2	Brand 3
1	2	1
3	2	2
3	4	4
0	3	2
1	0	3
0	4	4

14.6 Xr14-06 Many college and university students obtain summer jobs. A statistics professor wanted to determine whether students in different degree programs earn different amounts. A random sample of 5 students in the BA, BSc, and BBA programs were asked to report what they earned the previous summer. The results (in $1,000s) are listed here. Can the professor infer at the 5% significance level that students in different degree programs differ in their summer earnings?

B.A.	B.Sc.	B.B.A.
3.3	3.9	4.0
2.5	5.1	6.2
4.6	3.9	6.3
5.4	6.2	5.9
3.9	4.8	6.4

14.7 Xr14-07 Spam is the price we pay for being able to easily communicate by e-mail. Does spam affect everyone equally? In a preliminary study, university professors, administrators, and students were randomly sampled. Each person was asked to count the number of spam messages received that day. The results follow. Can we infer at the 2.5% significance level that the differing university communities differ in the amount of spam they receive in their e-mails?

Professors	Administrators	Students
7	5	12
4	9	4
0	12	5
3	16	18
18	10	15

14.8 Xr14-08 A management scientist believes that one way of judging whether a computer came equipped with enough memory is to determine the age of the computer. In a preliminary study, random samples of computer users were asked to identify the brand of computer and its age (in months). The categorized responses are shown here. Do these data provide sufficient evidence to conclude that there are differences in age between the computer brands? (Use $\alpha = .05$.)

IBM	Dell	Hewlett-Packard	Other
17	8	6	24
10	4	15	12
13	21	8	15

Exercises 14.9–14.32 require the use of a computer and software. Use a 5% significance level unless specified otherwise. The answers to Exercises 14.9–14.20 may be calculated manually. See Appendix A for the sample statistics.

14.9 Xr14-09 Because there are no national or regional standards, it is difficult for university admission committees to compare graduates of different high schools. University administrators have noted that an 80% average at a high school with low standards may be equivalent to a 70% average at another school with higher standards of grading. In an effort to more equitably compare applications, a pilot study was initiated. Random samples of students who were admitted the previous year from four local high schools were drawn. All the students entered the business program with averages between 70% and 80%. Their average grades in the first year at the university were computed.

a. Can the university admissions officer conclude that there are differences in grading standards between the four high schools?
b. What are the required conditions for the test conducted in part (a)?
c. Does it appear that the required conditions of the test in part (a) are satisfied?

14.10 Xr14-10 The friendly folks at the Internal Revenue Service (IRS) in the United States and Canada Revenue Agency (CRA) are always looking for ways to improve the wording and format of its tax return forms. Three new forms have been developed recently. To determine which, if any, are superior to the current form, 120 individuals were asked to participate in an experiment. Each of the three new forms and the currently used form were filled out by 30 different people. The amount of time (in minutes) taken by each person to complete the task was recorded.

a. What conclusions can be drawn from these data?
b. What are the required conditions for the test conducted in part (a)?
c. Does it appear that the required conditions of the test in part (a) are satisfied?

14.11 Xr14-11 Are proficiency test scores affected by the education of the child's parents? (Proficiency tests are administered to a sample of students in private and public schools. Test scores can range from 0 to 500.) To answer this question, a random sample of 9-year-old children was drawn. Each child's test score and the educational level of the parent with the higher level were recorded. The education categories are less than high school, high school graduate, some college, and college graduate. Can we infer that there are differences in test scores between children whose parents have different educational levels? (Adapted from the *Statistical Abstract of the United States*, *2000*, Table 286.)

14.12 Xr14-12 A manufacturer of outdoor brass lamps and mailboxes has received numerous complaints about premature corrosion. The manufacturer has identified the cause of the problem as the low-quality lacquer used to coat the brass. He decides to replace his current lacquer supplier with one of five possible alternatives. To judge which is best, he uses each of the five lacquers to coat 25 brass mailboxes and puts all 125 mailboxes outside. He records, for each, the number of days until the first sign of corrosion is observed.

a. Is there sufficient evidence at the 1% significance level to allow the manufacturer to conclude that differences exist between the five lacquers?
b. What are the required conditions for the test conducted in part (a)?
c. Does it appear that the required conditions of the test in part (a) are satisfied?

14.13 Xr14-13 In early 2001, the economy was slowing down and companies were laying off workers. A Gallup poll asked a random sample of workers how long it would be before they had significant financial hardships if they lost their jobs and couldn't find new ones. They also classified their income. The classifications are

More than $50,000
$30,000 to $50,000
$20,000 to $30,000
Less than $20,000

Can we infer that differences exist between the four groups?

14.14 Xr14-14 In the introduction to this chapter, we mentioned that the first use of the analysis of variance was in the 1920s. It was employed to determine whether different amounts of fertilizer yielded different amounts of crop. Suppose that a scientist at an agricultural college wanted to redo the original experiment using three different types of fertilizer. Accordingly, she applied fertilizer A to 20 1-acre plots of land, fertilizer B to another 20 plots, and fertilizer C to yet another 20 plots of land. At the end of the growing season, the crop yields were recorded. Can the scientist infer that differences exist between the crop yields?

14.15 Xr14-15 A study performed by a Columbia University professor (described in *Report on Business*, August 1991) counted the number of times per minute professors from three different departments said "uh" or "ah" during lectures to fill gaps between words. The data derived from observing 100 minutes from each of the three departments were recorded. If we assume that the more frequent use of "uh" and "ah" results in more boring lectures, can we conclude that some departments' professors are more boring than others?

14.16 Xr14-16 Does the level of success of publicly traded companies affect the way their board members are paid? Publicly traded companies were divided into four quarters using the rate of return in their stocks to differentiate among the companies. The annual payment (in $1,000s) to their board members was recorded. Can we infer that the amount of payment differs between the four groups of companies?

14.17 Xr14-17 In 1994, the chief executive officers of the major tobacco companies testified before a U.S. Senate subcommittee. One of the accusations made was that tobacco firms added nicotine to their cigarettes, which made them even more addictive to smokers. Company scientists argued that the amount of nicotine in cigarettes depended completely on the size of the tobacco leaf: During poor growing seasons, the tobacco leaves would be smaller than in normal or good growing seasons. However, because the amount of nicotine in a leaf is a fixed quantity, smaller leaves would result in cigarettes having more nicotine (because a greater fraction of the leaf would be used to make a cigarette). To examine the issue, a university chemist took random samples of tobacco leaves that were grown in greenhouses where the amount of water was allowed to vary. Three different groups of tobacco leaves were grown. Group 1 leaves were grown with about an average season's rainfall. Group 2 leaves were given about 67% of group 1's water, and group 3 leaves were given 33% of group 1's water. The size of the leaf (in grams) and the amount of nicotine in each leaf were measured.
a. Test to determine whether the leaf sizes differ between the three groups.
b. Test to determine whether the amounts of nicotine differ in the three groups.

14.18 Xr14-18 There is a bewildering number of breakfast cereals on the market. Each company produces several different products in the belief that there are distinct markets. For example, there is a market composed primarily of children, another for diet-conscious adults, and another for health-conscious adults. Each cereal the companies produce has at least one market as its target. However, consumers make their own decisions, which may or may not match the target predicted by the cereal maker. In an attempt to distinguish between consumers, a survey of adults between the ages of 25 and 65 was undertaken. Each was asked several questions, including age, income, and years of education, as well as which brand of cereal they consumed most frequently. The cereal choices are

1. Sugar Smacks, a children's cereal
2. Special K, a cereal aimed at dieters
3. Fiber One, a cereal that is designed and advertised as healthy
4. Cheerios, a combination of healthy and tasty

The results of the survey were recorded using the following format:

Column 1: Cereal choice
Column 2: Age of respondent
Column 3: Annual household income
Column 4: Years of education

a. Determine whether there are differences between the ages of the consumers of the four cereals.
b. Determine whether there are differences between the incomes of the consumers of the four cereals.
c. Determine whether there are differences between the educational levels of the consumers of the four cereals.
d. Summarize your findings in parts (a) through (c) and prepare a report describing the differences between the four groups of cereal consumers.

APPLICATIONS in MARKETING

© Susan Van Etten

Test Marketing

In Chapter 13, we introduced test marketing, which allows us to determine whether changing some of the elements of the marketing mix yields different sales. In the next exercise, we apply the technique to discover the effect of different prices.

14.19 Xr14-19 A manufacturer of novelty items is undecided about the price to charge for a new product. The marketing manager knows that it should sell for about $10 but is unsure of whether sales will vary significantly if it is priced at either $9 or $11. To conduct a pricing experiment, she distributes the new product to a sample of 60 stores belonging to a certain chain of variety stores. These 60 stores are all located in similar neighborhoods. The manager randomly selects 20 stores in which to sell the item at $9, 20 stores to sell it at $10, and the remaining 20 stores to sell it at $11. Sales at the end of the trial period were recorded. What should the manager conclude?

APPLICATIONS in MARKETING

© John Eder/Getty Images

Marketing Segmentation

Section 12.4 introduced market segmentation. In Chapter 13 we demonstrated how to use statistical analyses to determine whether two segments differ in their buying behavior. The next exercise requires you to apply the analysis of variance to determine whether several segments differ.

14.20 Xr14-20 After determining in Exercise 13.155 that teenagers watch more movies than do 20 to 30 year olds, teenagers were further segmented into three age groups: 12 to 14, 15 to 16, and 17 to 19. Random samples were drawn from each segment, and the number of movies each teenager saw last year was recorded. Do these data allow a marketing manager of a movie studio to conclude that differences exist between the three segments?

GENERAL SOCIAL SURVEY EXERCISES

14.21 GSS2002* GSS2004* GSS2006* GSS2008* Have educational levels kept uniform over the years 2002, 2004, 2006, and 2008? Conduct a test to determine whether the number of years of education (EDUC) differ in the four years.

14.22 GSS2008* Television networks and their advertisers are constantly measuring viewers to determine their likes and dislikes and how much time adults spend watching television per day. Do the data from the General Social Survey in 2008 allow us to infer that the amount of television (TVHOURS) differs by race (RACE)?

14.23 GSS2008* How are income and degree related? The General Social Survey asked respondents to identify the highest degree completed (DEGREE: 0 = Left high school, 1 = High school, 2 = Junior college, 3 = Bachelor's degree, 4 = Graduate degree). Is there enough statistical evidence to conclude that there are differences in income (INCOME) between people with different completed degrees?

14.24 GSS2002* GSS2004* GSS2006* GSS2008* Has the amount of time Americans devote to work weekly (HRS) changed over the years 2002, 2004, 2006, and 2008? Perform a statistical analysis to answer the question.

14.25 GSS2008* Who earns more money: married people, single people, widows and widowers, or divorcees? Conduct an appropriate statistical technique to determine whether there is enough evidence to conclude that incomes (INCOME) vary by marital status (MARITAL).

14.26 GSS2002* GSS2004* GSS2006* GSS2008* Has the amount of television American adults watch been constant over the years 2002, 2004, 2006, and 2008, or has the amount varied? Test to determine whether the number of hours of television per day (TVHOURS) changed over the 8-year span.

American National Election Survey Exercises

14.27 ANES2008* Repeat the chapter-opening example using the data from the American National Election Survey of 2008. Test to determine whether annual incomes (INCOME) differ between the seven political views (LIBCON).

14.28 ANES2008* Who has the most and least education among Democrats, Independents, and Republicans? Conduct a statistical test to determine if there is evidence of a difference in education (EDUC) between the three political affiliations (PARTY3).

14.29 ANES2008* Who reads newspapers more: Democrats, Independents, or Republicans? Test to determine whether differences exist in the number of days reading a newspaper (DAYS9) between the three political affiliations (PARTY3).

14.30 ANES2008* How are income and degree related? The American National Election Survey asked respondents who reported at least 13 years of education to identify the highest degree completed (DEGREE: 0 = No degree earned; 1 = Bachelor's degree; 2 = Master's degree; 3 = PhD, etc.; 4 = LLB, JD; 5 = MD, DDS, etc.; 6 = JDC, STD, THD; 7 = Associate's degree). Is there enough statistical evidence to conclude that there are differences in income (INCOME) between people with different completed degrees?

14.31 ANES2008* Are marital status and education related? If so, we would expect that the amount of education in at least two of the marital status categories to be different. Conduct a statistical procedure to determine whether there is enough evidence to infer that the amount of education (EDUC) differs between marital status (MARITAL) categories.

14.32 ANES2008* How definite is one's intention to vote in the presidential election, and is that intention related to party affiliation? To answer this question, conduct a test to determine whether the intention to vote (DEFINITE) varies between Democrats, Independents, and Republicans (PARTY3).

14.2 Multiple Comparisons

When we conclude from the one-way analysis of variance that at least two treatment means differ, we often need to know which treatment means are responsible for these differences. For example, if an experiment is undertaken to determine whether different locations within a store produce different mean sales, the manager would be keenly interested in determining which locations result in significantly higher sales and which locations result in lower sales. Similarly, a stockbroker would like to know which one of several mutual funds outperforms the others, and a television executive would like to know which television commercials hold the viewers' attention and which are ignored.

Although it may appear that all we need to do is examine the sample means and identify the largest or the smallest to determine which population means are largest or smallest, this is not the case. To illustrate, suppose that in a five-treatment analysis of variance, we discover that differences exist and that the sample means are as follows:

$$\bar{x}_1 = 20 \quad \bar{x}_2 = 19 \quad \bar{x}_3 = 25 \quad \bar{x}_4 = 22 \quad \bar{x}_5 = 17$$

The statistics practitioner wants to know which of the following conclusions are valid:

1. μ_3 is larger than the other means.
2. μ_3 and μ_4 are larger than the other means.
3. μ_5 is smaller than the other means.
4. μ_5 and μ_2 are smaller than the other means.
5. μ_3 is larger than the other means, and μ_5 is smaller than the other means.

From the information we have, it is impossible to determine which, if any, of the statements are true. We need a statistical method to make this determination. The technique is called **multiple comparisons**.

EXAMPLE 14.2

DATA Xm14-02

Comparing the Costs of Repairing Car Bumpers

Because of foreign competition, North American automobile manufacturers have become more concerned with quality. One aspect of quality is the cost of repairing damage caused by accidents. A manufacturer is considering several new types of bumpers. To test how well they react to low-speed collisions, 10 bumpers of each of four different types were installed on mid-size cars, which were then driven into a wall at 5 miles per hour. The cost of repairing the damage in each case was assessed. The data are shown below.

a. Is there sufficient evidence at the 5% significance level to infer that the bumpers differ in their reactions to low-speed collisions?

b. If differences exist, which bumpers differ?

Bumper 1	Bumper 2	Bumper 3	Bumper 4
610	404	599	272
354	663	426	405
234	521	429	197
399	518	621	363
278	499	426	297
358	374	414	538
379	562	332	181
548	505	460	318
196	375	494	412
444	438	637	499

SOLUTION

IDENTIFY

The problem objective is to compare four populations. The data are interval, and the samples are independent. The correct statistical method is the one-way analysis of variance, which we perform using Excel and Minitab.

COMPUTE

EXCEL

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	Bumper 1	10	3800	380.0	16,924		
6	Bumper 2	10	4859	485.9	8,197		
7	Bumper 3	10	4838	483.8	10,426		
8	Bumper 4	10	3482	348.2	14,049		
9							
10							
11	ANOVA						
12	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
13	Between Groups	150,884	3	50,295	4.06	0.0139	2.8663
14	Within Groups	446,368	36	12,399			
15							
16	Total	597,252	39				

MINITAB

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

```
Source  DF      SS     MS     F      P
Factor   3  150884  50295  4.06  0.014
Error   36  446368  12399
Total   39  597252

S = 111.4   R-Sq = 25.26%   R-Sq(adj) = 19.03%

                               Individual 95% CIs For Mean Based on Pooled StDev
Level      N   Mean  StDev   ---------+-----------------+-----------------+-----------------+--------
Bumper 1  10  380.0  130.1           (---------------*--------------)
Bumper 2  10  485.9   90.5                                (----------------*----------------)
Bumper 3  10  483.8  102.1                                (--------------*----------------)
Bumper 4  10  348.2  118.5    (-----------------*--------------)
                             ---------+-----------------+-----------------+-----------------+--------
                                    320               400               480               560
```

INTERPRET

The test statistic is $F = 4.06$ and the p-value $= .0139$. There is enough statistical evidence to infer that there are differences between some of the bumpers. The question is now, Which bumpers differ?

There are several statistical inference procedures that deal with this problem. We will present three methods that allow us to determine which population means differ. All three methods apply to the one-way experiment only.

Fisher's Least Significant Difference (LSD) Method

The **least significant difference (LSD)** method was briefly introduced in Section 14.1 (page 535). To determine which population means differ, we could perform a series of t-tests of the difference between two means on all pairs of population means to determine which are significantly different. In Chapter 13, we introduced the equal-variances t-test of the difference between two means. The test statistic and confidence interval estimator are, respectively,

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$(\bar{x}_1 - \bar{x}_2) \pm t_{\alpha/2}\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

with degrees of freedom $\nu = n_1 + n_2 - 2$.

Recall that s_p^2 is the pooled variance estimate, which is an unbiased estimator of the variance of the two populations. (Recall that the use of these techniques requires that the population variances be equal.) In this section, we modify the test statistic and interval estimator.

Earlier in this chapter, we pointed out that MSE is an unbiased estimator of the common variance of the populations we're testing. Because MSE is based on all the observations in the k samples, it will be a better estimator than s_p^2 (which is based on only two samples). Thus, we could draw inferences about every pair of means by substituting MSE for s_p^2 in the formulas for test statistic and confidence interval estimator shown previously. The number of degrees of freedom would also change to $\nu = n - k$ (where n is the total sample size). The test statistic to determine whether μ_i and μ_j differ is

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

The confidence interval estimator is

$$(\bar{x}_i - \bar{x}_j) \pm t_{\alpha/2}\sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$$

with degrees of freedom $\nu = n - k$.

We define the least significant difference LSD as

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

A simple way of determining whether differences exist between each pair of population means is to compare the absolute value of the difference between their two sample means and LSD. In other words, we will conclude that μ_i and μ_j differ if

$$|\bar{x}_i - \bar{x}_j| > \text{LSD}$$

LSD will be the same for all pairs of means if all k sample sizes are equal. If some sample sizes differ, LSD must be calculated for each combination.

In Section 14.1 we argued that this method is flawed because it will increase the probability of committing a Type I error. That is, it is more likely than the analysis of variance to conclude that a difference exists in some of the population means when in fact none differ. On page 535, we calculated that if $k = 6$ and all population means are equal, the probability of erroneously inferring at the 5% significance level that at least two means differ is about 54%. The 5% figure is now referred to as the *comparisonwise Type I error rate*. The true probability of making at least one Type I error is called the *experimentwise Type I error rate*, denoted α_E. The experimentwise Type I error rate can be calculated as

$$\alpha_E = 1 - (1 - \alpha)^C$$

Here C is the number of pairwise comparisons, which can be calculated by $C = k(k - 1)/2$. Mathematicians have proven that

$$\alpha_E \le C\alpha$$

which means that if we want the probability of making at least one Type I error to be no more than α_E, we simply specify $\alpha = \alpha_E/C$. The resulting procedure is called the **Bonferroni adjustment**.

Bonferroni Adjustment to LSD Method

The adjustment is made by dividing the specified experimentwise Type I error rate by the number of combinations of pairs of population means. For example, if $k = 6$, then

$$C = \frac{k(k-1)}{2} = \frac{6(5)}{2} = 15$$

If we want the true probability of a Type I error to be no more than 5%, we divide this probability by C. Thus, for each test we would use a value of α equal to

$$\alpha = \frac{\alpha_E}{C} = \frac{.05}{15} = .0033$$

We use Example 14.2 to illustrate Fisher's LSD method and the Bonferroni adjustment. The four sample means are

$$\bar{x}_1 = 380.0$$
$$\bar{x}_2 = 485.9$$
$$\bar{x}_3 = 483.8$$
$$\bar{x}_4 = 348.2$$

The pairwise absolute differences are

$$|\bar{x}_1 - \bar{x}_2| = |380.0 - 485.9| = |-105.9| = 105.9$$
$$|\bar{x}_1 - \bar{x}_3| = |380.0 - 483.8| = |-103.8| = 103.8$$
$$|\bar{x}_1 - \bar{x}_4| = |380.0 - 348.2| = |31.8| = 31.8$$
$$|\bar{x}_2 - \bar{x}_3| = |485.9 - 483.8| = |2.1| = 2.1$$
$$|\bar{x}_2 - \bar{x}_4| = |485.9 - 348.2| = |137.7| = 137.7$$
$$|\bar{x}_3 - \bar{x}_4| = |483.8 - 348.2| = |135.6| = 135.6$$

From the computer output, we learn that MSE = 12,399 and $\nu = n - k = 40 - 4 = 36$. If we conduct the LSD procedure with $\alpha = .05$ we find $t_{\alpha/2,n-k} = t_{.025,36} \approx t_{.025,35} = 2.030$. Thus,

$$t_{\alpha/2}\sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} = 2.030\sqrt{12{,}399\left(\frac{1}{10} + \frac{1}{10}\right)} = 101.09$$

We can see that four pairs of sample means differ by more than 101.09. In other words, $|\bar{x}_1 - \bar{x}_2| = 105.9$, $|\bar{x}_1 - \bar{x}_3| = 103.8$, $|\bar{x}_2 - \bar{x}_4| = 137.7$, and $|\bar{x}_3 - \bar{x}_4| = 135.6$. Hence, μ_1 and μ_2, μ_1 and μ_3, μ_2 and μ_4, and μ_3 and μ_4 differ. The other two pairs—μ_1 and μ_4, and μ_2 and μ_3—do not differ.

If we perform the LSD procedure with the Bonferroni adjustment, the number of pairwise comparisons is 6 (calculated as $C = k(k - 1)/2 = 4(3)/2$). We set $\alpha = .05/6 = .0083$. Thus $t_{\alpha/2,36} = t_{.0042,36} = 2.794$ (available from Excel and difficult to approximate manually) and

$$\text{LSD} = t_{\alpha/2}\sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} = 2.794\sqrt{12{,}399\left(\frac{1}{10} + \frac{1}{10}\right)} = 139.13$$

Now no pair of means differ because all the absolute values of the differences between sample means are less than 139.19.

The drawback to the LSD procedure is that we increase the probability of at least one Type I error. The Bonferroni adjustment corrects this problem. However, recall that the probabilities of Type I and Type II errors are inversely related. The Bonferroni adjustment uses a smaller value of α, which results in an increased probability of a Type II error. A Type II error occurs when a difference between population means exists, yet we cannot detect it. This may be the case in this example. The next multiple comparison method addresses this problem.

Tukey's Multiple Comparison Method

A more powerful test is **Tukey's multiple comparison method**. This technique determines a critical number similar to LSD for Fisher's test, denoted by ω (Greek letter *omega*) such that, if any pair of sample means has a difference greater than ω, we conclude that the pair's two corresponding population means are different.

The test is based on the Studentized range, which is defined as the variable

$$q = \frac{\bar{x}_{\max} - \bar{x}_{\min}}{s/\sqrt{n}}$$

where $\bar{x}_{\max}$ and $\bar{x}_{\min}$ are the largest and smallest sample means, respectively, assuming that there are no differences between the population means. We define ω as follows.

Critical Number ω

$$\omega = q_\alpha(k,\nu)\sqrt{\frac{\text{MSE}}{n_g}}$$

where

k = Number of treatments
n = Number of observations ($n = n_1 + n_2 + \cdots + n_k$)
ν = Number of degrees of freedom associated with MSE ($\nu = n - k$)
n_g = Number of observations in each of k samples
α = Significance level
$q_\alpha(k, \nu)$ = Critical value of the Studentized range

Theoretically, this procedure requires that all sample sizes be equal. However, if the sample sizes are different, we can still use this technique provided that the sample sizes are at least similar. The value of n_g used previously is the *harmonic mean* of the sample sizes; that is,

$$n_g = \frac{k}{\frac{1}{n_1} + \frac{1}{n_2} + \cdots + \frac{1}{n_k}}$$

Table 7 in Appendix B provides values of $q_\alpha(k,\nu)$ for a variety of values of k and ν, and for $\alpha = .01$ and $.05$. Applying Tukey's method to Example 14.2, we find

$$\begin{aligned} k &= 4 \\ n_1 = n_2 = n_3 &= n_4 = n_g = 10 \\ \nu &= n - k = 40 - 4 = 36 \\ \text{MSE} &= 12{,}399 \\ q_{.05}(4, 37) &\approx q_{.05}(4, 40) = 3.79 \end{aligned}$$

Thus,

$$\omega = q_\alpha(k,\nu)\sqrt{\frac{\text{MSE}}{n_g}} = (3.79)\sqrt{\frac{12{,}399}{10}} = 133.45$$

There are two absolute values larger than 133.45. Hence, we conclude that μ_2 and μ_4, and μ_3 and μ_4 differ. The other four pairs do not differ.

EXCEL

	A	B	C	D	E
1	**Multiple Comparisons**				
2					
3				LSD	Omega
4	Treatment	Treatment	Difference	Alpha = 0.05	Alpha = 0.05
5	*Bumper 1*	*Bumper 2*	–105.9	100.99	133.45
6		*Bumper 3*	–103.8	100.99	133.45
7		*Bumper 4*	31.8	100.99	133.45
8	*Bumper 2*	*Bumper 3*	2.1	100.99	133.45
9		*Bumper 4*	137.7	100.99	133.45
10	*Bumper 3*	*Bumper 4*	135.6	100.99	133.45

Tukey and Fisher's LSD with the Bonferroni Adjustment (α = .05/6 = 0083)

	A	B	C	D	E
1	**Multiple Comparisons**				
2					
3				LSD	Omega
4	Treatment	Treatment	Difference	Alpha = 0.0083	Alpha = 0.05
5	*Bumper 1*	*Bumper 2*	–105.9	139.11	133.45
6		*Bumper 3*	–103.8	139.11	133.45
7		*Bumper 4*	31.8	139.11	133.45
8	*Bumper 2*	*Bumper 3*	2.1	139.11	133.45
9		*Bumper 4*	137.7	139.11	133.45
10	*Bumper 3*	*Bumper 4*	135.6	139.11	133.45

The printout includes ω (Tukey's method), the differences between sample means for each combination of populations, and Fisher's LSD. (The Bonferroni adjustment is made by specifying another value for α.)

INSTRUCTIONS

1. Type or import the data into adjacent columns. (Open Xm14-02.)
2. Click **Add-Ins, Data Analysis Plus,** and **Multiple Comparisons**.
3. Specify the **Input Range** (A1:D11). Type the value of α. To use the Bonferroni adjustment divide α by $C = k(k-1)/2$. For Tukey, Excel computes ω only for $\alpha = .05$.

MINITAB

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

Individual confidence level = 98.38%

Bumper 1 subtracted from:

           Lower  Center  Upper  ---------+---------+---------+---------+-
Bumper 2   -28.3   105.9  240.1                (--------*--------)
Bumper 3   -30.4   103.8  238.0                (--------*--------)
Bumper 4  -166.0   -31.8  102.4        (--------*--------)
                                 ---------+---------+---------+---------+-
                                       -150         0       150       300

Bumper 2 subtracted from:

           Lower  Center  Upper  ---------+---------+---------+---------+-
Bumper 3  -136.3    -2.1  132.1          (--------*--------)
Bumper 4  -271.9  -137.7   -3.5  (--------*--------)
                                 ---------+---------+---------+---------+--
                                       -150         0       150       300

Bumper 3 subtracted from:

           Lower  Center  Upper  ---------+---------+---------+---------+-
Bumper 4  -269.8  -135.6   -1.4  (--------*--------)
                                 ---------+---------+---------+---------+-
                                       -150         0       150       300

Fisher 99.17% Individual Confidence Intervals
All Pairwise Comparisons

Simultaneous confidence level = 96.04%

Bumper 1 subtracted from:

           Lower  Center  Upper  ---------+---------+---------+---------+-
Bumper 2   -33.2   105.9  245.0                (--------*--------)
Bumper 3   -35.3   103.8  242.9                (--------*--------)
Bumper 4  -170.9   -31.8  107.3        (--------*--------)
                                 ---------+---------+---------+---------+-
                                       -150         0       150       300

Bumper 2 subtracted from:

           Lower  Center  Upper  ---------+---------+---------+---------+-
Bumper 3  -141.2    -2.1  137.0        (--------*--------)
Bumper 4  -276.8  -137.7    1.4  (--------*--------)
                                 ---------+---------+---------+---------+-
                                       -150         0       150       300

Bumper 3 subtracted from:

           Lower  Center  Upper  ---------+---------+---------+---------+-
Bumper 4  -274.4  -135.6    3.5  (--------*--------)
                                 ---------+---------+---------+---------+-
                                       -150         0       150       300
```

Minitab reports the results of Tukey's multiple comparisons by printing interval estimates of the differences between each pair of means. The estimates are computed by calculating the pairwise difference between sample means minus ω for the lower limit and plus ω for the upper limit. The calculations are described in the following table.

Tukey's Method

Pair of Population Means Compared	Difference	Lower Limit	Upper Limit
Bumper 2–Bumper 1	105.9	−28.3	240.1
Bumper 3–Bumper 1	103.8	−30.4	238.0
Bumper 4–Bumper 1	−31.8	−166.0	102.4
Bumper 3–Bumper 2	−2.1	−136.3	132.1
Bumper 4–Bumper 2	−137.7	−271.9	−3.5
Bumper 4–Bumper 3	−135.6	−269.8	−1.4

A similar calculation is performed for Fisher's method replacing ω by LSD.

Fisher's Method

Pair of Population Means Compared	Difference	Lower Limit	Upper Limit
Bumper 2–Bumper 1	105.9	−33.2	245.0
Bumper 3–Bumper 1	103.8	−35.3	242.9
Bumper 4–Bumper 1	−31.8	−170.9	107.3
Bumper 3–Bumper 2	−2.1	−141.2	137.0
Bumper 4–Bumper 2	−137.7	−276.8	1.4
Bumper 4–Bumper 3	−135.6	274.7	3.5

We interpret the test results in the following way. If the interval includes 0, there is not enough evidence to infer that the pair of means differ. If the entire interval is above or the entire interval is below 0, we conclude that the pair of means differ.

INSTRUCTIONS

1. Type or import the data either in stacked or unstacked format. (Open Xm14-02.)
2. Click **Stat, ANOVA**, and **Oneway (Unstacked) . . . .**
3. Type or **Select** the variables in the **Responses (in separate columns)** box (Bumper 1, Bumper 2, Bumper 3, Bumper 4).
4. Click **Comparisons . . .** Select Tukey's method and specify α. Select Fisher's method and specify α. For the Bonferroni adjustment divide α by $C = k(k - 1)/2$.

INTERPRET

Using the Bonferroni adjustment of Fisher's LSD method, we discover that none of the bumpers differ. Tukey's method tells us that bumper 4 differs from both bumpers 2 and 3. Based on this sample, bumper 4 appears to have the lowest cost of repair. Because there was not enough evidence to conclude that bumpers 1 and 4 differ, we would consider using bumper 1 if it has other advantages over bumper 4.

Which Multiple Comparison Method to Use

Unfortunately, no one procedure works best in all types of problems. Most statisticians agree with the following guidelines:

If you have identified two or three pairwise comparisons that you wish to make before conducting the analysis of variance, use the Bonferroni method. This means that if there are 10 populations in a problem but you're particularly interested in comparing, say, populations 3 and 7 and populations 5 and 9, use Bonferroni with $C = 2$.

If you plan to compare all possible combinations, use Tukey.

When do we use Fisher's LSD? If the purpose of the analysis is to point to areas that should be investigated further, Fisher's LSD method is indicated.

Incidentally, to employ Fisher's LSD or the Bonferroni adjustment, you must perform the analysis of variance first. Tukey's method can be employed instead of the analysis of variance.

EXERCISES

Developing an Understanding of Statistical Concepts

14.33 a. Use Fisher's LSD method with $\alpha = .05$ to determine which population means differ in the following problem.

$k = 3$ $n_1 = 10$ $n_2 = 10$ $n_3 = 10$
MSE = 700 $\bar{x}_1 = 128.7$ $\bar{x}_2 = 101.4$ $\bar{x}_3 = 133.7$

b. Repeat part (a) using the Bonferroni adjustment.
c. Repeat part (a) using Tukey's multiple comparison method.

14.34 a. Use Fisher's LSD procedure with $\alpha = .05$ to determine which population means differ given the following statistics:

$k = 5$ $n_1 = 5$ $n_2 = 5$ $n_3 = 5$
MSE = 125 $\bar{x}_1 = 227$ $\bar{x}_2 = 205$ $\bar{x}_3 = 219$
$n_4 = 5$ $n_5 = 5$
$\bar{x}_4 = 248$ $\bar{x}_5 = 202$

b. Repeat part (a) using the Bonferroni adjustment.
c. Repeat part (a) using Tukey's multiple comparison method

Applications

14.35 Apply Tukey's method to determine which brands differ in Exercise 14.5.

14.36 Refer to Exercise 14.6.
a. Employ Fisher's LSD method to determine which degrees differ (use $\alpha = .10$).
b. Repeat part (a) using the Bonferroni adjustment.

Exercises 14.37–14.50 require the use of a computer and software. Use a 5% significance level unless specified otherwise. The answers to Exercises 14.37–14.42 may be calculated manually. See Appendix A for the sample statistics.

14.37 Xr14-09 a. Apply Fisher's LSD method with the Bonferroni adjustment to determine which schools differ in Exercise 14.9.
b. Repeat part (a) applying Tukey's method instead.

14.38 Xr14-10 a. Apply Tukey's multiple comparison method to determine which forms differ in Exercise 14.10.
b. Repeat part (a) applying the Bonferroni adjustment.

14.39 Xr14-39 Police cars, ambulances, and other emergency vehicles are required to carry road flares. One of the most important features of flares is their burning times. To help decide which of four brands on the market to use, a police laboratory technician measured the burning time for a random sample of 10 flares of each brand. The results were recorded to the nearest minute.
a. Can we conclude that differences exist between the burning times of the four brands of flares?
b. Apply Fisher's LSD method with the Bonferroni adjustment to determine which flares are better.
c. Repeat part (b) using Tukey's method.

14.40 Xr14-12 Refer to Exercise 14.12.
a. Apply Fisher's LSD method with the Bonferroni adjustment to determine which lacquers differ.
b. Repeat part (a) applying Tukey's method instead.

14.41 Xr14-41 An engineering student who is about to graduate decided to survey various firms in Silicon Valley to see which offered the best chance for early promotion and career advancement. He surveyed 30 small firms (size level is based on gross revenues), 30 medium-size firms, and 30 large firms and determined how much time must elapse before an average engineer can receive a promotion.
a. Can the engineering student conclude that speed of promotion varies between the three sizes of engineering firms?
b. If differences exist, which of the following is true? Use Tukey's method.
i. Small firms differ from the other two.
ii. Medium-size firms differ from the other two.
iii. Large firms differ from the other two.
iv. All three firms differ from one another.
v. Small firms differ from large firms.

14.42 Xr14-14 a. Apply Tukey's multiple comparison method to determine which fertilizers differ in Exercise 14.14.
b. Repeat part (a) applying the Bonferroni adjustment.

GENERAL SOCIAL SURVEY EXERCISES

14.43 GSS2002* GSS2004* GSS2006* GSS2008* Refer to Exercise 14.21. Use an appropriate statistical technique to determine which years differ with respect to the amount of education (EDUC).

14.44 GSS2008* Refer to Exercise 14.22. Use Tukey's multiple comparison method to determine which races differ in the amount of television watched (TVHOURS).

14.45 GSS2008* Refer to Exercise 14.23. Use a suitable multiple comparison method to determine which degrees differ in annual incomes (INCOME).

14.46 GSS2008* Refer to Exercise 14.25. Apply a suitable multiple comparison method to determine which categories of marital status differ.

American National Election Survey Exercises

14.47 ANES2008* Refer to Exercise 14.27. Apply Tukey's multiple comparison method to determine which positions on the liberal–conservative spectrum (LIBCON) differ with respect to income (INCOME).

14.48 ANES2008* Refer to Exercise 14.28. Use a multiple comparison method to determine which of the three parties (PARTY3) differ with respect to education (EDUC).

14.49 ANES2008* Refer to Exercise 14.31. Use Tukey's multiple comparison method to find which categories of marital status (MARITAL) differ with respect to education (EDUC).

14.50 ANES2008* Refer to Exercise 14.32. Use an appropriate multiple comparison method to determine which of the three parties (PARTY3) differ with respect to intention to vote (DEFINITE).

14.3 Analysis of Variance Experimental Designs

Since we introduced the matched pairs experiment in Section 13.3, the experimental design has been one of the factors that determines which technique we use. Statistics practitioners often design experiments to help extract the information they need to assist them in making decisions. The one-way analysis of variance introduced in Section 14.1 is only one of many different experimental designs of the analysis of variance. For each type of experiment, we can describe the behavior of the response variable using a mathematical expression or model. Although we will not exhibit the mathematical expressions in this chapter (we introduce models in Chapter 16), we think it is useful for you to be aware of the elements that distinguish one experimental design or model from another. In this section, we present some of these elements; in so doing, we introduce two of the experimental designs that will be presented later in this chapter.

Single-Factor and Multifactor Experimental Designs

As we pointed out in Section 14.1, the criterion by which we identify populations is called a *factor*. The experiment described in Section 14.1 is a single-factor analysis of variance because it addresses the problem of comparing two or more populations defined on the basis of only one factor. A **multifactor experiment** is one in which two or more factors define the treatments. The experiment described in Example 14.1 is a single-factor design because we had one treatment: age of the head of the household. In other words, the factor is the age, and the four age categories were the levels of this factor.

Suppose that we can also look at the gender of the household head in another study. We would then develop a two-factor analysis of variance in which the first factor, age, has four levels, and the second factor, gender, has two levels. We will discuss two-factor experiments in Section 14.5.

Independent Samples and Blocks

In Section 13.3, we introduced statistical techniques where the data were gathered from a matched pairs experiment. This type of experimental design reduces the variation within the samples, making it easier to detect differences between the two populations.

When the problem objective is to compare more than two populations, the experimental design that is the counterpart of the matched pairs experiment is called the **randomized block design**. The term *block* refers to a matched group of observations from each population. Suppose that in Examples 13.4 and 13.5 we had wanted to compare the salary offers for finance, marketing, accounting, and operations management majors. To redo Example 13.5 we would conduct a randomized block experiment where the blocks are the 25 GPA groups and the treatments are the four MBA majors.

Once again, the experimental design should reduce the variation in each treatment to make it easier to detect differences.

We can also perform a blocked experiment by using the same subject (person, plant, and store) for each treatment. For example, we can determine whether sleeping pills are effective by giving three brands of pills to the same group of people to measure the effects. Such experiments are called **repeated measures** designs. Technically, this is a different design than the randomized block. However, the data are analyzed in the same way for both designs. Hence, we will treat repeated measures designs as randomized block designs.

The randomized block experiment is also called the **two-way analysis of variance**. In Section 14.4, we introduce the technique used to calculate the test statistic for this type of experiment.

Fixed and Random Effects

If our analysis includes all possible levels of a factor, the technique is called a **fixed-effects analysis of variance**. If the levels included in the study represent a random sample of all the levels that exist, the technique is called a **random-effects analysis of variance**. In Example 14.2, there were only four possible bumpers. Consequently, the study is a fixed-effects experiment. However, if there were other bumpers besides the four described in the example, and we wanted to know whether there were differences in repair costs between all bumpers, the application would be a random-effects experiment. Here's another example.

To determine whether there is a difference in the number of units produced by the machines in a large factory, 4 machines out of 50 in the plant are randomly selected for study. The number of units each produces per day for 10 days will be recorded. This experiment is a random-effects experiment because we selected a random sample of four machines and the statistical results thus allow us to determine whether there are differences between the 50 machines.

In some experimental designs, there are no differences in calculations of the test statistic between fixed and random effects. However, in others, including the two-factor experiment presented in Section 14.5, the calculations are different.

14.4 Randomized Block (Two-Way) Analysis of Variance

The purpose of designing a randomized block experiment is to reduce the within-treatments variation to more easily detect differences between the treatment means. In the one-way analysis of variance, we partitioned the total variation into the between-treatments and the within-treatments variation; that is,

$$\text{SS(Total)} = \text{SST} + \text{SSE}$$

In the randomized block design of the analysis of variance, we partition the total variation into three sources of variation,

$$\text{SS(Total)} = \text{SST} + \text{SSB} + \text{SSE}$$

where **SSB**, the **sum of squares for blocks**, measures the variation between the blocks. When the variation associated with the blocks is removed, SSE is reduced, making it easier to determine whether differences exist between the treatment means.

At this point in our presentation of statistical inference, we will deviate from our usual procedure of solving examples in three ways: manually, using Excel, and using Minitab. The calculations for this experimental design and for the experiment presented in the next section are so time consuming that solving them by hand adds little to your understanding of the technique. Consequently, although we will continue to present the concepts by discussing how the statistics are calculated, we will solve the problems only by computer.

To help you understand the formulas, we will use the following notation:

$\bar{x}[T]_j$ = Mean of the observations in the jth treatment ($j = 1, 2, \ldots, k$)

$\bar{x}[B]_i$ = Mean of the observations in the ith block ($i = 1, 2, \ldots, b$)

b = Number of blocks

Table 14.4 summarizes the notation we use in this experimental design.

TABLE 14.4 Notation for the Randomized Block Analysis of Variance

	TREATMENTS				
BLOCK	**1**	**2**		***k***	**BLOCK MEAN**
1	x_{11}	x_{12}	...	x_{1k}	$\bar{x}[B]_1$
2	x_{21}	x_{22}	...	x_{2k}	$\bar{x}[B]_2$
⋮	⋮	⋮		⋮	⋮
b	x_{b1}	x_{b2}	...	x_{bk}	$\bar{x}[B]_b$
Treatment mean	$\bar{x}[T]_1$	$\bar{x}[T]_2$	...	$\bar{x}[T]_k$	

The definitions of SS(Total) and SST in the randomized block design are identical to those in the independent samples design. SSE in the independent samples design is equal to the sum of SSB and SSE in the randomized block design.

Sums of Squares in the Randomized Block Experiment

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

$$\text{SST} = \sum_{j=1}^{k} b(\bar{x}[T]_j - \bar{\bar{x}})^2$$

$$\text{SSB} = \sum_{i=1}^{b} k(\bar{x}[B]_i - \bar{\bar{x}})^2$$

$$\text{SSE} = \sum_{j=1}^{k}\sum_{i=1}^{b}(x_{ij} - \bar{x}[T]_j - \bar{x}[B]_i + \bar{\bar{x}})^2$$

The test is conducted by determining the mean squares, which are computed by dividing the sums of squares by their respective degrees of freedom.

Mean Squares for the Randomized Block Experiment

$$\text{MST} = \frac{\text{SST}}{k - 1}$$

$$\text{MSB} = \frac{\text{SSB}}{b - 1}$$

$$\text{MSE} = \frac{\text{SSE}}{n - k - b + 1}$$

Finally, the test statistic is the ratio of mean squares, as described in the box.

Test Statistic for the Randomized Block Experiment

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

which is F-distributed with $\nu_1 = k - 1$ and $\nu_2 = n - k - b + 1$ degrees of freedom.

An interesting, and sometimes useful, by-product of the test of the treatment means is that we can also test to determine whether the block means differ. This will allow us to determine whether the experiment should have been conducted as a randomized block design. (If there are no differences between the blocks, the randomized block design is less likely to detect real differences between the treatment means.) Such a discovery could be useful in future similar experiments. The test of the block means is almost identical to that of the treatment means except the test statistic is

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

which is F-distributed with $\nu_1 = b - 1$ and $\nu_2 = n - k - b + 1$ degrees of freedom.

As with the one-way experiment, the statistics generated in the randomized block experiment are summarized in an ANOVA table, whose general form is exhibited in Table 14.5.

TABLE **14.5** ANOVA Table for the Randomized Block Analysis of Variance

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-STATISTIC
Treatments	$k - 1$	SST	MST = SST/$(k - 1)$	F = MST/MSE
Blocks	$b - 1$	SSB	MSB = SSB/$(b - 1)$	F = MSB/MSE
Error	$n - k - b + 1$	SSE	MSE = SSE/$(n - k - b + 1)$	
Total	$n - 1$	SS(Total)		

EXAMPLE 14.3 Comparing Cholesterol-Lowering Drugs

DATA
Xm14-03

Many North Americans suffer from high levels of cholesterol, which can lead to heart attacks. For those with very high levels (above 280), doctors prescribe drugs to reduce cholesterol levels. A pharmaceutical company has recently developed four such drugs. To determine whether any differences exist in their benefits, an experiment was organized. The company selected 25 groups of four men, each of whom had cholesterol levels in excess of 280. In each group, the men were matched according to age and weight. The drugs were administered over a 2-month period, and the reduction in cholesterol was recorded. Do these results allow the company to conclude that differences exist between the four new drugs?

Group	Drug 1	Drug 2	Drug 3	Drug 4
1	6.6	12.6	2.7	8.7
2	7.1	3.5	2.4	9.3
3	7.5	4.4	6.5	10
4	9.9	7.5	16.2	12.6
5	13.8	6.4	8.3	10.6
6	13.9	13.5	5.4	15.4
7	15.9	16.9	15.4	16.3
8	14.3	11.4	17.1	18.9
9	16	16.9	7.7	13.7
10	16.3	14.8	16.1	19.4
11	14.6	18.6	9	18.5
12	18.7	21.2	24.3	21.1
13	17.3	10	9.3	19.3
14	19.6	17	19.2	21.9
15	20.7	21	18.7	22.1
16	18.4	27.2	18.9	19.4
17	21.5	26.8	7.9	25.4
18	20.4	28	23.8	26.5
19	21.9	31.7	8.8	22.2
20	22.5	11.9	26.7	23.5
21	21.5	28.7	25.2	19.6
22	25.2	29.5	27.3	30.1
23	23	22.2	17.6	26.6
24	23.7	19.5	25.6	24.5
25	28.4	31.2	26.1	27.4

SOLUTION

IDENTIFY

The problem objective is to compare four populations, and the data are interval. Because the researchers recorded the cholesterol reduction for each drug for each member of the similar groups of men, we identify the experimental design as randomized block. The response variable is the cholesterol reduction, the treatments are the drugs, and the blocks are the 25 similar groups of men. The hypotheses to be tested are as follows.

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1: At least two means differ

COMPUTE

EXCEL

	A	B	C	D	E	F	G
36	ANOVA						
37	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
38	Rows	3848.66	24	160.36	10.11	9.70E-15	1.6695
39	Columns	195.95	3	65.32	4.12	0.0094	2.7318
40	Error	1142.56	72	15.87			
41							
42	Total	5187.17	99				

Note the use of scientific notation for one of the *p*-values. The number 9.70E-15 (E stands for *exponent*) is 9.70 multiplied by 10 raised to the power -15, that is, 9.70×10^{-15}. You can increase or decrease the number of decimal places, and you can convert the number into a regular number, but you would need many decimal places, which is why Excel uses scientific notation when the number is very small. (Excel also uses scientific notation for very large numbers.)

The output includes block and treatment statistics (sums, averages, and variances, which are not shown here), and the ANOVA table. The *F*-statistic to determine whether differences exist between the four drugs (**Columns**) is 4.12. Its *p*-value is .0094. The other *F*-statistic, 10.11 (p-value $= 9.70 \times 10^{-15}$ = virtually 0), indicates that there are differences between the groups of men (**Rows**).

INSTRUCTIONS

1. Type or import the data into adjacent columns*. (Open Xm14-03.)
2. Click **Data, Data Analysis . . .** , and **Anova: Two-Factor Without Replication**.
3. Specify the **Input Range** (A1:E26). Click **Labels** if applicable. If you do, both the treatments and blocks must be labeled (as in Xm14-03). Specify the value of α (.05).

MINITAB

Two-way ANOVA: Reduction versus Group, Drug

Analysis of Variance for Reduction

Source	DF	SS	MS	F	P
Group	24	3848.7	160.4	10.11	0.000
Drug	3	196.0	65.3	4.12	0.009
Error	72	1142.6	15.9		
Total	99	5187.2			

The *F*-statistic for **Drug** is 4.12 with a *p*-value of .009. The *F*-statistic for the blocks (**Group**) is 10.11, with a *p*-value of 0.

INSTRUCTIONS

The data must be in stacked format in three columns. One column contains the responses, another contains codes for the levels of the blocks, and a third column contains codes for the levels of the treatments.

1. Click **Stat, ANOVA,** and **Twoway . . . .**
2. Specify the **Responses, Row factor,** and **Column factor**

*If one or more columns contain a blank (representing missing data) the entire row must be deleted.

INTERPRET

A Type I error occurs when you conclude that differences exist when, in fact, they do not. A Type II error is committed when the test reveals no difference when at least two means differ. It would appear that both errors are equally costly. Accordingly, we judge the p-value against a standard of 5%. Because the p-value = .0094, we conclude that there is sufficient evidence to infer that at least two of the drugs differ. An examination reveals that cholesterol reduction is greatest using drugs 2 and 4. Further testing is recommended to determine which is better.

Checking the Required Conditions

The F-test of the randomized block design of the analysis of variance has the same requirements as the independent samples design. That is, the random variable must be normally distributed and the population variances must be equal. The histograms (not shown) appear to support the validity of our results; the reductions appear to be normal. The equality of variances requirement also appears to be met.

Violation of the Required Conditions

When the response is not normally distributed, we can replace the randomized block analysis of variance with the Friedman test, which is introduced in Section 19.4.

Criteria for Blocking

In Section 13.3, we listed the advantages and disadvantages of performing a matched pairs experiment. The same comments are valid when we discuss performing a blocked experiment. The purpose of blocking is to reduce the variation caused by differences between the experimental units. By grouping the experimental units into homogeneous blocks with respect to the response variable, the statistics practitioner increases the chances of detecting actual differences between the treatment means. Hence, we need to find criteria for blocking that significantly affect the response variable. For example, suppose that a statistics professor wants to determine which of four methods of teaching statistics is best. In a one-way experiment, he might take four samples of 10 students, teach each sample by a different method, grade the students at the end of the course, and perform an F-test to determine whether differences exist. However, it is likely that there are very large differences between the students within each class that may hide differences between classes. To reduce this variation, the statistics professor must identify variables that are linked to a student's grade in statistics. For example, overall ability of the student, completion of mathematics courses, and exposure to other statistics courses are all related to performance in a statistics course.

The experiment could be performed in the following way. The statistics professor selects four students at random whose average grade before statistics is 95–100. He then randomly assigns the students to one of the four classes. He repeats the process with students whose average is 90–95, 85–90, . . . , and 50–55. The final grades would be used to test for differences between the classes.

Any characteristics that are related to the experimental units are potential blocking criteria. For example, if the experimental units are people, we may block according to age, gender, income, work experience, intelligence, residence (country, county, or city), weight, or height. If the experimental unit is a factory and we're measuring number of units produced hourly, blocking criteria include workforce experience, age of the plant, and quality of suppliers.

Developing an Understanding of Statistical Concepts

As we explained previously, the randomized block experiment is an extension of the matched pairs experiment discussed in Section 13.3. In the matched pairs experiment, we simply remove the effect of the variation caused by differences between the experimental units. The effect of this removal is seen in the decrease in the value of the standard error (compared to the standard error in the test statistic produced from independent samples) and the increase in the value of the t-statistic. In the randomized block experiment of the analysis of variance, we actually measure the variation between the blocks by computing SSB. The sum of squares for error is reduced by SSB, making it easier to detect differences between the treatments. In addition, we can test to determine whether the blocks differ—a procedure we were unable to perform in the matched pairs experiment.

To illustrate, let's return to Examples 13.4 and 13.5, which were experiments to determine whether there was a difference in starting salaries offered to finance and marketing MBA majors. (In fact, we tested to determine whether finance majors draw higher salary offers than do marketing majors. However, the analysis of variance can test only for differences.) In Example 13.4 (independent samples), there was insufficient evidence to infer a difference between the two types of majors. In Example 13.5 (matched pairs experiment), there was enough evidence to infer a difference. As we pointed out in Section 13.3, matching by grade point average allowed the statistics practitioner to more easily discern a difference between the two types of majors. If we repeat Examples 13.4 and 13.5 using the analysis of variance, we come to the same conclusion. The Excel outputs are shown here. (Minitab's printouts are similar.)

Excel Analysis of Variance Output for Example 13.4

	A	B	C	D	E	F	G
9	ANOVA						
10	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
11	Between Groups	338,130,013	1	338,130,013	1.09	0.3026	4.0427
12	Within Groups	14,943,884,470	48	311,330,926			
13							
14	Total	15,282,014,483	49				

Excel Analysis of Variance Output for Example 13.5

	A	B	C	D	E	F	G
34	ANOVA						
35	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
36	Rows	21,415,991,654	24	892,332,986	40.39	4.17E-14	1.9838
37	Columns	320,617,035	1	320,617,035	14.51	0.0009	4.2597
38	Error	530,174,605	24	22,090,609			
39							
40	Total	22,266,783,295	49				

In Example 13.4, we partition the total sum of squares [SS(Total) = 15,282,014,483] into two sources of variation: SST = 338,130,013 and SSE = 14,943,884,470. In Example 13.5, the total sum of squares is SS(Total) = 22,266,783,295, SST (sum of squares for majors) = 320,617,035, SSB (sum of squares for GPA) = 21,415,991,654, and SSE = 530,174,605. As you can see, the sums of squares for treatments are approximately equal (338,130,013 and 320,617,035). However, the two calculations differ in the sums of squares for error. SSE in Example 13.5 is much smaller than SSE in Example 13.4 because the randomized block experiment allows us to measure and remove the effect of the variation between MBA students with the same majors. The sum of squares for blocks (sum of squares for GPA groups) is 21,415,991,654, a statistic that measures how much variation exists between the salary offers within majors. As a result of removing this variation, SSE is small. Thus, we conclude in Example 13.5 that the salary offers differ between majors whereas there was not enough evidence in Example 13.4 to draw the same conclusion.

Notice that in both examples the t-statistic squared equals the F-statistic. in Example 13.4, $t = 1.04$, which when squared equals 1.09, which is the F-statistic (rounded). In Example 13.5, $t = 3.81$, which when squared equals 14.51, the F-statistic for the test of the treatment means. Moreover, the p-values are also the same.

We now complete this section by listing the factors that we need to recognize to use this experiment of the analysis of variance.

Factors That Identify the Randomized Block of the Analysis of Variance

1. **Problem objective:** Compare two or more populations
2. **Data type:** Interval
3. **Experimental design:** Blocked samples

EXERCISES

Developing an Understanding of Statistical Concepts

14.51 The following statistics were generated from a randomized block experiment with $k = 3$ and $b = 7$:

SST = 100 SSB = 50 SSE = 25

a. Test to determine whether the treatment means differ. (Use $\alpha = .05$.)
b. Test to determine whether the block means differ. (Use $\alpha = .05$.)

14.52 A randomized block experiment produced the following statistics:

$k = 5$ $b = 12$ SST = 1,500 SSB = 1,000 SS(Total) = 3,500

a. Test to determine whether the treatment means differ. (Use $\alpha = .01$.)
b. Test to determine whether the block means differ. (Use $\alpha = .01$.)

14.53 Suppose the following statistics were calculated from data gathered from a randomized block experiment with $k = 4$ and $b = 10$:

SS(Total) = 1,210 SST = 275 SSB = 625

a. Can we conclude from these statistics that the treatment means differ? (Use $\alpha = .01$.)
b. Can we conclude from these statistics that the block means differ? (Use $\alpha = .01$.)

14.54 A randomized block experiment produced the following statistics.

$k = 3$ $b = 8$ SST = 1,500 SS(Total) = 3,500

a. Test at the 5% significance level to determine whether the treatment means differ given that SSB = 500.
b. Repeat part (a) with SSB = 1,000.
c. Repeat part (a) with SSB = 1,500.
d. Describe what happens to the test statistic as SSB increases.

14.55 Xr14-55 a. Assuming that the data shown here were generated from a randomized block experiment, calculate SS(Total), SST, SSB, and SSE.
b. Assuming that the data below were generated from a one-way (independent samples) experiment, calculate SS(Total), SST, and SSE.

c. Why does SS(Total) remain the same for both experimental designs?
d. Why does SST remain the same for both experimental designs?
e. Why does SSB + SSE in part (a) equal SSE in part (b)?

Treatment		
1	2	3
7	12	8
10	8	9
12	16	13
9	13	6
12	10	11

14.56 Xr14-56 a. Calculate SS(Total), SST, SSB, and SSE, assuming that the accompanying data were generated from a randomized block experiment.
b. Calculate SS(Total), SST, and SSE, assuming that the data below were generated from a one-way (independent samples) experiment.
c. Explain why SS(Total) remains the same for both experimental designs.
d. Explain why SST remains the same for both experimental designs.
e. Explain why SSB + SSE in part (a) equals SSE in part (b).

Treatment			
1	2	3	4
6	5	4	4
8	5	5	6
7	6	5	6

Applications

14.57 Xr14-57 As an experiment to understand measurement error, a statistics professor asks four students to measure the height of the professor, a male student, and a female student. The differences (in centimeters) between the correct dimension and the ones produced by the students are listed here. Can we infer that there are differences in the errors between the subjects being measured? (Use $\alpha = .05$.)

	Errors in Measuring Heights of		
Student	Professor	Male Student	Female Student
1	1.4	1.5	1.3
2	3.1	2.6	2.4
3	2.8	2.1	1.5
4	3.4	3.6	2.9

14.58 Xr14-58 How well do diets work? In a preliminary study, 20 people who were more than 50 pounds overweight were recruited to compare four diets. The people were matched by age. The oldest four became block 1, the next oldest four became block 2, and so on. The number of pounds that each person lost are listed in the following table. Can we infer at the 1% significance level that there are differences between the four diets?

	Diet			
Block	1	2	3	4
1	5	2	6	8
2	4	7	8	10
3	6	12	9	2
4	7	11	16	7
5	9	8	15	14

Exercises 14.59–14.67 require the use of a computer and software. Use a 5% significance level unless specified otherwise. The answers to Exercises 14.59–14.65 may be calculated manually. See Appendix A for the sample statistics.

14.59 Xr14-59 In recent years, lack of confidence in the U.S. Postal Service has led many companies to send all of their correspondence by private courier. A large company is in the process of selecting one of 3 possible couriers to act as its sole delivery method. To help in making the decision, an experiment was performed in which letters were sent using each of the 3 couriers at 12 different times of the day to a delivery point across town. The number of minutes required for delivery was recorded.
a. Can we conclude that there are differences in delivery times between the three couriers?
b. Did the statistics practitioner choose the correct design? Explain.

14.60 Xr14-60 Refer to Exercise 14.14. Despite failing to show that differences in the three types of fertilizer exist, the scientist continued to believe that there were differences, and that the differences were masked by the variation between the plots of land. Accordingly, she conducted another experiment. In the second experiment, she found 20 three-acre plots of land scattered across the county. She divided each into three plots and applied the three types of fertilizer on each of the 1-acre plots. The crop yields were recorded.
a. Can the scientist infer that there are differences between the three types of fertilizer?
b. What do these test results reveal about the variation between the plots?

14.61 Xr14-61 A recruiter for a computer company would like to determine whether there are differences in sales ability between business, arts, and science graduates. She takes a random sample of 20 business graduates who have been working for the company for the past 2 years. Each is then matched with an arts graduate and a science graduate with similar educational and working experience. The commission earned by each (in $1,000s) in the last year was recorded.

a. Is there sufficient evidence to allow the recruiter to conclude that there are differences in sales ability between the holders of the three types of degrees?
b. Conduct a test to determine whether an independent samples design would have been a better choice.
c. What are the required conditions for the test in part (a)?
d. Are the required conditions satisfied?

14.62 Xr14-62 Exercise 14.10 described an experiment that involved comparing the completion times associated with four different income tax forms. Suppose the experiment is redone in the following way. Thirty people are asked to fill out all four forms. The completion times (in minutes) are recorded.

a. Is there sufficient evidence at the 1% significance level to infer that differences in the completion times exist between the four forms?
b. Comment on the suitability of this experimental design in this problem.

14.63 Xr14-63 The advertising revenues commanded by a radio station depend on the number of listeners it has. The manager of a station that plays mostly hard rock music wants to learn more about its listeners—mostly teenagers and young adults. In particular, he wants to know whether the amount of time they spend listening to radio music varies by the day of the week. If the manager discovers that the mean time per day is about the same, he will schedule the most popular music evenly throughout the week. Otherwise, the top hits will be played mostly on the days that attract the greatest audience. An opinion survey company is hired, and it randomly selects 200 teenagers and asks them to record the amount of time spent listening to music on the radio for each day of the previous week. What can the manager conclude from these data?

14.64 Xr14-64 Do medical specialists differ in the amount of time they devote to patient care? To answer this question, a statistics practitioner organized a study. The numbers of hours of patient care per week were recorded for five specialists. The experimental design was randomized blocks. The physicians were blocked by age. (Adapted from the *Statistical Abstract of the United States, 2000*, Table 190.)

a. Can we infer that there are differences in the amount of patient care between medical specialties?
b. Can we infer that blocking by age was appropriate?

14.65 Xr14-65 Refer to Exercise 14.9. Another study was conducted in the following way. Students from each of the high schools who were admitted to the business program were matched according to their high school averages. The average grades in the first year were recorded. Can the university admissions officer conclude that there are differences in grading standards between the four high schools?

AMERICAN NATIONAL ELECTION SURVEY EXERCISES

14.66 ANES2008* Is there sufficient evidence to infer that there are differences between the number of days Americans watch national news on television (DAYS1), watch local television news in the afternoon or early evening (DAYS2), watch local television news in the late evening (DAYS3), and read a daily newspaper (DAYS4)?

Warning: There are blanks representing missing data that must be removed.

14.67 ANES2004* Repeat Exercise 14.66 for 2004.

14.5 TWO-FACTOR ANALYSIS OF VARIANCE

In Section 14.1, we addressed problems where the data were generated from single-factor experiments. In Example 14.1, the treatments were the four age categories. Thus, there were four levels of a single factor. In this section, we address the problem where the experiment features two factors. The general term for such data-gathering procedures is **factorial experiment**. In factorial experiments, we can examine the effect on the response variable of two or more factors, although in this book we address the problem of only two factors. We can use the analysis of variance to determine whether the levels of each factor are different from one another.

We will present the technique for fixed effects only. That means we will address problems where all the levels of the factors are included in the experiment. As was the case with the randomized block design, calculating the test statistic in this type of experiment is quite time consuming. As a result, we will use Excel and Minitab to produce our statistics.

EXAMPLE 14.4*

Comparing the Lifetime Number of Jobs by Educational Level

DATA
Xm14-04

One measure of the health of a nation's economy is how quickly it creates jobs. One aspect of this issue is the number of jobs individuals hold. As part of a study on job tenure, a survey was conducted in which Americans aged between 37 and 45 were asked how many jobs they have held in their lifetimes. Also recorded were gender and educational attainment. The categories are

Less than high school (E1)

High school (E2)

Some college/university but no degree (E3)

At least one university degree (E4)

The data are shown for each of the eight categories of gender and education. Can we infer that differences exist between genders and educational levels?

Male E1	Male E2	Male E3	Male E4	Female E1	Female E2	Female E3	Female E4
10	12	15	8	7	7	5	7
9	11	8	9	13	12	13	9
12	9	7	5	14	6	12	3
16	14	7	11	6	15	3	7
14	12	7	13	11	10	13	9
17	16	9	8	14	13	11	6
13	10	14	7	13	9	15	10
9	10	15	11	11	15	5	15
11	5	11	10	14	12	9	4
15	11	13	8	12	13	8	11

SOLUTION

IDENTIFY

We begin by treating this example as a one-way analysis of variance. Notice that there are eight treatments. However, the treatments are defined by two different factors. One factor is gender, which has two levels. The second factor is educational attainment, which has four levels.

We can proceed to solve this problem in the same way we did in Section 14.1: We test the following hypotheses.

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8$$
$$H_1: \text{At least two means differ}$$

*Adapted from the *Statistical Abstract of the United States, 2006*, Table 598.

COMPUTE

EXCEL

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	Male E1	10	126	12.60	8.27		
6	Male E2	10	110	11.00	8.67		
7	Male E3	10	106	10.60	11.60		
8	Male E4	10	90	9.00	5.33		
9	Female E1	10	115	11.50	8.28		
10	Female E2	10	112	11.20	9.73		
11	Female E3	10	94	9.40	16.49		
12	Female E4	10	81	8.10	12.32		
13							
14							
15	ANOVA						
16	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
17	Between Groups	153.35	7	21.91	2.17	0.0467	2.1397
18	Within Groups	726.20	72	10.09			
19							
20	Total	879.55	79				

MINITAB

One-way ANOVA: Male E1, Male E2, Male E3, Male E4, Female E1, Female E2, ...

```
Source  DF     SS    MS     F      P
Factor   7  153.4  21.9  2.17  0.047
Error   72  726.2  10.1
Total   79  879.5

S = 3.176   R-Sq = 17.44%   R-Sq(adj) = 9.41%


                                     Individual 95% CIs For Mean Based on
                                     Pooled StDev
Level       N    Mean   StDev   ---------+---------+---------+---------+-----
Male E1    10  12.600   2.875                      (---------*---------)
Male E2    10  11.000   2.944                (---------*---------)
Male E3    10  10.600   3.406              (---------*---------)
Male E4    10   9.000   2.309        (---------*---------)
Female E1  10  11.500   2.877                  (---------*---------)
Female E2  10  11.200   3.120                 (---------*---------)
Female E3  10   9.400   4.061          (---------*---------)
Female E4  10   8.100   3.510    (---------*---------)
                                ---------+---------+---------+---------+-----
                                       7.5      10.0      12.5      15.0

Pooled StDev = 3.176
```

INTERPRET

The value of the test statistic is $F = 2.17$ with a p-value of .0467. We conclude that there are differences in the number of jobs between the eight treatments.

This statistical result raises more questions—namely, can we conclude that the differences in the mean number of jobs are caused by differences between males and females? Or are they caused by differences between educational levels? Or, perhaps, are there combinations, called **interactions**, of gender and education that result in especially high or low numbers? To show how we test for each type of difference, we need to develop some terminology.

A **complete factorial experiment** is an experiment in which the data for all possible combinations of the levels of the factors are gathered. That means that in Example 14.4 we measured the number of jobs for all eight combinations. This experiment is called a complete 2×4 factorial experiment.

In general, we will refer to one of the factors as factor A (arbitrarily chosen). The number of levels of this factor will be denoted by a. The other factor is called factor B, and its number of levels is denoted by b. This terminology becomes clearer when we present the data from Example 14.4 in another format. Table 14.6 depicts

TABLE **14.6** Two-Way Classification for Example 14.4

	MALE	FEMALE
Less than high school	10	7
	9	13
	12	14
	16	6
	14	11
	17	14
	13	13
	9	11
	11	14
	15	12
High School	12	7
	11	12
	9	6
	14	15
	12	10
	16	13
	10	9
	10	15
	5	12
	11	13
Less than bachelor's degree	15	5
	8	13
	7	12
	7	3
	7	13
	9	11
	14	15
	15	5
	11	9
	13	8
At least one bachelor's degree	8	7
	9	9
	5	3
	11	7
	13	9
	8	6
	7	10
	11	15
	10	4
	8	11

the layout for a *two-way classification*, which is another name for the complete factorial experiment. The number of observations for each combination is called a **replicate**. The number of replicates is denoted by r. In this book, we address only problems in which the number of replicates is the same for each treatment. Such a design is called **balanced**.

Thus, we use a complete factorial experiment where the number of treatments is ab with r replicates per treatment. In Example 14.4, $a = 2$, $b = 4$, and $r = 10$. As a result, we have 10 observations for each of the eight treatments.

If you examine the ANOVA table, you can see that the total variation is SS(Total) = 879.55, the sum of squares for treatments is SST = 153.35, and the sum of squares for error is SSE = 726.20. The variation caused by the treatments is measured by SST. To determine whether the differences result from factor A, factor B, or some interaction between the two factors, we need to partition SST into three sources. These are SS(A), SS(B), and SS(AB).

For those whose mathematical confidence is high, we have provided an explanation of the notation as well as the definitions of the sums of squares. Learning how the sums of squares are calculated is useful but hardly essential to your ability to conduct the tests. Uninterested readers should jump to the box on page 569 where we describe the individual F-tests.

How the Sums of Squares for Factors A and B and Interaction are Computed

To help you understand the formulas, we will use the following notation:

x_{ijk} = kth observation in the ijth treatment

$\bar{x}[AB]_{ij}$ = Mean of the response variable in the ijth treatment (mean of the treatment when the factor A level is i and the factor B level is j)

$\bar{x}[A]_i$ = Mean of the observations when the factor A level is i

$\bar{x}[B]_j$ = Mean of the observations when the factor B level is j

$\bar{\bar{x}}$ = Mean of all the observations

a = Number of factor A levels

b = Number of factor B levels

r = Number of replicates

In this notation, $\bar{x}[AB]_{11}$ is the mean of the responses for factor A level 1 and factor B level 1. The mean of the responses for factor A level 1 is $\bar{x}[A]_1$. The mean of the responses for factor B level 1 is $\bar{x}[B]_1$.

Table 14.7 describes the notation for the two-factor analysis of variance.

TABLE 14.7 Notation for Two-Factor Analysis of Variance

	Factor A				
Factor B	1	2	. . .	a	
1	x_{111} x_{112} . . . x_{11r} $\bar{x}[\text{AB}]_{11}$	x_{211} x_{212} . . . x_{21r} $\bar{x}[\text{AB}]_{21}$		x_{a11} x_{a12} . . . x_{a1r} $\bar{x}[\text{AB}]_{a1}$	$\bar{x}[\text{B}]_1$
2	x_{121} x_{122} . . . x_{12r} $\bar{x}[\text{AB}]_{12}$	x_{221} x_{222} . . . x_{22r} $\bar{x}[\text{AB}]_{22}$		x_{a21} x_{a22} . . . x_{a2r} $\bar{x}[\text{AB}]_{a2}$	$\bar{x}[\text{B}]_2$
. . .					
b	x_{1b1} x_{1b2} . . . x_{1br} $\bar{x}[\text{AB}]_{1b}$	x_{2b1} x_{2b2} . . . x_{2br} $\bar{x}[\text{AB}]_{2b}$		x_{ab1} x_{ab2} . . . x_{abr} $\bar{x}[\text{AB}]_{ab}$	$\bar{x}[\text{B}]_b$
	$\bar{x}[\text{A}]_1$	$\bar{x}[\text{A}]_2$		$\bar{x}[\text{A}]_a$	$\bar{\bar{x}}$

The sums of squares are defined as follows.

Sums of Squares in the Two-Factor Analysis of Variance

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

$$\text{SS(A)} = rb\sum_{i=1}^{a}(\bar{x}[\text{A}]_i - \bar{\bar{x}})^2$$

$$\text{SS(B)} = ra\sum_{j=1}^{b}(\bar{x}[\text{B}]_j - \bar{\bar{x}})^2$$

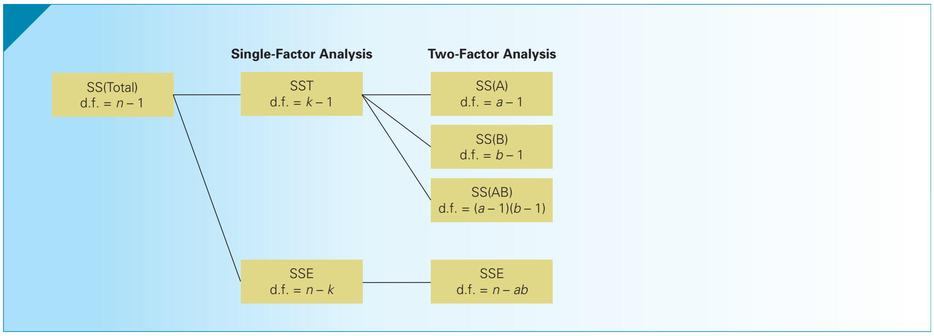
$$\text{SS(AB)} = r\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}[\text{AB}]_{ij} - \bar{x}[\text{A}]_i - \bar{x}[\text{B}]_j + \bar{\bar{x}})^2$$

$$\text{SSE} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(x_{ijk} - \bar{x}[\text{AB}]_{ij})^2$$

To compute SS(A), we calculate the sum of the squared differences between the factor A level means, which are denoted $\bar{x}[\text{A}]_i$, and the grand mean, $\bar{\bar{x}}$. The sum of squares for factor B, SS(B), is defined similarly. The interaction sum of squares, SS(AB), is calculated by taking each treatment mean (a treatment consists of a combination of a level of factor A and a level of factor B), subtracting the factor A level mean, subtracting the factor B level mean, adding the grand mean, squaring this quantity, and adding. The sum of squares for error, SSE, is calculated by subtracting the treatment means from the observations, squaring, and adding.

To test for each possibility, we conduct several F-tests similar to the one performed in Section 14.1. Figure 14.4 illustrates the partitioning of the total sum of squares that leads to the F-tests. We've included in this figure the partitioning used in the one-way study. When the one-way analysis of variance allows us to infer that differences between the treatment means exist, we continue our analysis by partitioning the treatment sum of squares into three sources of variation. The first is sum of squares for factor A, which we label SS(A), which measures the variation between the levels of factor A. Its degrees of freedom are $a - 1$. The second is the sum of squares for factor B, whose degrees of freedom are $b - 1$. SS(B) is the variation between the levels of factor B. The interaction sum of squares is labeled SS(AB), which is a measure of the amount of variation between the combinations of factors A and B; its degrees of freedom are $(a - 1) \times (b - 1)$. The sum of squares for error is SSE, and its degrees of freedom are $n - ab$. (Recall that n is the total sample size, which in this experiment is $n = abr$.) Notice that SSE and its number of degrees of freedom are identical in both partitions. As in the previous experiment, SSE is the variation within the treatments.

FIGURE **14.4** Partitioning SS(Total) in Single-Factor and Two-Factor Analysis of Variance



F-Tests Conducted in Two-Factor Analysis of Variance

Test for Differences between the Levels of Factor A

H_0: The means of the a levels of factor A are equal

H_1: At least two means differ

$$\text{Test statistic: } F = \frac{\text{MS(A)}}{\text{MSE}}$$

Test for Differences between the Levels of Factor B

H_0: The means of the b levels of factor B are equal

H_1: At least two means differ

$$\text{Test statistic: } F = \frac{\text{MS(B)}}{\text{MSE}}$$

Test for Interaction between Factors A and B

H_0: Factors A and B do not interact to affect the mean responses

H_1: Factors A and B do interact to affect the mean responses

$$\text{Test statistic: } F = \frac{\text{MS(AB)}}{\text{MSE}}$$

Required Conditions

1. The distribution of the response is normally distributed.
2. The variance for each treatment is identical.
3. The samples are independent.

As in the two previous experimental designs of the analysis of variance, we summarize the results in an ANOVA table. Table 14.8 depicts the general form of the table for the complete factorial experiment.

TABLE **14.8** ANOVA Table for the Two-Factor Experiment

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-STATISTIC
Factor A	$a - 1$	SS(A)	MS(A) = SS(A)/(a − 1)	F = MS(A)/MSE
Factor B	$b - 1$	SS(B)	MS(B) = SS(B)/(b − 1)	F = MS(B)/MSE
Interaction	$(a - 1)(b - 1)$	SS(AB)	MS(AB) = SS(AB)/[(a − 1)(b − 1)]	F = MS(AB)/MSE
Error	$n - ab$	SSE	MSE = SSE/(n − ab)	
Total	$n - 1$	SS(Total)		

We'll illustrate the techniques using the data in Example 14.4. All calculations will be performed by Excel and Minitab.

EXCEL

	A	B	C	D	E	F	G
1	Anova: Two-Factor with Replication						
2							
3	SUMMARY	Male	Female	Total			
4	*Less than HS*						
5	Count	10	10	20			
6	Sum	126	115	241			
7	Average	12.6	11.5	12.1			
8	Variance	8.27	8.28	8.16			
9							
10	*High School*						
11	Count	10	10	20			
12	Sum	110	112	222			
13	Average	11.0	11.2	11.1			
14	Variance	8.67	9.73	8.73			
15							
16	*Less than Bachelor's*						
17	Count	10	10	20			
18	Sum	106	94	200			
19	Average	10.6	9.4	10.0			
20	Variance	11.6	16.49	13.68			
21							
22	*Bachelor's or more*						
23	Count	10	10	20			
24	Sum	90	81	171			
25	Average	9.0	8.1	8.6			
26	Variance	5.33	12.32	8.58			
27							
28	*Total*						
29	Count	40	40				
30	Sum	432	402				
31	Average	10.8	10.1				
32	Variance	9.50	12.77				
33							
34	ANOVA						
35	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
36	Sample	135.85	3	45.28	4.49	0.0060	2.7318
37	Columns	11.25	1	11.25	1.12	0.2944	3.9739
38	Interaction	6.25	3	2.08	0.21	0.8915	2.7318
39	Within	726.20	72	10.09			
40							
41	Total	879.55	79				

In the ANOVA table, **Sample** refers to factor B (educational level) and **Columns** refers to factor A (gender). Thus, MS(B) = 45.28, MS(A) = 11.25, MS(AB) = 2.08, and MSE = 10.09. The F-statistics are 4.49 (educational level), 1.12 (gender), and .21 (interaction).

INSTRUCTIONS

1. Type or import the data using the same format as Xm14-04a. (*Note:* You must label the rows and columns as we did.)
2. Click **Data, Data Analysis,** and **Anova:Two-Factor with Replication**.
3. Specify the **Input Range** (A1:C41). Type the number of replications in the **Rows per sample** box (10).
4. Specify a value for α (.05).

MINITAB

Two-way ANOVA: Jobs versus Gender, Education

```
Source       DF      SS       MS     F      P
Gender        1    11.25  11.2500  1.12  0.294
Education     3   135.85  45.2833  4.49  0.006
Interaction   3     6.25   2.0833  0.21  0.892
Error        72   726.20  10.0861
Total        79   879.55

S = 3.176   R-Sq = 17.44%   R-Sq(adj) = 9.41%

                   Individual 95% CIs For Mean Based on Pooled StDev
Gender   Mean   -+-------------+-------------+-------------+------------
1        10.80               (--------------------*---------------------)
2        10.05   (---------------------*-------------------)
                -+-------------+-------------+-------------+------------
                9.10          9.80         10.50         11.20

                    Individual 95% CIs For Mean Based on Pooled StDev
Education   Mean   --------+-------------+-------------+-------------+------
1          12.05                              (------------*------------)
2          11.10                       (-----------*------------)
3          10.00              (------------*-----------)
4           8.55   (-----------*------------)
                   --------+-------------+-------------+-------------+------
                          8.0           9.6          11.2          12.8
```

INSTRUCTIONS

1. Type or import the data in stacked format in three columns. One column contains the responses, another contains codes for the levels of factor A, and a third column contains codes for the levels of factor B. (Open Xm14-04b.)
2. Click **Stat, ANOVA,** and **Twoway . . . .**
3. Specify the **Responses** (Jobs), **Row factor** (Gender), and **Column factor (**Education).
4. To produce the graphics check **Display means**.

Test for Differences in Number of Jobs between Men and Women

H_0: The means of the two levels of factor A are equal
H_1: At least two means differ

$$\text{Test statistic: } F = \frac{\text{MS(A)}}{\text{MSE}}$$

Value of the test statistic: From the computer output, we have

MS(A) = 11.25, MSE = 10.09, and $F = 11.25/10.09 = 1.12$ (p-value = .2944)

There is not evidence at the 5% significance level to infer that differences in the number of jobs exist between men and women.

Test for Differences in Number of Jobs between Education Levels

H_0: The means of the four levels of factor B are equal
H_1: At least two means differ

Test statistic: $F = \dfrac{\text{MS(B)}}{\text{MSE}}$

Value of the test statistic: From the computer output, we find

MS(B) = 45.28 and MSE = 10.09. Thus, $F = 45.28/10.09 = 4.49$ (p-value = .0060).

There is sufficient evidence at the 5% significance level to infer that differences in the number of jobs exist between educational levels.

Test for Interaction between Factors A and B

H_0: Factors A and B do not interact to affect the mean number of jobs
H_1: Factors A and B do interact to affect the mean number of jobs

Test statistic: $F = \dfrac{\text{MS(AB)}}{\text{MSE}}$

Value of the test statistic: From the printouts,

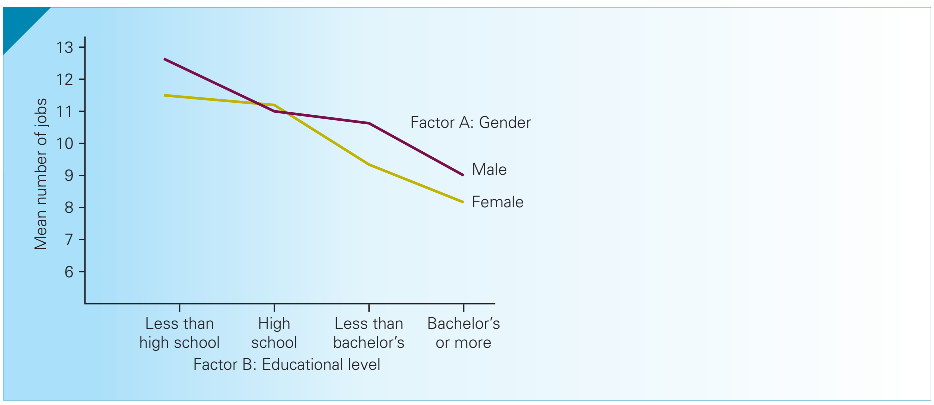
MS(AB) = 2.08, MSE = 10.09, and $F = 2.08/10.09 = .21$ (p-value = .8915).

There is not enough evidence to conclude that there is an interaction between gender and education.

INTERPRET

Figure 14.5 is a graph of the mean responses for each of the eight treatments. As you can see, there are small (not significant) differences between males and females. There are significant differences between men and women with different educational backgrounds. Finally, there is no interaction.

FIGURE **14.5** Mean Responses for Example 14.4



What Is Interaction?

To more fully understand interaction we have changed the sample associated with men who have not finished high school (Treatment 1). We subtracted 6 from the original numbers so that the sample in treatment 1 is

4, 3, 6, 10, 8, 11, 7, 3, 5, 9

The new data are stored in Xm14-04c (Excel format) and Xm14-04d (Minitab format).The mean is 6.6. Here are the Excel and Minitab ANOVA tables.

EXCEL

	A	B	C	D	E	F	G
35	ANOVA						
36	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
37	Sample	75.85	3	25.28	2.51	0.0657	2.7318
38	Columns	11.25	1	11.25	1.12	0.2944	3.9739
39	Interaction	120.25	3	40.08	3.97	0.0112	2.7318
40	Within	726.20	72	10.09			
41							
42	Total	933.55	79				

MINITAB

Two-way ANOVA: Jobs versus Gender, Education

Source	DF	SS	MS	F	P
Gender	1	11.25	11.2500	1.12	0.294
Education	3	75.85	25.2833	2.51	0.066
Interaction	3	120.25	40.0833	3.97	0.011
Error	72	726.20	10.0861		
Total	79	933.55			

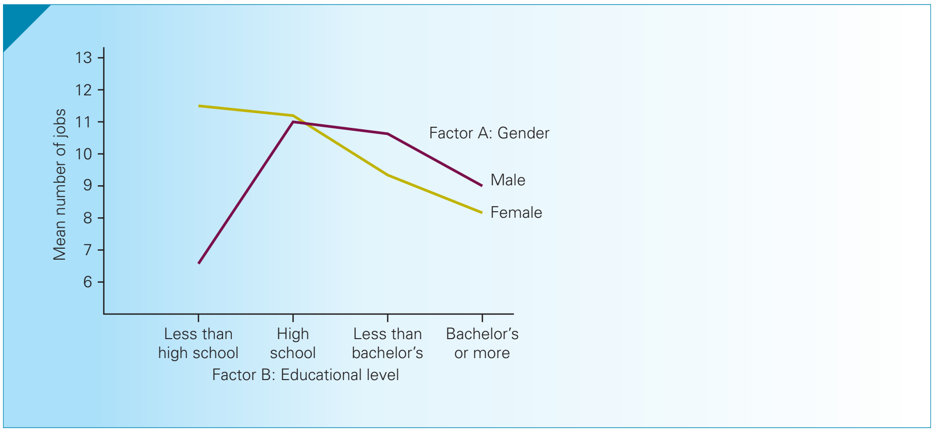
INTERPRET

In this example there is not enough evidence (at the 5% significance level) to infer that there are differences between men and women and between the educational levels. However, there is sufficient evidence to conclude that there is interaction between gender and education.

	Male	Female
Less than high school	6.6	11.5
High school	11.0	11.2
Less than bachelor's	10.6	9.4
Bachelor's or more	9.0	8.1

Compare Figures 14.5 and 14.6. In Figure 14.5, the lines joining the response means for males and females are quite similar. In particular we see that the lines are almost parallel. However, in Figure 14.6 the lines are no longer almost parallel. It is apparent that

FIGURE **14.6** Mean Responses for Example 14.4a



the mean of treatment 1 is smaller; the pattern is different. For whatever reason, in this case men with less than high school have a smaller number of jobs.

Conducting the Analysis of Variance for the Complete Factorial Experiment

In addressing the problem outlined in Example 14.4, we began by conducting a one-way analysis of variance to determine whether differences existed between the eight treatment means. This was done primarily for pedagogical reasons to enable you to see that when the treatment means differ, we need to analyze the reasons for the differences. However, in practice, we generally do not conduct this test in the complete factorial experiment (although it should be noted that some statistics practitioners prefer this "two-stage" strategy). We recommend that you proceed directly to the two-factor analysis of variance.

In the two versions of Example 14.4, we conducted the tests of each factor and then the test for interaction.

However, if there is evidence of interaction, the tests of the factors are irrelevant. There may or may not be differences between the levels of factor A and of the levels of factor B. Accordingly, we change the order of conducting the F-tests.

Order of Testing in the Two-Factor Analysis of Variance

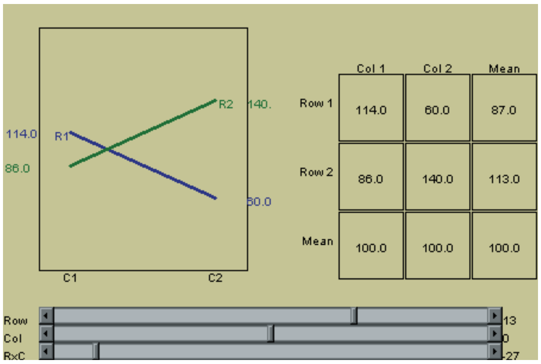
Test for interaction first. If there is enough evidence to infer that there is interaction, do not conduct the other tests.

If there is not enough evidence to conclude that there is interaction, proceed to conduct the F-tests for factors A and B.

SEEING STATISTICS

applet 17 Plots of Two-Way ANOVA Effects

This applet provides a graph similar to those in Figures 14.5 and 14.6. There are three sliders: one for rows, one for columns, and one for interaction. Moving the top slider changes the difference between the row means. The second slider changes the difference between the column means. The third slider allows us to see the effects of interaction.



Applet Exercises

Label the columns factor A and the rows factor B. Move the sliders to arrange for each of the following differences. Describe what the resulting figure tells you about differences between levels of factor A, levels of factor B, and interaction.

	ROW	COL	R × C
17.1	−30	0	0
17.2	0	25	0
17.3	0	0	−20
17.4	25	−30	0
17.5	30	0	30
17.6	30	0	−30
17.7	0	20	20
17.8	0	20	−20
17.9	30	30	30
17.10	30	30	−30

Developing an Understanding of Statistical Concepts

You may have noticed that there are similarities between the two-factor experiment and the randomized block experiment. In fact, when the number of replicates is one, the calculations are identical. (Minitab uses the same command.) This raises the question, What is the difference between a factor in a multifactor study and a block in a randomized block experiment? In general, the difference between the two experimental designs is that in the randomized block experiment, blocking is performed specifically to reduce variation, whereas in the two-factor model the effect of the factors on the response variable is of interest to the statistics practitioner. The criteria that define the blocks are always characteristics of the experimental units. Consequently, factors that are characteristics of the experimental units will be treated not as factors in a multifactor study, but as blocks in a randomized block experiment.

Let's review how we recognize the need to use the procedure described in this section.

Factors That Identify the Independent Samples Two-Factor Analysis of Variance

1. **Problem objective**: Compare two or more populations (populations are defined as the combinations of the levels of two factors)
2. **Data type**: Interval
3. **Experimental design**: Independent samples

EXERCISES

14.68 A two-factor analysis of variance experiment was performed with $a = 3$, $b = 4$, and $r = 20$. The following sums of squares were computed:

SS(Total) = 42,450 SS(A) = 1,560
SS(B) = 2,880 SS(AB) = 7,605

a. Determine the one-way ANOVA table.
b. Test at the 1% significance level to determine whether differences exist between the 12 treatments.
c. Conduct whatever test you deem necessary at the 1% significance level to determine whether there are differences between the levels of factor A, the levels of factor B, or interaction between factors A and B.

14.69 A statistics practitioner conducted a two-factor analysis of variance experiment with a = 4, $b = 3$, and $r = 8$. The sums of squares are listed here:

SS(Total) = 9,420 SS(A) = 203 SS(B) = 859
SS(AB) = 513

a. Test at the 5% significance level to determine whether factors A and B interact.
b. Test at the 5% significance level to determine whether differences exist between the levels of factor A.
c. Test at the 5% significance level to determine whether differences exist between the levels of factor B.

14.70 Xr14-70 The following data were generated from a 2×2 factorial experiment with three replicates:

	Factor B	
Factor A	**1**	**2**
1	6	12
	9	10
	7	11
2	9	15
	10	14
	5	10

a. Test at the 5% significance level to determine whether factors A and B interact.
b. Test at the 5% significance level to determine whether differences exist between the levels of factor A.
c. Test at the 5% significance level to determine whether differences exist between the levels of factor B.

14.71 Xr14-71 The data shown here were taken from a 2×3 factorial experiment with four replicates:

	Factor B	
Factor A	**1**	**2**
1	23	20
	18	17
	17	16
	20	19
2	27	29
	23	23
	21	27
	28	25
3	23	27
	21	19
	24	20
	16	22

a. Test at the 5% significance level to determine whether factors A and B interact.
b. Test at the 5% significance level to determine whether differences exist between the levels of factor A.
c. Test at the 5% significance level to determine whether differences exist between the levels of factor B.

14.72 Xr14-72 Refer to Example 14.4. We've revised the data by adding 2 to each of the numbers of the men. What do these data tell you?

14.73 Xr14-73 Refer to Example 14.4. We've altered the data by subtracting 4 from the numbers of treatment 8. What do these data tell you?

Applications

The following exercises require the use of a computer and software.

14.74 Xr14-74 Refer to Exercise 14.10. Suppose that the experiment is redone in the following way. Thirty taxpayers fill out each of the four forms. However, 10 taxpayers in each group are in the lowest income bracket, 10 are in the next income bracket, and the remaining 10 are in the highest bracket. The amount of time needed to complete the returns is recorded.

Column 1: Group number
Column 2: Times to complete form 1 (first 10 rows = low income, next 10 rows = next income bracket, and last 10 rows = highest bracket)
Column 3: Times to complete form 2 (same format as column 2)
Column 4: Times to complete form 3 (same format as column 2)
Column 5: Times to complete form 4 (same format as column 2)

a. How many treatments are there in this experiment?
b. How many factors are there? What are they?
c. What are the levels of each factor?
d. Is there evidence at the 5% significance level of interaction between the two factors?
e. Can we conclude at the 5% significance level that differences exist between the four forms?
f. Can we conclude at the 5% significance level that taxpayers in different brackets require different amounts of time to complete their tax forms?

14.75 Xr14-75 Detergent manufacturers frequently make claims about the effectiveness of their products. A consumer protection service decided to test the five best-selling brands of detergent, where each manufacturer claims that its product produces the "whitest whites" in all water temperatures. The experiment was conducted in the following way. One hundred fifty white sheets were equally soiled. Thirty sheets were washed in each brand—l0 with cold water, 10 with warm water, and 10 with hot water. After washing, the "whiteness" scores for each sheet were measured with laser equipment.

Column 1: Water temperature code
Column 2: Scores for detergent 1 (first 10 rows = cold water, middle 10 rows = warm, and last 10 rows = hot)
Column 2: Scores for detergent 2 (same format as column 2)
Column 3: Scores for detergent 3 (same format as column 2)
Column 4: Scores for detergent 4 (same format as column 2)
Column 5: Scores for detergent 5 (same format as column 2)

a. What are the factors in this experiment?
b. What is the response variable?
c. Identify the levels of each factor.
d. Perform a statistic analysis using a 5% significance level to determine whether there is sufficient statistical evidence to infer that there are differences in whiteness scores between the five detergents, differences in whiteness scores between the three water temperatures, or interaction between detergents and temperatures.

14.76 Xr14-76 Headaches are one of the most common, but least understood, ailments. Most people get headaches several times per month; over-the-counter medication is usually sufficient to eliminate their pain. However, for a significant proportion of people, headaches are debilitating and make their lives almost unbearable. Many such people have investigated a wide spectrum of possible treatments, including narcotic drugs, hypnosis, biofeedback, and acupuncture, with little or no success. In the last few years, a promising new treatment has been developed. Simply described, the treatment involves a series of injections of a local anesthetic to the occipital nerve (located in the back of the neck). The current treatment procedure is to schedule the injections once a week for 4 weeks. However, it has been suggested that another procedure may be better—one that features one injection every other day for a total of four injections. In addition, some physicians recommend other combinations of drugs that may increase the effectiveness of the injections. To analyze the problem, an experiment was organized. It was decided to test for a difference between the two schedules of injection and to determine whether there are differences between four drug mixtures. Because of the possibility of an interaction between the schedule and the drug, a complete factorial experiment was chosen. Five headache patients were randomly selected for each combination of schedule and drug. Forty patients were treated, and each was asked to report the frequency, duration, and severity of his or her headache prior to treatment and for the 30 days following the last injection. An index ranging from 0 to 100 was constructed for each patient, with 0 indicating no headache pain and 100 specifying the worst headache pain. The improvement in the headache index for each patient was recorded and reproduced in the accompanying table. (A negative value indicates a worsening condition.) (The author is grateful to Dr. Lorne Greenspan for his help in writing this example.)

a. What are the factors in this experiment?
b. What is the response variable?
c. Identify the levels of each factor.
d. Analyze the data and conduct whichever tests you deem necessary at the 5% significance level to determine whether there is sufficient statistical evidence to infer that there are differences in the improvement in the headache index between the two schedules, differences in the improvement in the headache index between the four drug mixtures, or interaction between schedules and drug mixtures.

Improvement in Headache Index

	Drug Mixture			
Schedule	**1**	**2**	**3**	**4**
One Injection	17	24	14	10
Every Week	6	15	9	−1
(Four Weeks)	10	10	12	0
	12	16	0	3
	14	14	6	−1
One Injection	18	−2	20	−2
Every Two Days	9	0	16	7
(Four Days)	17	17	12	10
	21	2	17	6
	15	6	18	7

14.77 Xr14-77 Most college instructors prefer to have their students participate actively in class. Ideally, students will ask their professor questions and answer their professor's questions, making the classroom experience more interesting and useful. Many professors seek ways to encourage their students to participate in class. A statistics professor at a community college in upper New York state believes that several external factors affect student participation. He believes that the time of day and the configuration of seats are two such factors. Consequently, he organized the following experiment. Six classes of about 60 students each were scheduled for one semester. Two classes were scheduled at 9 A.M., two at 1 P.M., and two at 4 P.M. At each of the three times, one of the classes was assigned to a room where the seats were arranged in rows of 10 seats. The other class was a U-shaped, tiered room, where students not only face the instructor but also face their fellow students. In each of the six classrooms, over 5 days, student participation was measured by counting the number of times students asked and answered questions. These data are displayed in the accompanying table.

a. How many factors are there in this experiment? What are they?
b. What is the response variable?
c. Identify the levels of each factor.
d. What conclusions can the professor draw from these data?

	Time		
Class Configuration	**9 A.M.**	**1 P.M.**	**4 P.M.**
Rows	10	9	7
	7	12	12
	9	12	9
	6	14	20
	8	8	7
U-Shape	15	4	7
	18	4	4
	11	7	9
	13	4	8
	13	6	7

14.6 (Optional) Applications in Operations Management: Finding and Reducing Variation

In the introduction to Example 12.3, we pointed out that variation in the size, weight, or volume of a product's components causes the product to fail or not function properly. Unfortunately, it is impossible to eliminate all variation. Designers of products and the processes that make the products understand this phenomenon. Consequently, when they specify the length, weight, or some other measurable characteristic of the product, they allow for some variation, which is called the *tolerance*. For example, the diameters of the piston rings of a car are supposed to be .826 millimeter (mm) with a tolerance of .006 mm; that is, the product will function provided that the diameter is between $.826 - .006 = .820$ and $.826 + .006 = .832$ mm. These quantities are called the *lower* and *upper specification limits* (LSL and USL), respectively.

Suppose that the diameter of the piston rings is actually a random variable that is normally distributed with a mean of .826 and a standard deviation of .003 mm. We can compute the probability that a piston ring's diameter is between the specification limits. Thus,

$$\begin{aligned} P(.820 < X < .832) &= P\left(\frac{.820 - .826}{.003} < \frac{X - \mu}{\sigma} < \frac{.832 - .826}{.003}\right) \\ &= P(-2.0 < Z < 2.0) \\ &= .9772 - .0228 \\ &= .9544 \end{aligned}$$

The probability that the diameter does not meet specifications is $1 - .9544 = .0456$. This probability is a measure of the process capability.

If we can decrease the standard deviation, a greater proportion of piston rings will have diameters that meet specification. Suppose that the operations manager has decreased the diameter's standard deviation to .002. The proportion of piston rings that do not meet specifications is .0026. When the probabilities are quite low, we express the probabilities as the number of defective units per million or per billion. Thus, if the standard deviation is .002, the number of defective piston rings is expected to be 2,600 per million. The goal of many firms is to reduce the standard deviation so that the lower specification and upper specification limits are at least 6 standard deviations away from the mean. If the standard deviation is .001, the proportion of nonconforming piston rings is $1 - P(-6 < Z < 6)$, which is 2 per billion. (Incidentally, this figure is often erroneously quoted as 3.4 per million.) The goal is called *six sigma*. Figure 14.7 depicts the proportion of conforming and nonconforming piston rings for $\sigma = .003$, .002, and .001.

Another way to measure how well the process works is the process capability index, denoted by C_p, *which is defined as*

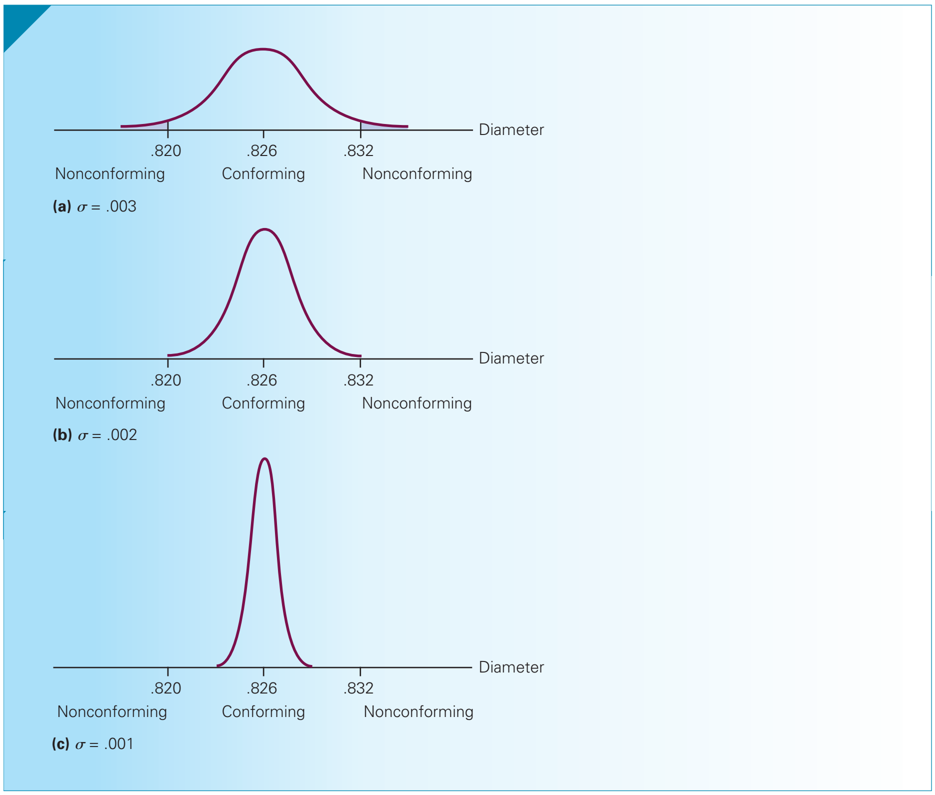
$$C_p = \frac{\text{USL} - \text{LSL}}{6\sigma}$$

Thus, in the illustration USL = .832 and LSL = .820. If the standard deviation is .002 then

$$C_p = \frac{\text{USL} - \text{LSL}}{6\sigma} = \frac{.832 - .820}{6(.002)} = 1.0$$

The larger the process capability index, the more capable is the process in meeting specifications. A value of 1.0 describes a production process where the specification limits are equal to 3 standard deviations above and below the mean. A process capability index of 2.0 means that the upper and lower limits are 6 standard deviations above and below the mean. This is the goal for many firms.

FIGURE **14.7** **Proportion of Conforming and Nonconforming Piston Rings**



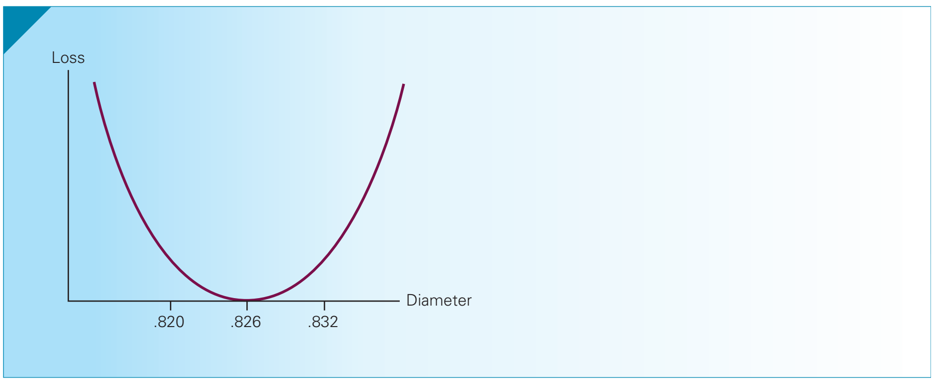
In practice, the standard deviation must be estimated from the data. We will address this issue again in Chapter 21.

Taguchi Loss Function

Historically, operations managers applied the "goalpost" philosophy, a name derived from the game of football. If the ball is kicked *anywhere* between the goalposts, the kick is equally as successful as one that is in the center of the goalposts. Under this philosophy, a piston ring that has a diameter of .821 works as well as one that is exactly .826. In other words, the company sustains a loss only when the product falls outside the goalposts. Products that lie between the goalposts suffer no financial loss. For many firms, this philosophy has now been replaced by the Taguchi loss function (named for Genichi Taguchi, a Japanese statistician whose ideas and techniques permeate any discussion of statistical applications in quality management).

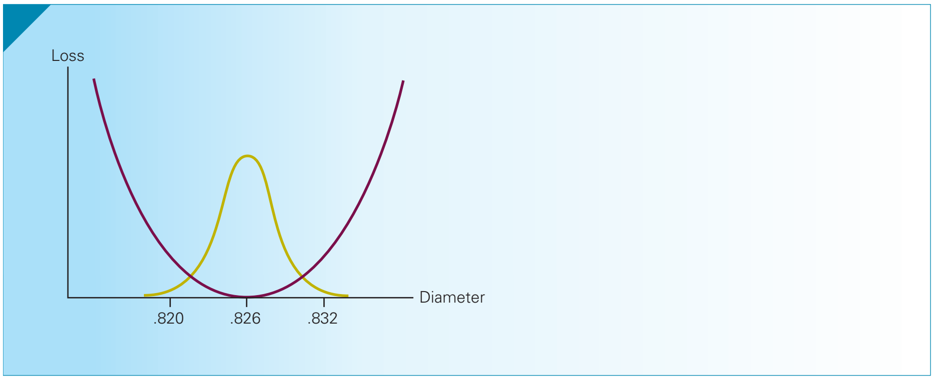
Products whose length or weight fall within the tolerances of their specifications do not all function in exactly the same way. There is a difference between a product that barely falls between the goalposts and one that is in the exact center. The Taguchi loss function recognizes that any deviation from the target value results in a financial loss. In addition, the farther the product's variable is from the target value, the greater the loss. The piston ring described previously is specified to have a diameter of exactly .826 mm, an amount specified by the manufacturer to work at the optimum level. Any deviation will cause that part and perhaps other parts to wear out prematurely. Although customers will not know the reason for the problem, they will know that the unit had to be replaced. The greater the deviation, the more quickly the part will wear and need replacing. If the part is under warranty, the company will incur a loss in replacing it. If the warranty has expired, customers will have to pay to replace the unit, causing some degree of displeasure that may result in them buying another company's product in the future. In either case, the company loses money. Figure 14.8 depicts the loss function. As you can see, any deviation from the target value results in some loss, with large deviations resulting in larger losses.

FIGURE **14.8** Taguchi Loss Function



Management scientists have shown that the loss function can be expressed as a function of the production process mean and variance. In Figure 14.9 we describe a normal distribution of the diameter of the machined part with a target value of .826 mm. When

FIGURE 14.9 Taguchi Loss Function and the Distribution of Piston Rings



the mean of the distribution is .826, any loss is caused by the variance. The statistical techniques introduced in Chapter 21 are usually employed to center the distribution on the target value. However, reducing the variance is considerably more difficult. To reduce variation, it is necessary to first find the sources of variation. We do so by conducting experiments. The principles are quite straightforward, drawing on the concepts developed in the previous section.

An important function of operations management is production design in which decisions are made about how a product is manufactured. The objective is to produce the highest quality product at a reasonable cost. This objective is achieved by choosing the machines, materials, methods, and "manpower" (personnel), the so-called 4 M's. By altering some or all of these elements, the operations manager can alter the size, weight, or volume and, ultimately, the quality of the product.

EXAMPLE 14.5

DATA
Xm14-05

Causes of Variation

A critical component in an aircraft engine is a steel rod that must be 41.387 cm long. The operations manager has noted that there has been some variation in the lengths. In some cases, the steel rods had to be discarded or reworked because they were either too short or too long. The operations manager believes that some of the variation is caused by the way the production process has been designed. Specifically, he believes that the rods vary from machine to machine and from operator to operator. To help unravel the truth, he organizes an experiment. Each of the three operators produces five rods on each of the four machines. The lengths are measured and recorded. Determine whether the machines or the operators (or both) are indeed sources of variation.

SOLUTION

IDENTIFY

The response variable is the length of the rods. The two factors are the operators and the machines. There are three levels of operators and four levels of machines. The model we employ is the two-factor model with interaction. The computer output is shown here.

COMPUTE

EXCEL

	A	B	C	D	E	F	G
1	Anova: Two-Factor With Replication						
2							
3	ANOVA						
4	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
5	Sample	0.0151	2	0.0076	6.98	0.0022	3.1907
6	Columns	0.0034	3	0.0011	1.04	0.3856	2.7981
7	Interaction	0.0046	6	0.0008	0.71	0.6394	2.2946
8	Within	0.0520	48	0.0011			
9							
10	Total	0.0751	59				

MINITAB

Xm14-00a stores the data in Minitab format.

Two-way Analysis of Variance

Analysis of Variance for Rods

Source	DF	SS	MS	F	P
Machines	3	3363	1121	1.04	0.386
Operator	2	15133	7566	6.98	0.002
Interaction	6	4646	774	0.71	0.639
Error	48	51995	1083		
Total	59	75137			

INTERPRET

The test for interaction yields $F = .71$ and a p-value of .6394. There is not enough evidence to infer that the two factors interact. The F-statistic for the operator factor (Sample) is 6.98 (p-value = .0022). The F-statistic for the machine factor (Columns) is 1.04 (p-value = .3856). We conclude that there are differences only between the levels of the operators. Thus, the only source of variation here is the different operators. The operations manager can now focus on reducing or eliminating this variation. For example, the manager may use only one operator in the future or investigate why the operators differ.

The causes of variation example that opened this chapter illustrate this strategy. Because we have limited our discussion to the two-factor model, the example features this experimental design. It should be understood, however, that more complicated models are needed to fully investigate sources of variation.

Design of Experiments and Taguchi Methods

In the example just discussed, the experiment used only two factors. In practice, there are frequently many more factors. The problem is that the total number of treatments or combinations can be quite high, making any experimentation both time consuming and expensive. For example, if there are 10 factors each with 2 levels, the number of treatments is $2^{10} = 1{,}024$. If we measure each treatment with 10 replicates, the number of observations, 10,240, makes this experiment prohibitive. Fortunately, it is possible to reduce this number considerably. Through the use of *orthogonal arrays*, we can conduct *fractional factorial experiments* that can produce useful results at a small fraction of the cost. The experimental designs and statistical analyses are beyond the level of this book. Interested readers can find a variety of books at different levels of mathematical and statistical sophistication to learn more about this application.

EXERCISES

Applications

The following exercises require the use of a computer and software. Use a 5% significance level.

14.78 Xr14-78 The headrests on a car's front seats are designed to protect the driver and front-seat passenger from whiplash when the car is hit from behind. The frame of the headrest is made from metal rods. A machine is used to bend the rod into a U shape exactly 440 millimeters wide. The width is critical; too wide or too narrow, and the rod won't fit into the holes drilled into the car seat frame. The company has experimented with several different metal alloys in the hope of finding a material that will result in more headrest frames that fit. Another possible source of variation is the machines used. To learn more about the process, the operations manager conducts an experiment. Both of the machines are used to produce 10 headrests from each of the five metal alloys now being used. Each frame is measured, and the data (in millimeters) are recorded using the format shown here. Analyze the data to determine whether the alloys, machines, or both are sources of variation.

Column 1: Machine 1, rows 1 to 10 alloy A, rows 11 to 20 alloy B
Column 2: Machine 2, rows 1 to 10 alloy A, rows 11 to 20 alloy B

14.79 Xr14-79 A paint manufacturer is attempting to improve the process that fills the 1-gallon containers. The foreperson has suggested that the nozzle can be made from several different alloys. Furthermore, the way that the process "knows" when to stop the flow of paint can be accomplished in two ways: by setting a predetermined amount or by measuring the amount of paint already in the can. To determine what factors lead to variation, an experiment is conducted. For each of the four alloys that could be used to make the nozzles and the two measuring devices, five cans are filled. The amount of paint in each container is precisely measured. The data in liters were recorded in the following way:

Column 1: Device 1, rows 1 to 5 alloy A, rows 6 to 10 alloy B, etc.
Column 2: Device 2, rows 1 to 5 alloy A, rows 6 to 10 alloy B, etc.

Can we infer that the alloys, the measuring devices, or both are sources of variation?

14.80 Xr14-80 The marketing department of a firm that manufactures office furniture has ascertained that there is a growing market for a specialized desk that houses the various parts of a computer system. The operations manager is summoned to put together a plan that will produce high-quality desks at low cost. The characteristics of the desk have been dictated by the marketing department, which has specified the material that the desk will be made from and the machines used to produce the parts. However, three methods can be utilized. Moreover, because of the complexity of the operation, the manager realizes that it is possible that different skill levels of the workers can yield different results. Accordingly, he organized an experiment. Workers from each of three skill levels were chosen. These groups were further divided into two subgroups. Each subgroup assembled the desks using methods A and B. The amount of time taken to assemble each of eight desks was recorded as follows. Columns 1 and 2 contain the times for methods A and B; rows 1 to 8, 9 to16, and 17 to 24 store the times for the three skill levels. What can we infer from these data?

CHAPTER SUMMARY

The analysis of variance allows us to test for differences between populations when the data are interval. The analyses of the results of three different experimental designs were presented in this chapter. The one-way analysis of variance defines the populations on the basis of one factor. The second experimental design also defines the treatments on the basis of one factor. However, the randomized block design uses data gathered by observing the results of a matched or blocked experiment (two-way analysis of variance). The third design is the two-factor experiment wherein the treatments are defined as the combinations of the levels of two factors. All the analyses of variance are based on partitioning the total sum of squares into sources of variation from which the mean squares and *F*-statistics are computed.

In addition, we introduced three multiple comparison methods that allow us to determine which means differ in the one-way analysis of variance.

Finally, we described an important application in operations management that employs the analysis of variance.

IMPORTANT TERMS

Analysis of variance 526
Treatment means 526
One-way analysis of variance 526
Response variable 528
Responses 528
Experimental units 528
Factor 528
Level 528
Between-treatments variation 528
Sum of squares for treatments (SST) 528
Within-treatments variation 529
Sum of squares for error (SSE) 529
Mean squares 530
Mean squares for treatments 530
Mean squares for error 530
F-statistic 531
Analysis of variance (ANOVA) table 531
Total variation 532
SS(Total) 532
Completely randomized design 534
Multiple comparisons 544
Least significance difference (LSD) 546
Bonferroni adjustment 547
Tukey's multiple comparison method 548
Multifactor experiment 553
Randomized block design 554
Repeated measures 554
Two-way analysis of variance 554
Fixed-effects analysis of variance 554
Random-effects analysis of variance 554
Sum of squares for blocks 555
Factorial experiment 563
Interactions 565
Complete factorial experiment 566
Replicate 567
Balanced 567

SYMBOLS

Symbol	Pronounced	Represents
$\bar{\bar{x}}$	x double bar	Overall or grand mean
q		Studentized range
ω	Omega	Critical value of Tukey's multiple comparison method
$q_\alpha(k,\nu)$	q sub alpha k ν	Critical value of the Studentized range
n_g		Number of observations in each of k samples
$\bar{x}[T]_j$	x bar T sub j	Mean of the jth treatment
$\bar{x}[B]_i$	x bar B sub i	Mean of the ith block
$\bar{x}[AB]_{ij}$	x bar A B sub i j	Mean of the ijth treatment
$\bar{x}[A]_i$	x bar A sub i	Mean of the observations when the factor A level is i
$\bar{x}[B]_j$	x bar B sub j	Mean of the observations when the factor B level is j

FORMULAS

One-way analysis of variance

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

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

$$\text{MST} = \frac{\text{SST}}{k-1}$$

$$\text{MST} = \frac{\text{SSE}}{n-k}$$

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

Least significant difference comparison method

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

Tukey's multiple comparison method

$$\omega = q_\alpha(k,\nu)\sqrt{\frac{\text{MSE}}{n_g}}$$

Two-way analysis of variance (randomized block design of experiment)

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

$$\text{SST} = \sum_{j=1}^{k} b(\bar{x}[T]_j - \bar{\bar{x}})^2$$

$$\text{SSB} = \sum_{i=1}^{b} k(\bar{x}[B]_i - \bar{\bar{x}})^2$$

$$\text{SSE} = \sum_{j=1}^{k}\sum_{i=1}^{b}(x_{ij} - \bar{x}[T]_j - \bar{x}[B]_i + \bar{\bar{x}})^2$$

$$\text{MST} = \frac{\text{SST}}{k-1}$$

$$\text{MSB} = \frac{\text{SSB}}{b-1}$$

$$\text{MSE} = \frac{\text{SSE}}{n-k-b+1}$$

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

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

Two-factor analysis of variance

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

$$\text{SS(A)} = rb\sum_{i=1}^{a}(\bar{x}[A]_i - \bar{\bar{x}})^2$$

$$\text{SS(B)} = ra\sum_{j=1}^{b}(\bar{x}[B]_j - \bar{\bar{x}})^2$$

$$\text{SS(AB)} = r\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{x}[AB]_{ij} - \bar{x}[A]_i - \bar{x}[B]_j + \bar{\bar{x}})^2$$

$$\text{SSE} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{r}(x_{ijk} - \bar{x}[AB]_{ij})^2$$

$$\text{MS(A)} = \frac{\text{SS(A)}}{a-1}$$

$$\text{MS(B)} = \frac{\text{SS(B)}}{b-1}$$

$$\text{MS(AB)} = \frac{\text{SS(AB)}}{(a-1)(b-1)}$$

$$F = \frac{\text{MS(A)}}{\text{MSE}}$$

$$F = \frac{\text{MS(B)}}{\text{MSE}}$$

$$F = \frac{\text{MS(AB)}}{\text{MSE}}$$

COMPUTER OUTPUT AND INSTRUCTIONS

Technique	Excel	Minitab
One-way ANOVA	533	533
Multiple comparisons (LSD, Bonferroni adjustment, and Tukey)	549	550
Two-way (randomized block) ANOVA	558	558
Two-factor ANOVA	570	571

CHAPTER EXERCISES

The following exercises require the use of a computer and software. Use a 5% significance level.

14.81 Xr14-81 Each year billions of dollars are lost because of worker injuries on the job. Costs can be decreased if injured workers can be rehabilitated quickly. As part of an analysis of the amount of time taken for workers to return to work, a sample was taken of male blue-collar workers aged 35 to 45 who suffered a common wrist fracture. The researchers believed that the mental and physical condition of the individual affects recovery time. Each man was asked to complete a questionnaire that measured whether he tended to be optimistic or pessimistic. The men's physical condition was also evaluated and categorized as very physically fit, average, or in poor condition. The number of days until the wrist returned to full function was measured for each individual. These data were recorded in the following way:

Column 1: Time to recover for optimists (rows 1–10) = very fit, rows 11–20 = in average condition, rows 21–30 = poor condition

Column 2: Time to recover for pessimists (same format as column 1)

a. What are the factors in this experiment? What are the levels of each factor?
b. Can we conclude that pessimists and optimists differ in their recovery times?
c. Can we conclude that physical condition affects recovery times?

14.82 Xr14-82 In the past decade, American companies have spent nearly $1 trillion on computer systems. However, productivity gains have been quite small. During the 1980s, productivity in U.S. service industries (where most computers are used) grew by only .7% annually. In the 1990s, this figure rose to 1.5%. (Source: *New York Times Service*, February 22, 1995.) The problem of small productivity increases may be caused by the trouble employees experience in learning how to use the computer. Suppose that in an experiment to examine the problem, 100 firms were studied. Each company had bought a new computer system 5 years earlier. The companies reported their increase in productivity over the 5-year period and were also classified as offering extensive employee training, some employee training, little employee training, or no formal employee training in the use of computers. (There were 25 firms in each group.)

a. Can we conclude that differences in productivity gain exist between the four groups of companies?
b. If there are differences, what are they?

14.83 Xr14-83 The possible imposition of a residential property tax has been a sensitive political issue in a large city that consists of five boroughs. Currently, property tax is based on an assessment system that dates back to 1950. This system has produced numerous inequities whereby newer homes tend to be assessed at higher values than older homes. A new system based on the market value of the house has been proposed. Opponents of the plan argue that residents of some boroughs would have to pay considerably more on the average, while residents of other boroughs would pay less. As part of a study examining this issue, several homes in each borough were assessed under both plans. The percentage increase (a decrease is represented by a negative increase) in each case was recorded.

a. Can we conclude that there are differences in the effect the new assessment system would have on the five boroughs?
b. If differences exist, which boroughs differ? Use Tukey's multiple comparison method.
c. What are the required conditions for your conclusions to be valid?
d. Are the required conditions satisfied?

14.84 Xr14-84 The editor of the student newspaper was in the process of making some major changes in the newspaper's layout. He was also contemplating changing the typeface of the print used. To help himself make a decision, he set up an experiment in which 20 individuals were asked to read four newspaper pages, with each page printed in a different typeface. If the reading speed differed, then the typeface that was read fastest would be used. However, if there was not enough evidence to allow the editor to conclude that such differences existed, the current typeface would be continued. The times (in seconds) to completely read one page were recorded. What should the editor do?

14.85 Xr14-85 In marketing children's products, it is extremely important to produce television commercials that hold the attention of the children who view them. A psychologist hired by a marketing research firm wants to determine whether differences in attention span exist between children watching advertisements for different types of products. One hundred fifty children less than 10 were recruited for an experiment. One-third watched a 60-second commercial for a new computer game, one-third watched a commercial for a breakfast cereal, and one-third watched a commercial for children's clothes. Their attention spans (in seconds) were measured and recorded. Do these data provide enough evidence to conclude that there are differences in attention span between the three products advertised?

14.86 Xr14-86 On reconsidering the experiment in Exercise 14.85, the psychologist decides that the age of the child may influence the attention span. Consequently, the experiment is redone in the following way. Three children of each age (10 year olds, 9 year olds, 8 year olds, 7 year olds, 6 year olds, 5 year olds, and 4 year olds) are randomly assigned to watch one of the commercials, and their attention spans are measured. Do the results indicate that there are differences in the abilities of the products advertised to hold children's attention?

14.87 Xr14-87 It is important for salespeople to be knowledgeable about how people shop for certain products. Suppose that a new car salesperson believes that the age and gender of a car shopper affect the way he or she makes an offer on a car. He records the initial offers made by a group of men and women shoppers on a $30,000 Honda Accord. Besides the gender of the shopper, the salesman also notes the age category.

The amount of money below the asking price that each person offered initially for the car was recorded using the following format: Column 1 contains the data for the less than 30 group, the first 25 rows store the results for female shoppers, and the last 25 rows are the male shoppers. Columns 2 and 3 store the data for the 30–45 and older than 45 categories, respectively. What can we conclude from these data?

14.88 Xr14-88 Many of you reading this page probably learned how to read using the whole-language method. This strategy maintains that the natural and effective way is to be exposed to whole words in context. Students learn how to read by recognizing words they have seen before. In the past generation, this has been the dominant teaching strategy throughout North America. It replaced phonics, wherein children were taught to sound out the letters to form words. The whole language method was instituted with little or no research and has been severely criticized in the past. A recent study may have resolved the question of which method should be employed. Barbara Foorman, an educational psychologist at the University of Houston described the experiment at the annual meeting of the American Association for the Advancement of Science. The subjects were 375 low-achieving, poor, first-grade students in Houston schools. The students were divided into three groups. One was educated according to the whole language philosophy, a second group was taught using a pure phonics strategy, and the third was taught employing a mixed or embedded phonics technique. At the end of the term, students were asked to read words on a list of 50 words. The number of words each child could read was recorded.

a. Can we infer that differences exist between the effects of the three teaching strategies?
b. If differences exist, identify which method appears to be best.

14.89 Xr14-89 Are babies who are exposed to music before their birth smarter than those who are not? And, if so, what kind of music is best? Researchers at the University of Wisconsin conducted an experiment with rats. The researchers selected a random sample of pregnant rats and divided the sample into three groups. Mozart works were played to one group, a second group was exposed to white noise (a steady hum with no musical elements), and the third group listened to Philip Glass music (very simple compositions). The researchers then trained the young rats to run a maze in search of food. The amount of time for the rats to complete the maze was measured for all three groups.

a. Can we infer from these data that there are differences between the three groups?
b. If there are differences, determine which group is best.

14.90 Xr14-90 Increasing tuition has resulted in some students being saddled with large debts on graduation. To examine this issue, a random sample of recent graduates was asked to report whether they had student loans; if so, how much was the debt at graduation? Those who reported they owed money were also asked whether their degrees were BAs, BScs, BBAs, or other. Can we conclude that debt levels differ between the four types of degree?

14.91 Xr14-91 Studies indicate that single male investors tend to take the most risk, whereas married female investors tend to be conservative. This raises the question, Which does best? The risk-adjusted returns for single and married men, and for single and married women were recorded. Can we infer that differences exist between the four groups of investors?

14.92 Xr14-92 Like all other fine restaurants Ye Olde Steak House in Windsor, Ontario, attempts to have three "seatings" on weekend nights. Three seatings means that each table gets three different sets of customers. Obviously, any group that lingers over dessert and coffee may result in the loss of one seating and profit for the restaurant. In an effort to determine which types of groups tend to linger, a random sample of 150 groups was drawn. For each group, the number of members and the length of time that the group stayed were recorded in the following way.

Column A: Length of time for 2 people
Column B: Length of time for 3 people
Column C: Length of time for 4 people
Column D: Length of time for more than 4 people

Do these data allow us to infer that the length of time in the restaurant depends on the size of the party?

14.93 Xr14-93 When the stock market has a large 1-day decline, does it bounce back the next day or does the bad news endure? To answer this question, an economist examined a random sample of daily changes to the Toronto Stock Index (TSE). He recorded the percent change. He classified declines as

down by less than 0.5%
down by 0.5% to 1.5%
down by 1.5% to 2.5%
down by more than 2.5%

For each of these days, he recorded the percent loss the following day. Do these data allow us to infer that there are differences in changes to the TSE depending on the loss the previous day? (This exercise is based on a study undertaken by Tim Whitehead, an economist for Left Bank Economics, a consulting firm near Paris, Ontario.)

14.94 Xr14-94 Stock market investors are always seeking the "Holy Grail," a sign that tells them the market has bottomed out or achieved its highest level. There are several indicators. One is the buy signal developed by Gerald Appel, who believed that a bottom has been reached when the difference between the weekly close of the New York Stock Exchange index and the 10-week moving average (see Chapter 20) is −4.0 points or more. Another bottom indicator is based on identifying a certain pattern in the line chart of the stock market index. As an experiment, a financial analyst randomly selected 100 weeks. For each week, he determined whether there was an Appel buy, a chart buy, or no indication. For each type of week, he recorded the percentage change over the next 4 weeks. Can we infer that the two buy indicators are not useful?

14.95 Xr14-95 Millions of North Americans spend up to several hours a day commuting to and from work. Aside from the wasted time, are there other negative effects associated with fighting traffic? A study by Statistics Canada may shed light on the issue. A random sample of adults was surveyed. Among other questions, each was asked how much time he or she slept and how much time was spent commuting. The categories for commuting time are 1 to 30 minutes, 31 to 60 minutes, and more than 60 minutes. Is there sufficient evidence to conclude that the amount of sleep differs between commuting categories?

The following exercises use data files associated with three exercises seen previously in this book.

14.96 Xr12-126* In Exercise 12.126, marketing managers for the JC Penney department store chain segmented the market for women's apparel on the basis of personal and family values. The segments are labeled Conservative, Traditional, and Contemporary. Recall that the classification was done on the basis of questionnaires. In addition to identifying the segment via the questionnaire, each woman was also asked to report family income (in $1,000s). Do these data allow us to infer that family incomes differ between the three market segments?

14.97 Xr13-21* Exercise 13.21 addressed the problem of determining whether the distances young (less than 25) males and females drive annually differ. Included in the data is also the number of accidents that each person was involved in the past 2 years. Responses are 0, 1, or 2 or more. Do the data allow us to infer that the distances driven differ between the drivers who have had 0, 1, or 2 or more accidents?

14.98 Xr13-111* The objective in Exercise 13.111 was to determine whether various market segments were more likely to use the Quik Lube service. Included with the data is also the age (in months) of the car. Do the data allow us to conclude that there are differences in the age between the four market segments?

CASE 14.1 Comparing Three Methods of Treating Childhood Ear Infections*


©AP Photo/Chris Carlson

DATA C14-01

Acute otitis media, an infection of the middle ear, is a common childhood illness. There are various ways to treat the problem. To help determine the best way, researchers conducted an experiment. One hundred and eighty children between 10 months and 2 years with recurrent acute otitis media were divided into three equal groups. Group 1 was treated by surgically removing the adenoids (adenoidectomy), the second was treated with the drug Sulfafurazole, and the third with a placebo. Each child was tracked for 2 years, during which time all symptoms and episodes of acute otitis media were recorded. The data were recorded in the following way:

Column 1: ID number
Column 2: Group number
Column 3: Number of episodes of the illness
Column 4: Number of visits to a physician because of any infection
Column 5: Number of prescriptions
Column 6: Number of days with symptoms of respiratory infection

a. Are there differences between the three groups with respect to the number of episodes, number of physician visits, number of prescriptions, and number of days with symptoms of respiratory infection?

b. Assume that you are working for the company that makes the drug Sulfafurazole. Write a report to the company's executives discussing your results.

*This case is adapted from the *British Medical Journal*, February 2004.

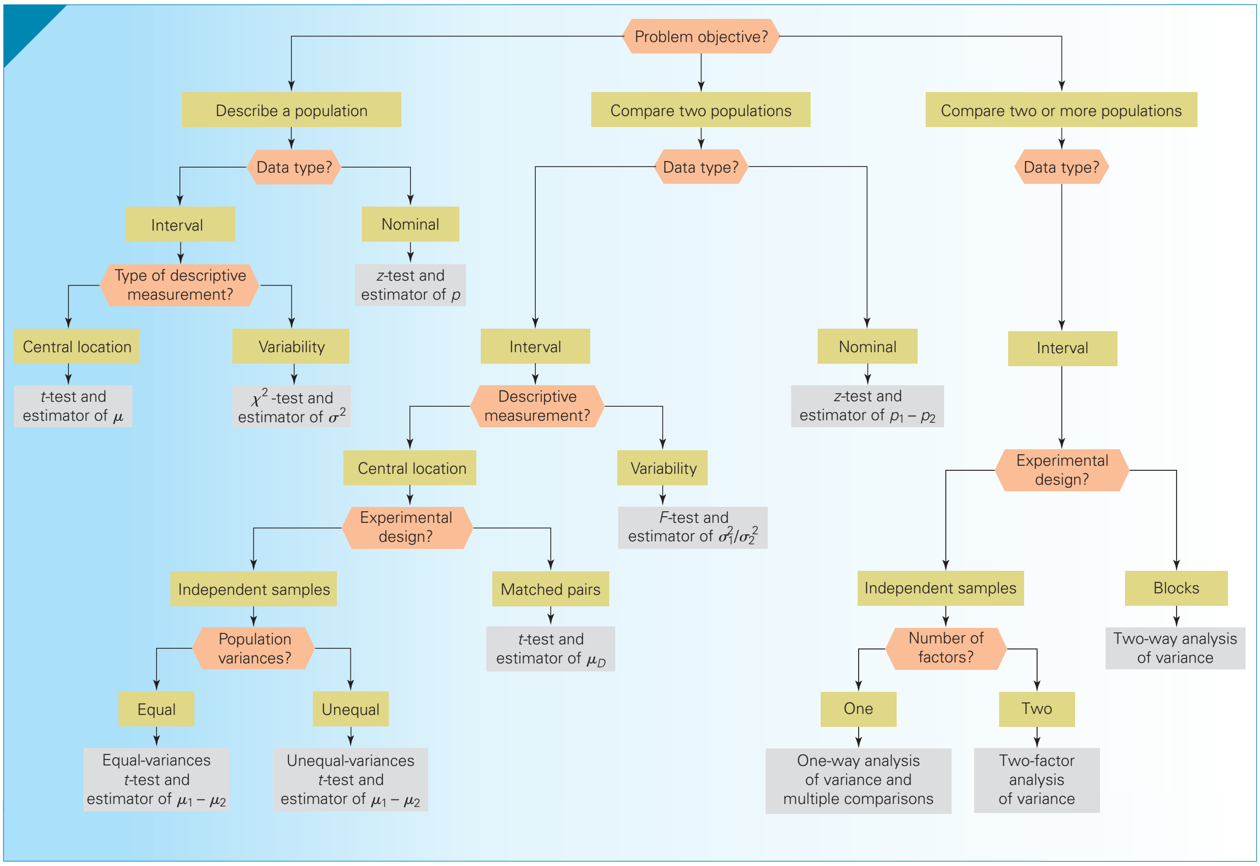
APPENDIX 14 / REVIEW OF CHAPTERS 12 TO 14

The number of techniques introduced in Chapters 12 to 14 is up to 20. As we did in Appendix 13, we provide a table of the techniques with formulas and required conditions, a flowchart to help you identify the correct technique, and 25 exercises to give you practice in how to choose the appropriate method. The table and the flowchart have been amended to include the three analysis of variance techniques introduced in this chapter and the three multiple comparison methods.

TABLE A14.1 Summary of Statistical Techniques in Chapters 12 to 14

t-test of μ
Estimator of μ (including estimator of $N\mu$)
χ^2 test of σ^2
Estimator of σ^2
z-test of p
Estimator of p (including estimator of Np)
Equal-variances t-test of $\mu_1 - \mu_2$
Equal-variances estimator of $\mu_1 - \mu_2$
Unequal-variances t-test of $\mu_1 - \mu_2$
Unequal-variances estimator of $\mu_1 - \mu_2$
t-test of μ_D
Estimator of μ_D
F-test of σ_1^2/σ_2^2
Estimator of σ_1^2/σ_2^2
z-test of $p_1 - p_2$ (Case 1)
z-test of $p_1 - p_2$ (Case 2)
Estimator of $p_1 - p_2$
One-way analysis of variance (including multiple comparisons)
Two-way (randomized blocks) analysis of variance
Two-factor analysis of variance

FIGURE **A14.1** Summary of Statistical Techniques in Chapters 12 to 14



Exercises

Note that as we did in Appendix 13, we do not specify a significance level in exercises requiring a test of hypothesis. We leave this decision to you. After analyzing the issues raised in the exercise, use your own judgment to determine whether the p-value is small enough to reject the null hypothesis.

A14.1 XrA14-01 Sales of a product may depend on its placement in a store. Candy manufacturers frequently offer discounts to retailers who display their products more prominently than competing brands. To examine this phenomenon more carefully, a candy manufacturer (with the assistance of a national chain of restaurants) planned the following experiment. In 20 restaurants, the manufacturer's brand was displayed behind the cashier's counter with all the other brands (this was called position 1). In another 20 restaurants, the brand was placed separately but close to the other brands (position 2). In a third group of 20 restaurants, the candy was placed in a special display next to the cash register (position 3). The number of packages sold during 1 week at each restaurant was recorded. Is there sufficient evidence to infer that sales of candy differ according to placement?

A14.2 XrA14-02 Advertising is critical in the residential real estate industry. Agents are always seeking ways to increase sales through improved advertising methods. A particular agent believes that he can increase the number of inquiries (and thus the probability of making a sale) by describing the house for sale without indicating its asking price. To support his belief, he conducted an experiment in which 100 houses for sale were advertised in two ways—with and without the asking price. The number of inquiries for each house was recorded as well as whether the customer saw the ad with or without the asking price shown. Do these data allow the real estate agent to infer that ads with no price shown are more effective in generating interest in a house?

A14.3 XrA14-03 A professor of statistics hands back his graded midterms in class by calling out the name of each student and personally handing the exam over to its owner. At the end of the process, he notes that there are several exams left over, the result of students missing that class. He forms the theory that the absence is caused by a poor performance by those students on the test. If the theory is correct, the leftover papers will have lower marks than those papers handed back. He recorded the marks (out of 100) for the leftover papers and the marks of the returned papers. Do the data support the professor's theory?

A14.4 XrA14-04 A study was undertaken to determine whether a drug commonly used to treat epilepsy could help alcoholics to overcome their addiction. The researchers took a sample of 103 hardcore alcoholics. Fifty-five drinkers were given topiramate and the remaining 48 were given a placebo. The following variables were recorded after 6 months:

Column 1: Identification number
Column 2: 1 = Topiramate and 2 = placebo
Column 3: Abstain from alcohol for one month (1 = no, 2 = yes)
Column 4: Did not binge in final month (1 = no, 2 = yes)

Do these data provide sufficient evidence to infer that topiramate is effective in

a. causing abstinence for the first month?
b. causing alcoholics to refrain from binge drinking in the final month?

A14.5 XrA14-05 Health-care costs in the United States and Canada are concerns for citizens and politicians. The question is, How can we devise a system wherein people's medical bills are covered but individuals attempt to reduce costs? An American company has come up with a possible solution. Golden Rule is an insurance company in Indiana with 1,300 employees. The company offered its employees a choice of programs. One choice was a medical savings account (MSA) plan. Here's how it works. To ensure that a major illness or accident does not financially destroy an employee, Golden Rule offers catastrophic insurance—a policy that covers all expenses above $2,000 per year. At the beginning of the year, the company deposits $1,000 (for a single employee) and $2,000 (for an employee with a family) into the MSA. For minor expenses, the employee pays from his or her MSA. As an incentive for the employee to spend wisely, any money left in the MSA at the end of the year can be withdrawn by the employee. To determine how well it works, a random sample of employees who opted for the medical savings account plan was compared to employees who chose the regular plan. At the end of the year, the medical expenses for each employee were recorded. Critics of MSA say that the plan leads to poorer health care, and as a result employees are less likely to be in excellent health. To address this issue, each employee was examined. The results of the examination were recorded where 1 = excellent health and 2 = not in excellent health

a. Can we infer from these data that MSA is effective in reducing costs?
b. Can we infer that the critics of MSA are correct?

A14.6 XrA14-06 Discrimination in hiring has been illegal for many years. It is illegal to discriminate against any person on the basis of race, gender, or religion. It is also illegal to discriminate because of a person's handicap if it in no way prevents that person from performing that job. In recent years, the definition of "handicap" has widened. Several applicants have successfully sued companies because they were denied employment for no other reason than that they were overweight. A study was conducted to examine attitudes toward overweight people. The experiment involved showing a number of subjects videotape of an applicant being interviewed for a job. Before the interview, the subject was given a description of the job. Following the interview, the subject was asked to score the applicant in terms of how well the applicant was suited for the job. The score was out of 100, where higher scores described greater suitability. (The scores are interval data.) The same procedure was repeated for each subject. However, the gender and weight (average and overweight) of the applicant varied. The results were recorded using the following format:

Column 1: Score for average weight males
Column 2: Score for overweight males
Column 3: Score for average weight females
Column 4: Score for overweight females

a. Can we infer that the scores of the four groups of applicants differ?
b. Are the differences detected in part (a) because of weight, gender, or some interaction?

A14.7 XrA14-07 Most automobile repair shops now charge according to a schedule that is claimed to be based on average times. This means that instead of determining the actual time to make a repair and multiplying this value by their hourly rate, repair shops determine the cost from a schedule that is calculated from average times. A critic of this policy is examining how closely this schedule adheres to the actual time to complete a job. He randomly selects five jobs. According to the schedule, these jobs should take 45 minutes, 60 minutes, 80 minutes, 100 minutes, and 125 minutes, respectively. The critic then takes a random sample of repair shops and records the actual times for each of 20 cars for each job. For each job, can we infer that the time specified by the schedule is greater than the actual time?

A14.8 XrA14-08 Automobile insurance appraisers examine cars that have been involved in accidental collisions and estimate the cost of repairs. An insurance executive claims that there are significant differences in the estimates from different appraisers. To support his claim, he takes a random sample of 25 cars that have recently been damaged in accidents. Three appraisers then estimated the repair costs of each car. The estimates were recorded for each appraiser. From the data, can we conclude that the executive's claim is true?

A14.9 XrA14-09 The widespread use of salt on roads in Canada and the northern United States during the winter and acid precipitation throughout the year combine to cause rust on cars. Car manufacturers and other companies offer rustproofing services to help purchasers preserve the value of their cars. A consumer protection agency decides to determine whether there are any differences between the rust protection provided by automobile manufacturers and that provided by two competing types of rustproofing services. As an experiment, 60 identical new cars are selected. Of these, 20 are rustproofed by the manufacturer. Another 20 are rustproofed using a method that applies a liquid to critical areas of the car. The liquid hardens, forming a (supposedly) lifetime bond with the metal. The last 20 are treated with oil and are retreated every 12 months. The cars are then driven under similar conditions in a Minnesota city. The number of months until the first rust appears was recorded. Is there sufficient evidence to conclude that at least one rustproofing method is different from the others?

A14.10 XrA14-10 One of the ways in which advertisers measure the value of television commercials is by telephone surveys conducted shortly after commercials are aired. Respondents who watched a certain television station at a given time period, during which the commercial appeared, are asked whether they can recall the name of the product in the commercial. Suppose an advertiser wants to compare the recall proportions of two commercials. The first commercial is relatively inexpensive. A second commercial shown a week later is quite expensive to produce. The advertiser decides that the second commercial is viable only if its recall proportion is more than 15% higher than the recall proportion of the first commercial. Two surveys of 500 television viewers each were conducted after each commercial was aired. Each person was asked whether he or she remembered the product name. The results are stored in columns 1 (commercial 1) and 2 (commercial 2) (2 = remembered the product name, 1 = did not remember the product name). Can we infer that the second commercial is viable?

A14.11 XrA14-11 In the door-to-door selling of vacuum cleaners, various factors influence sales. The Birk Vacuum Cleaner Company considers its sales pitch and overall package to be extremely important. As a result, it often thinks of new ways to sell its product. Because the company's management develops so many new sales pitches each year, there is a two-stage testing process. In stage 1, a new plan is tested with a relatively small sample. If there is sufficient evidence that the plan increases sales, a second, considerably larger, test is undertaken. In a stage 1 test to determine whether the inclusion of a "free" 10-year service contract increases sales, 100 sales representatives were selected at random from the company's list of several thousand. The monthly sales of these representatives were recorded for 1 month before the use of the new sales pitch and for 1 month after its introduction. Should the company proceed to stage 2?

A14.12 XrA14-12 The cost of workplace injuries is high for the individual worker, for the company, and for society. It is in everyone's interest to rehabilitate the injured worker as quickly as possible. A statistician working for an insurance company has investigated the problem. He believes that physical condition is a major determinant in how quickly a worker returns to his or her job after sustaining an injury. To help determine whether he is on the right track, he organized an experiment. He took a random sample of male and female workers who were injured during the preceding year. He recorded their gender, their physical condition, and the number of working days until they returned to their job. These data were recorded in the following way. Columns 1 and 2 store the number of working days until return to work for men and women, respectively. In each column, the first 25 observations relate to those who are physically fit, the next 25 rows relate to individuals who are moderately fit, and the last 25 observations are for those who are in poor physical shape. Can we infer that the six groups differ? If differences exist, determine whether the differences result from gender, physical fitness, or some combination of gender and physical fitness.

A14.13 XrA14-13 Does driving an ABS-equipped car change the behavior of drivers? To help answer this question, the following experiment was undertaken. A random sample of 200 drivers who currently operate cars without ABS was selected. Each person was given an identical car to drive for 1 year. Half the sample were given cars that had ABS, and the other half were given cars with standard-equipment brakes. Computers on the cars recorded the average speed (in miles per hour) during the year. Can we infer that operating an ABS-equipped car changes the behavior of the driver?

A14.14 XrA14-14 We expect the demand for a product depends on its price: The higher the price, the lower the demand. However, this may not be entirely true. In an experiment conducted by professors at Northwestern University and MIT, a mail-order dress was available at the prices $34, $39, and $44. The number of dresses sold weekly over a 20-week period was recorded. The prices were randomized over 60 weeks. Conduct a test to determine whether demand differed and, if so, which price elicited the highest sales.

A14.15 XrA14-15 Researchers at the University of Washington conducted an experiment to determine whether the herbal remedy Echinacea is effective in treating children's colds and other respiratory infection (*National Post*, December 3, 2003). A sample of 524 children was recruited. Half the sample treated their colds with Echinacea, and the other half was given a placebo. For each infection, the duration of the colds (in days) were measured and recorded. Can we conclude that Echinacea is effective?

A14.16 XrA14-16 The marketing manager of a large ski resort wants to advertise that his ski resort has the shortest lift lines of any resort in the area. To avoid the possibility of a false advertising liability suit, he collects data on the times skiers wait in line at his resort and at each of two competing resorts on each of 14 days.

a. Can he conclude that there are differences in waiting times between the three resorts?
b. What are the required conditions for these techniques?
c. How would you check to determine that the required conditions are satisfied?

A14.17 XrA14-17 A popularly held belief about university professors is that they don't work very hard and that the higher their rank, the less work they do. A statistics student decided to determine whether the belief is true. She took a random sample of 20 university instructors in the faculties of business, engineering, arts, and sciences. In each sample of 20, 5 were instructors, 5 were assistant professors, 5 were associate professors, and 5 were full professors. Each professor was surveyed and asked to report confidentially the number of weekly hours of work. These data were recorded in the following way:

Column 1: hours of work for business professors (first 5 rows = instructors, next 5 rows = assistant professors, next 5 rows = associate professors, and last 5 rows = full professors)
Column 2: hours of work for engineering professors (same format as column 1)
Column 3: hours of work for arts professors (same format as column 1)
Column 4: hours of work for science professors (same format as column 1)

a. If we conduct the test under the single-factor analysis of variance, how many levels are there? What are they?
b. Test to determine whether differences exist using a single-factor analysis of variance.
c. If we conduct tests using the two-factor analysis of variance, what are the factors? What are their levels?
d. Is there evidence of interaction?
e. Are there differences between the four ranks of instructor?
f. Are there differences between the four faculties?

A14.18 XrA14-18 Billions of dollars are spent annually by Americans for the care and feeding of pets. A survey conducted by the American Veterinary Medical Association drew a random sample of 1,328 American households and asked whether they owned a pet and, if so, the type of animal. In addition, each was asked to report the veterinary expenditures for the previous 12 months. Column 1 contains the expenditures for dogs, and column 2 stores the expenditures for cats. The results are that 474 households reported that they owned at least one dog and 419 owned at least one cat. The latest census indicates that there are 112 million households in the United States. (*Source: Statistical Abstract of the United States, 2006*, Table 1232.)

a. Estimate with 95% confidence the total number of households owning at least one dog.
b. Repeat part (a) for cats.
c. Assume that there are 40 million households with at least one dog and estimate with 95% confidence the total amount spent on veterinary expenditures for dogs.
d. Assume that there are 35 million households with at least one cat and estimate with 95% confidence the total amount spent on veterinary expenditures for cats.

General Social Survey Exercises

A14.19 GSS2008* Can we infer from the data that the majority of Americans support capital punishment for murderers? (CAPPUN: 1 = Favor, 2 = Oppose)

A14.20 GSS2008* Test to determine whether Democrats and Republicans differ in their answers to the question, Have you ever taken any drugs by injection (heroine, cocaine, etc.)? (EVIDU: 1 = Yes, 2 = No)

A14.21 GSS2008* Is there enough evidence to infer that differences in the amount of television watched (TVHOURS) differs between classes (CLASS)?

A14.22 GSS2008* Do the data provide enough statistical evidence to conclude that differences in number of hours worked (HRS) exist between the three races (RACE)?

A14.23 GSS2006* GSS2008* Is there sufficient evidence to infer that on average Americans have aged (AGE) between 2006 and 2008?

American National Election Survey Exercises

A14.24 ANES2008* Can we conclude that differences in having access to the Internet (ACCESS: 1 = Yes, 5 = No) differs between Republicans and Democrats (PARTY: 1 Democrat, 2 = Republican)?

A14.25 ANES2008* Can we conclude that differences in age (AGE) exist between liberals, moderates, and conservatives (LIBCON3)?

This page intentionally left blank

© Bob Elsdale/Workbook Stock/Jupiterimages

15 CHI-SQUARED TESTS

15.1 *Chi-Squared Goodness-of-Fit Test*

15.2 *Chi-Squared Test of a Contingency Table*

15.3 *Summary of Tests on Nominal Data*

15.4 *(Optional) Chi-Squared Test for Normality*

Appendix 15 *Review of Chapters 12 to 15*

General Social Surveys

Has Support for Capital Punishment for Murderers Changed since 2002?

DATA
GSS2002*
GSS2004*
GSS2006*
GSS2008*

The issue of capital punishment for murderers in the United States has been argued for many years. A few states have abolished it, and others have kept their laws on the books but rarely use them. Where does the public stand on the issue, and has public support been constant or has it changed from year to year? One of the questions asked in the General Social Survey was

Do you favor capital punishment for murder (CAPPUN)? The responses are

1 = Favor, 2 = Oppose

Conduct a test to determine whether public support varies from year to year.

© Bob Daemmerich/PhotoEdit

On page 611 we solve this problem.

INTRODUCTION

We have seen a variety of statistical techniques that are used when the data are nominal. In Chapter 2, we introduced bar and pie charts, both graphical techniques to describe a set of nominal data. Later in Chapter 2, we showed how to describe the relationship between two sets of nominal data by producing a frequency table and a bar chart. However, these techniques simply describe the data, which may represent a sample or a population. In this chapter, we deal with similar problems, but the goal is to use statistical techniques to make inferences about populations from sample data.

This chapter develops two statistical techniques that involve nominal data. The first is a *goodness-of-fit test* applied to data produced by a *multinomial experiment*, a generalization of a binomial experiment. The second uses data arranged in a table (called a *contingency table*) to determine whether two classifications of a population of nominal data are statistically independent; this test can also be interpreted as a comparison of two or more populations. The sampling distribution of the test statistics in both tests is the chi-squared distribution introduced in Chapter 8.

15.1 CHI-SQUARED GOODNESS-OF-FIT TEST

This section presents another test designed to describe a population of nominal data. The first such test was introduced in Section 12.3, where we discussed the statistical procedure employed to test hypotheses about a population proportion. In that case, the nominal variable could assume one of only two possible values: success or failure. Our tests dealt with hypotheses about the proportion of successes in the entire population. Recall that the experiment that produces the data is called a *binomial experiment*. In this section, we introduce the **multinomial experiment**, which is an extension of the binomial experiment, wherein there are two or more possible outcomes per trial.

Multinomial Experiment

A multinomial experiment is one that possesses the following properties.

1. The experiment consists of a fixed number n of trials.
2. The outcome of each trial can be classified into one of k categories, called *cells*.
3. The probability p_i that the outcome will fall into cell i remains constant for each trial. Moreover, $p_1 + p_2 + \cdots + p_k = 1$
4. Each trial of the experiment is independent of the other trials.

When $k = 2$, the multinomial experiment is identical to the binomial experiment. Just as we count the number of successes (recall that we label the number of successes x) and failures in a binomial experiment, we count the number of outcomes falling into each of the k cells in a multinomial experiment. In this way, we obtain a set of observed frequencies $f_1, f_2, \ldots, f_k$ where f_i is the observed frequency of outcomes falling into cell i, for $i = 1, 2, \ldots, k$. Because the experiment consists of n trials and an outcome must fall into some cell,

$$f_1 + f_2 + \cdots + f_k = n$$

Just as we used the number of successes x (by calculating the sample proportion $\hat{p}$, which is equal to x/n) to draw inferences about p, so we use the observed frequencies to draw inferences about the cell probabilities. We'll proceed in what by now has become a standard procedure. We will set up the hypotheses and develop the test statistic and its sampling distribution. We'll demonstrate the process with the following example.

EXAMPLE 15.1 Testing Market Shares

Company A has recently conducted aggressive advertising campaigns to maintain and possibly increase its share of the market (currently 45%) for fabric softener. Its main competitor, company B, has 40% of the market, and a number of other competitors account for the remaining 15%. To determine whether the market shares changed after the advertising campaign, the marketing manager for company A solicited the preferences of a random sample of 200 customers of fabric softener. Of the 200 customers, 102 indicated a preference for company A's product, 82 preferred company B's fabric softener, and the remaining 16 preferred the products of one of the competitors. Can the analyst infer at the 5% significance level that customer preferences have changed from their levels before the advertising campaigns were launched?

SOLUTION

The population in question is composed of the brand preferences of the fabric softener customers. The data are nominal because each respondent will choose one of three possible answers: product A, product B, or other. If there were only two categories, or if we were interested only in the proportion of one company's customers (which we would label as successes and label the others as failures), we would identify the technique as the z-test of p. However, in this problem we're interested in the proportions of all three categories. We recognize this experiment as a multinomial experiment, and we identify the technique as the **chi-squared goodness-of-fit test**.

Because we want to know whether the market shares have changed, we specify those precampaign market shares in the null hypothesis.

$$H_0\text{:}\quad p_1 = .45, p_2 = .40, p_3 = .15$$

The alternative hypothesis attempts to answer our question, Have the proportions changed? Thus,

$$H_1\text{:}\quad \text{At least one } p_i \text{ is not equal to its specified value}$$

Test Statistic

If the null hypothesis is true, we would expect the number of customers selecting brand A, brand B, and other to be 200 times the proportions specified under the null hypothesis; that is,

$$e_1 = 200(.45) = 90$$

$$e_2 = 200(.40) = 80$$

$$e_3 = 200(.15) = 30$$

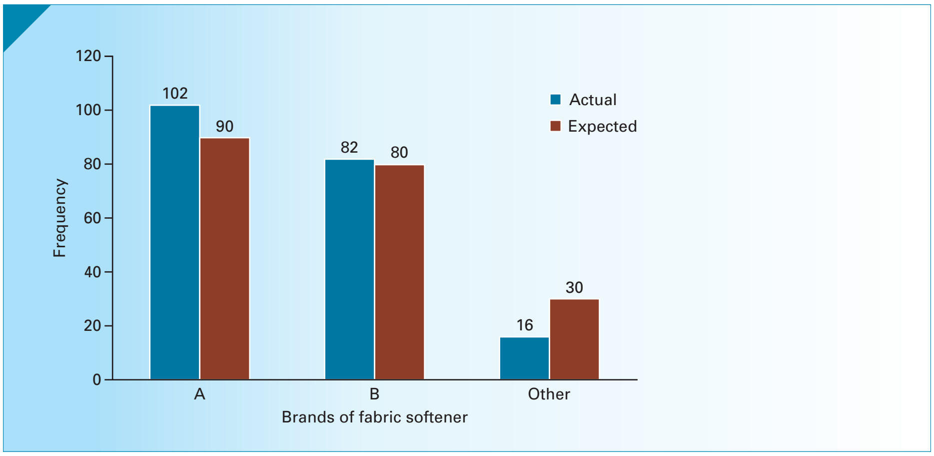
In general, the **expected frequency** for each cell is given by

$$e_i = np_i$$

This expression is derived from the formula for the expected value of a binomial random variable, introduced in Section 7.4.

Figure 15.1 is a bar chart (created by Excel) showing the comparison of actual and expected frequencies.

FIGURE **15.1** Bar Chart for Example 15.1



If the expected frequencies e_i and the **observed frequencies** f_i are quite different, we would conclude that the null hypothesis is false, and we would reject it. However, if the expected and observed frequencies are similar, we would not reject the null hypothesis. The test statistic defined in the box measures the similarity of the expected and observed frequencies.

Chi-Squared Goodness-of-Fit Test Statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

The sampling distribution of the test statistic is approximately chi-squared distributed with $\nu = k - 1$ degrees of freedom, provided that the sample size is large. We will discuss this required condition later. (The chi-squared distribution was introduced in Section 8.4.)

The following table demonstrates the calculation of the test statistic. Thus, the value $\chi^2 = 8.18$. As usual, we judge the size of this test statistic by specifying the rejection region or by determining the p-value.

Company	Observed Frequency f_i	Expected Frequency e_i	$(f_i - e_i)$	$\frac{(f_i - e_i)^2}{e_i}$
A	102	90	12	1.60
B	82	80	2	0.05
Other	16	30	−14	6.53
Total	200	200		$\chi^2 = 8.18$

When the null hypothesis is true, the observed and expected frequencies should be similar, in which case the test statistic will be small. Thus, a small test statistic supports the null hypothesis. If the null hypothesis is untrue, some of the observed and expected frequencies will differ and the test statistic will be large. Consequently, we want to reject the null hypothesis when χ^2 is greater than $\chi^2_{\alpha,k-1}$. In other words, the rejection region is

$$\chi^2 > \chi^2_{\alpha,k-1}$$

In Example 15.1, $k = 3$; the rejection region is

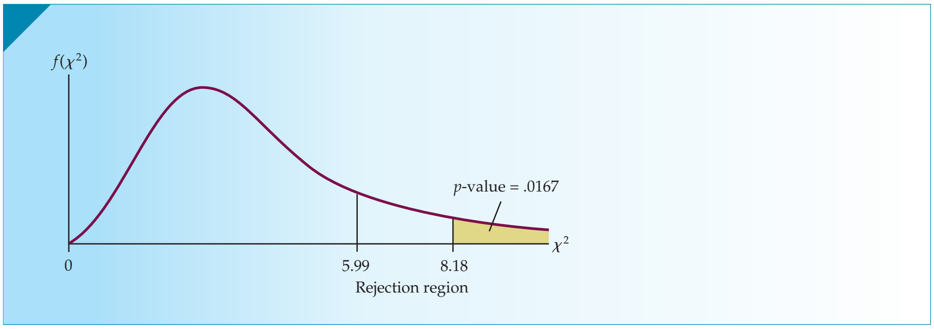
$$\chi^2 > \chi^2_{\alpha,k-1} = \chi^2_{.05,2} = 5.99$$

Because the test statistic is $\chi^2 = 8.18$, we reject the null hypothesis. The p-value of the test is

$$p\text{-value} = P(\chi^2 > 8.18)$$

Unfortunately, Table 5 in Appendix B does not allow us to perform this calculation (except for approximation by interpolation). The p-value must be produced by computer. Figure 15.2 depicts the sampling distribution, rejection region, and p-value.

FIGURE 15.2 Sampling Distribution for Example 15.1



EXCEL

The output from the commands listed here is the p-value of the test. It is .0167.

INSTRUCTIONS

1. Type the observed values into one column and the expected values into another column. (If you wish, you can type the cell probabilities specified in the null hypothesis and let Excel convert these into expected values by multiplying by the sample size.)
2. Activate an empty cell and type

= **CHITEST**(**[Actual_range], [Expected_range]**)

where the ranges are the cells containing the actual observations and the expected values.

You can also perform what-if analyses to determine for yourself the effect of changing some of the observed values and the sample size.

If we have the raw data representing the nominal responses we must first determine the frequency of each category (the observed values) using the **COUNTIF** function described on page 20.

MINITAB

Chi-Square Goodness-of-Fit Test for Observed Counts in Variable: C1

Category	Observed	Test Proportion	Expected	Contribution to Chi-Sq
1	102	0.45	90	1.60000
2	82	0.40	80	0.05000
3	16	0.15	30	6.53333

N	DF	Chi-Sq	P-Value
200	2	8.18333	0.017

INSTRUCTIONS

1. Click **Stat, Tables**, and **Chi-square Goodness-of-Fit Test (One Variable) . . . .**
2. Type the observed values into the **Observed counts:** box (102 82 16). If you have a column of data click **Categorical data:** and specify the column or variable name.
3. Click **Proportions specified by historical counts** and **Input constants**. Type the values of the proportions under the null hypothesis (.45 .40 .15).

INTERPRET

There is sufficient evidence at the 5% significance level to infer that the proportions have changed since the advertising campaigns were implemented. If the sampling was conducted properly, we can be quite confident in our conclusion. This technique has only one required condition, which is satisfied. (See the next subsection.) It is probably a worthwhile exercise to determine the nature and causes of the changes. The results of this analysis will determine the design and timing of other advertising campaigns.

Required Condition

The actual sampling distribution of the test statistic defined previously is discrete, but it can be approximated by the chi-squared distribution provided that the sample size is large. This requirement is similar to the one we imposed when we used the normal approximation to the binomial in the sampling distribution of a proportion. In that approximation we needed np and $n(1 - p)$ to be 5 or more. A similar rule is imposed for the chi-squared test statistic. It is called the *rule of five*, which states that the sample size must be large enough so that the expected value for each cell must be 5 or more. Where necessary, cells should be combined to satisfy this condition. We discuss this required condition and provide more details on its application in Keller's website Appendix Rule of Five.

Factors That Identify the Chi-Squared Goodness-of-Fit Test

1. **Problem objective**: Describe a single population
2. **Data type**: Nominal
3. **Number of categories**: 2 or more

EXERCISES

Developing an Understanding of Statistical Concepts

Exercises 15.1–15.6 are "what-if" analyses designed to determine what happens to the test statistic of the goodness-of-fit test when elements of the statistical inference change. These problems can be solved manually or using Excel's **CHITEST**.

15.1 Consider a multinomial experiment involving $n = 300$ trials and $k = 5$ cells. The observed frequencies resulting from the experiment are shown in the accompanying table, and the null hypothesis to be tested is as follows:

$$H_0:\ p_1 = .1, p_2 = .2, p_3 = .3, p_4 = .2, p_5 = .2$$

Test the hypothesis at the 1% significance level.

Cell	1	2	3	4	5
Frequency	24	64	84	72	56

15.2 Repeat Exercise 15.1 with the following frequencies:

Cell	1	2	3	4	5
Frequency	12	32	42	36	28

15.3 Repeat Exercise 15.1 with the following frequencies:

Cell	1	2	3	4	5
Frequency	6	16	21	18	14

15.4 Review the results of Exercises 15.1–15.3. What is the effect of decreasing the sample size?

15.5 Consider a multinomial experiment involving $n = 150$ trials and $k = 4$ cells. The observed frequencies resulting from the experiment are shown in the accompanying table, and the null hypothesis to be tested is as follows:

$$H_0:\ p_1 = .3, p_2 = .3, p_3 = .2, p_4 = .2$$

Cell	1	2	3	4
Frequency	38	50	38	24

Test the hypotheses, using $\alpha = .05$.

15.6 For Exercise 15.5, retest the hypotheses, assuming that the experiment involved twice as many trials ($n = 300$) and that the observed frequencies were twice as high as before, as shown here.

Cell	1	2	3	4
Frequency	76	100	76	48

Exercises 15.7–15.21 require the use of a computer and software. Use a 5% significance level unless specified otherwise. The answers to Exercises 15.7–15.16 may be calculated manually. See Appendix A for the sample statistics.

15.7 Xr15-07 The results of a multinomial experiment with $k = 5$ were recorded. Each outcome is identified by the numbers 1 to 5. Test to determine whether there is enough evidence to infer that the proportions of outcomes differ.

15.8 Xr15-08 A multinomial experiment was conducted with $k = 4$. Each outcome is stored as an integer from 1 to 4 and the results of a survey were recorded. Test the following hypotheses.

$$H_0:\ p_1 = .15, p_2 = .40, p_3 = .35, p_4 = .10$$

H_1: At least one p_i is not equal to its specified value

15.9 Xr15-09 To determine whether a single die is balanced, or fair, the die was rolled 600 times. Is there sufficient evidence to allow you to conclude that the die is not fair?

Applications

15.10 Xr15-10 Grades assigned by an economics instructor have historically followed a symmetrical distribution: 5% A's, 25% B's, 40% C's, 25% D's, and 5% F's. This year, a sample of 150 grades was drawn and the grades (1 = A, 2 = B, 3 = C, 4 = D, and 5 = F) were recorded. Can you conclude, at the 10% level of significance, that this year's grades are distributed differently from grades in the past?

15.11 Xr15-11 Pat Statsdud is about to write a multiple-choice exam but as usual knows absolutely nothing. Pat plans to guess one of the five choices. Pat has been given one of the professor's previous exams with the correct answers marked. The correct choices were recorded where 1 = (a), 2 = (b), 3 = (c), 4 = (d), and 5 = (e). Help Pat determine whether this professor does not randomly distribute the correct answer over the five choices? If this is true, how does it affect Pat's strategy?

15.12 Xr15-12 Financial managers are interested in the speed with which customers who make purchases on credit pay their bills. In addition to calculating the average number of days that unpaid bills (called *accounts receivable*) remain outstanding, they often prepare an aging schedule. An aging schedule classifies outstanding accounts receivable according to the time that has elapsed since billing and records the proportion of accounts receivable belonging to each classification. A large firm has determined its aging schedule for the past 5 years. These results are shown in the accompanying table. During the past few months, however, the economy has taken a downturn. The company would like to know whether the recession has affected the aging schedule. A random sample of 250 accounts receivable was drawn and each account was classified as follows:

1 = 0–14 days outstanding

2 = 15–29 days outstanding

3 = 30–59 days outstanding

4 = 60 or more days outstanding

Number of Days Outstanding	Proportion of Accounts Receivable Past 5 Years
0–14	.72
15–29	.15
30–59	.10
60 and more	.03

Determine whether the aging schedule has changed.

15.13 Xr15-13 License records in a county reveal that 15% of cars are subcompacts (1), 25% are compacts (2), 40% are midsize (3), and the rest are an assortment of other styles and models (4). A random sample of accidents involving cars licensed in the county was drawn. The type of car was recorded using the codes in parentheses. Can we infer that certain sizes of cars are involved in a higher than expected percentage of accidents?

15.14 Xr15-14 In an election held last year that was contested by three parties. Party A captured 31% of the vote, party B garnered 51%, and party C received the remaining votes. A survey of 1,200 voters asked each to identify the party that they would vote for in the next election. These results were recorded where 1 = party A, 2 = party B, and 3 = party C. Can we infer at the 10% significance level that voter support has changed since the election?

15.15 Xr15-15 In a number of pharmaceutical studies volunteers who take placebos (but are told they have taken a cold remedy) report the following side effects:

Headache (1)	5%
Drowsiness (2)	7%
Stomach upset (3)	4%
No side effect (4)	84%

A random sample of 250 people who were given a placebo (but who thought they had taken an anti-inflammatory) reported whether they had experienced each of the side effects. These responses were recorded using the codes in parentheses. Do these data provide enough evidence to infer that the reported side effects of the placebo for an anti-inflammatory differ from that of a cold remedy?

APPLICATIONS in MARKETING

© Paul Thomas/ The Image Bank/Getty Images

Market Segmentation

Market segmentation was introduced in Section 12.4, where a statistical technique was used to estimate the size of a segment. In Chapters 13 and 14, statistical procedures were applied to determine whether market segments differ in their purchases of products and services. Exercise 15.16 requires you to apply the chi-squared goodness-of-fit test to determine whether the relative sizes of segments have changed.

15.16 Xr12-125* Refer to Exercise 12.125 where the statistics practitioner estimated the size of market segments based on education among California adults. Suppose that census figures from 10 years ago showed the education levels and the proportions of California adults, as follows:

Level	Proportion
1. Did not complete high school	.23
2. Completed high school only	.40
3. Some college or university	.15
4. College or university graduate	.22

Determine whether there has been a change in these proportions.

American National Election Survey Exercise

15.17 ANES2008* According to the *Statistical Abstract of the United States, 2009*, Table 55, the proportions for each category of marital status in 2007 was

Never married (including partnered, not married) 25%
Married (including separated, but not divorced) 58%
Widowed 6%
Divorced 11%

Can we infer that the American National Election Survey in 2008 overrepresented at least one category of marital status (MARITAL)?

General Social Survey Exercises

According to the *Statistical Abstract of the United States, 2009*, Table 7, the racial mix in the United States in 2007 was

White 79%
Black 13%
Other 8%

15.18 GSS2008* Test to determine whether there is sufficient evidence that the General Social Survey in 2008 overrepresented at least one race (RACE).

15.19 GSS2006* Is there sufficient evidence to conclude that the General Social Survey in 2006 overrepresented at least one race (RACE)?

According to the *Statistical Abstract of the United States, 2009*, Table 55, the proportions for each category of marital status in 2007 was

Never married 25%
Married (including separated, but not divorced) 58%
Widowed 6%
Divorced 11%

15.20 GSS2008* Can we infer that the General Social Survey in 2008 overrepresented at least one category of marital status (MARITAL)?

15.21 GSS2006* Is there sufficient evidence to conclude that the General Social Survey in 2006 overrepresented at least one category of marital status (MARITAL)?

15.2 Chi-Squared Test of a Contingency Table

In Chapter 2, we developed the **cross-classification table** as a first step in graphing the relationship between two nominal variables (see page 32). Our goal was to determine whether the two variables were related. In this section we extend the technique to statistical inference. We introduce another chi-squared test, this one designed to satisfy two different problem objectives. The **chi-squared test of a contingency table** is used to determine whether there is enough evidence to infer that two nominal variables are related and to infer that differences exist between two or more populations of nominal variables. Completing both objectives entails classifying items according to two different criteria. To see how this is done, consider the following example.

EXAMPLE 15.2 Relationship between Undergraduate Degree and MBA Major

DATA
Xm15-02

The MBA program was experiencing problems scheduling its courses. The demand for the program's optional courses and majors was quite variable from one year to the next. In one year, students seem to want marketing courses; in other years, accounting or finance are the rage. In desperation, the dean of the business school turned to a statistics professor for assistance. The statistics professor believed that the problem may be the variability in the academic background of the students and that the undergraduate degree affects the choice of major. As a start, he took a random sample of last year's MBA students and recorded the undergraduate degree and the major selected in the graduate program. The undergraduate degrees were BA, BEng, BBA, and several others. There are three possible majors for the MBA students: accounting, finance, and marketing. The results were summarized in a cross-classification table, which is shown here. Can the statistician conclude that the undergraduate degree affects the choice of major?

	MBA Major			
Undergraduate Degree	**Accounting**	**Finance**	**Marketing**	**Total**
B.A.	31	13	16	60
B.Eng.	8	16	7	31
B.B.A.	12	10	17	39
Other	10	5	7	22
Total	61	44	47	152

SOLUTION

One way to solve the problem is to consider that there are two variables: undergraduate degree and MBA major. Both are nominal. The values of the undergraduate degree are BA, BEng, BBA, and other. The values of MBA major are accounting, finance, and marketing. The problem objective is to analyze the relationship between the two variables. Specifically, we want to know whether one variable is related to the other.

Another way of addressing the problem is to determine whether differences exist between BA's, BEng's, BBA's, and others. In other words, we treat the holders of each undergraduate degree as a separate population. Each population has three possible values represented by the MBA major. The problem objective is to compare four populations. (We can also answer the question by treating the MBA majors as populations and the undergraduate degrees as the values of the random variable.)

As you will shortly discover, both objectives lead to the same test. Consequently, we address both objectives at the same time.

The null hypothesis will specify that there is no relationship between the two variables. We state this in the following way:

H_0: The two variables are independent

The alternative hypothesis specifies one variable affects the other, expressed as

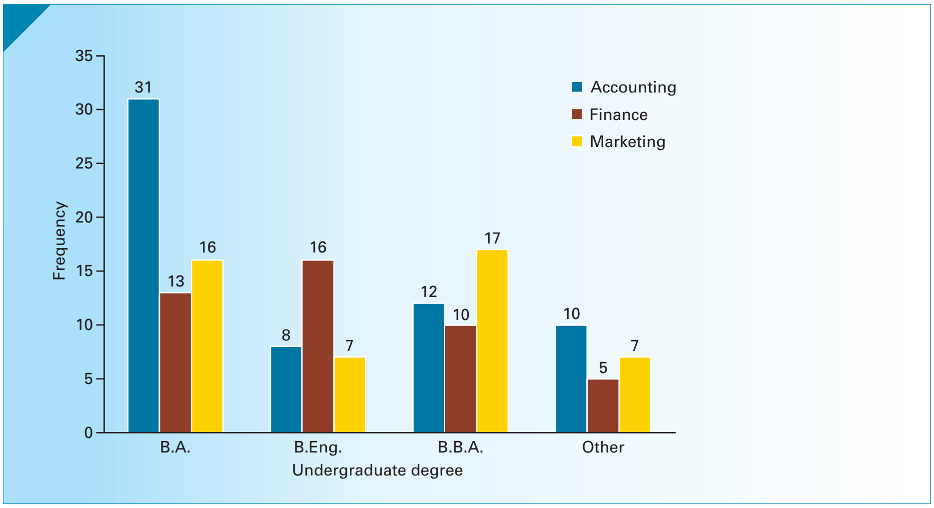
H_1: The two variables are dependent

Graphical Technique

Figure 15.3 depicts the graphical technique introduced in Chapter 2 to show the relationship (if any) between the two nominal variables.

The bar chart displays the data from the sample. It does appear that there is a relationship between the two nominal variables in the sample. However, to draw inferences about the population of MBA students we need to apply an inferential technique.

FIGURE 15.3 Bar Chart for Example 15.2



Test Statistic

The test statistic is the same as the one used to test proportions in the goodness-of-fit test; that is, the test statistic is

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

where k is the number of cells in the cross-classification table. If you examine the null hypothesis described in the goodness-of-fit test and the one described above, you will discover a major difference. In the goodness-of-fit test, the null hypothesis lists values for the probabilities p_i. The null hypothesis for the chi-squared test of a contingency table only states that the two variables are independent. However, we need the probabilities to compute the expected values e_i, which in turn are needed to calculate the value of the test statistic. (The entries in the table are the observed values f_i.) The question immediately arises, From where do we get the probabilities? The answer is that they must come from the data after we assume that the null hypothesis is true.

In Chapter 6 we introduced independent events and showed that if two events A and B are independent, the joint probability P(A and B) is equal to the product of P(A) and P(B). That is,

$$P(\text{A and B}) = P(\text{A}) \times P(\text{B})$$

The events in this example are the values each of the two nominal variables can assume. Unfortunately, we do not have the probabilities of A and B. However, these probabilities can be estimated from the data. Using relative frequencies, we calculate the estimated probabilities for the MBA major.

$$P(\text{Accounting}) = \frac{61}{152} = .401$$

$$P(\text{Finance}) = \frac{44}{152} = .289$$

$$P(\text{Marketing}) = \frac{47}{152} = .309$$

We calculate the estimated probabilities for the undergraduate degree.

$$P(\text{BA}) = \frac{60}{152} = .395$$

$$P(\text{BEng}) = \frac{31}{152} = .204$$

$$P(\text{BBA}) = \frac{39}{152} = .257$$

$$P(\text{Other}) = \frac{22}{152} = .145$$

Assuming that the null hypothesis is true, we can compute the estimated joint probabilities. To produce the expected values, we multiply the estimated joint probabilities by the sample size, $n = 152$. The results are listed in a **contingency table**, the word *contingency* derived by calculating the expected values contingent on the assumption that the null hypothesis is true (the two variables are independent).

Undergraduate	MBA Major			
Degree	Accounting	Finance	Marketing	Total
B.A.	$152 \times \frac{60}{152} \times \frac{61}{152} = 24.08$	$152 \times \frac{60}{152} \times \frac{44}{152} = 17.37$	$152 \times \frac{60}{152} \times \frac{47}{152} = 18.55$	60
B.Eng.	$152 \times \frac{31}{152} \times \frac{61}{152} = 12.44$	$152 \times \frac{31}{152} \times \frac{44}{152} = 8.97$	$152 \times \frac{31}{152} \times \frac{47}{152} = 9.59$	31
B.B.A.	$152 \times \frac{39}{152} \times \frac{61}{152} = 15.65$	$152 \times \frac{39}{152} \times \frac{44}{152} = 11.29$	$152 \times \frac{39}{152} \times \frac{47}{152} = 12.06$	39
Other	$152 \times \frac{22}{152} \times \frac{61}{152} = 8.83$	$152 \times \frac{22}{152} \times \frac{44}{152} = 6.37$	$152 \times \frac{22}{152} \times \frac{47}{152} = 6.80$	22
Total	61	44	47	152

As you can see, the expected value for each cell is computed by multiplying the row total by the column total and dividing by the sample size. For example, the BA and Accounting cell expected value is

$$152 \times \frac{60}{152} \times \frac{61}{152} = \frac{60 \times 61}{152} = 24.08$$

All the other expected values would be determined similarly.

Expected Frequencies for a Contingency Table

The expected frequency of the cell in row i and column j is

$$e_{ij} = \frac{\text{Row } i \text{ total} \times \text{Column } j \text{ total}}{\text{Sample size}}$$

The expected cell frequencies are shown in parentheses in the following table. As in the case of the goodness-of-fit test, the expected cell frequencies should satisfy the rule of five.

	MBA Major		
Undergraduate Degree	Accounting	Finance	Marketing
B.A.	31 (24.08)	13 (17.37)	16 (18.55)
B.Eng.	8 (12.44)	16 (8.97)	7 (9.59)
B.B.A.	12 (15.65)	10 (11.29)	17 12.06)
Other	10 (8.83)	5 (6.37)	7 (6.80)

We can now calculate the value of the test statistic:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} = \frac{(31 - 24.08)^2}{24.08} + \frac{(13 - 17.37)^2}{17.37} + \frac{(16 - 18.55)^2}{18.55}$$

$$+ \frac{(8 - 12.44)^2}{12.44} + \frac{(16 - 8.97)^2}{8.97} + \frac{(7 - 9.59)^2}{9.59} + \frac{(12 - 15.65)^2}{15.65}$$

$$+ \frac{(10 - 11.29)^2}{11.29} + \frac{(17 - 12.06)^2}{12.06} + \frac{(10 - 8.83)^2}{8.83}$$

$$+ \frac{(5 - 6.37)^2}{6.37} + \frac{(7 - 6.80)^2}{6.80}$$

$$= 14.70$$

Notice that we continue to use a single subscript in the formula of the test statistic when we should use two subscripts, one for the rows and one for the columns. We believe that it is clear, that for each cell we must calculate the squared difference between the observed and expected frequencies divided by the expected frequency. We don't believe that the satisfaction of using the mathematically correct notation overcomes the unnecessary complication.

Rejection Region and *p*-Value

To determine the rejection region we must know the number of degrees of freedom associated with the chi-squared statistic. The number of degrees of freedom for a contingency table with r rows and c columns is $\nu = (r - 1)(c - 1)$. For this example, the number of degrees of freedom is $\nu = (r - 1)(c - 1) = (4 - 1)(3 - 1) = 6$.

If we employ a 5% significance level, the rejection region is

$$\chi^2 > \chi^2_{\alpha,\nu} = \chi^2_{.05,6} = 12.6$$

Because $\chi^2 = 14.70$, we reject the null hypothesis and conclude that there is evidence of a relationship between undergraduate degree and MBA major.

The p-value of the test statistic is

$$P(\chi^2 > 14.70)$$

Unfortunately, we cannot determine the p-value manually.

Using the Computer

Excel and Minitab can produce the chi-squared statistic either from a cross-classification table whose frequencies have already been calculated or from raw data. The respective printouts are almost identical.

File Xm15-02 contains the raw data using the following codes:

Column1 (Undergraduate Degree)	Column 2 (MBA Major)
1 = B.A.	1 = Accounting
2 = B.Eng.	2 = Finance
3 = B.B.A.	3 = Marketing
4 = Other	

EXCEL

	A	B	C	D	E	F
1	Contingency Table					
2						
3		*Degree*				
4	*MBA Major*		1	2	3	TOTAL
5		1	31	13	16	60
6		2	8	16	7	31
7		3	12	10	17	39
8		4	10	5	7	22
9		TOTAL	61	44	47	152
10						
11						
12		chi-squared Stat			14.70	
13		df			6	
14		p-value			0.0227	
15		chi-squared Critical			12.5916	

INSTRUCTIONS (RAW DATA)

1. Type or import the data into two adjacent columns*. (Open Xm15-02.) The codes must be positive integers greater than 0.
2. Click **Add-Ins**, **Data Analysis Plus**, and **Contingency Table (Raw Data)**.
3. Specify the **Input Range** (A1:B153) and specify the value of α(.05).

INSTRUCTIONS (COMPLETED TABLE)

1. Type the frequencies into adjacent columns.
2. Click **Add-Ins**, **Data Analysis Plus**, and **Contingency Table**.
3. Specify the **Input Range**. Click **Labels** if the first row and first column of the input range contain the names of the categories. Specify the value for α.

MINITAB

Tabulated statistics: Degree, MBA Major

Rows: Degree Columns: MBA Major

	1	2	3	All
1	31	13	16	60
2	8	16	7	31
3	12	10	17	39
4	10	5	7	22
All	61	44	47	152

Cell Contents: Count

Pearson Chi-Square = 14.702, DF = 6, P-Value = 0.023
Likelihood Ratio Chi-Square = 13.781, DF = 6, P-Value = 0.032

INSTRUCTIONS (RAW DATA)

1. Type or import the data into two columns. (Open Xm15-02.)
2. Click **Stat**, **Tables**, and **Cross Tabulation and Chi-Square . . . .**

(Continued)

*If one or both columns contain a blank (representing missing data) the row must be deleted.

3. In the **Categorical variables** box, select or type the variables **For rows** (Degree) and **For columns** (MBA Major). Click **Chi-Square . . . .**
4. Under **Display** click **Chi-Square analysis.** Specify **Chi-Square analysis**.

INSTRUCTIONS (COMPLETED TABLE)

1. Type the observed frequencies into adjacent columns.
2. Click **Stat**, **Tables**, and **Chi-Square Test (Table in Worksheet) . . . .**
3. Select or type the names of the variables representing the columns.

INTERPRET

There is strong evidence to infer that the undergraduate degree and MBA major are related. This suggests that the dean can predict the number of optional courses by counting the number of MBA students with each type of undergraduate degree. We can see that BA's favor accounting courses, BEng's prefer finance, BBA's are partial to marketing, and others show no particular preference.

If the null hypothesis is true, undergraduate degree and MBA major are independent of one another. This means that whether an MBA student earned a BA, BEng, BBA, or other degree does not affect his or her choice of major program in the MBA. Consequently, there is no difference in major choice among the graduates of the undergraduate programs. If the alternative hypothesis is true, undergraduate degree does affect the choice of MBA major. Thus, there are differences between the four undergraduate degree categories.

Rule of Five

In the previous section, we pointed out that the expected values should be at least 5 to ensure that the chi-squared distribution provides an adequate approximation of the sampling distribution. In a contingency table where one or more cells have expected values of less than 5, we need to combine rows or columns to satisfy the rule of five. This subject is discussed in Keller's website Appendix Rule of Five.

Data Formats

In Example 15.2, the data were stored in two columns, one column containing the values of one nominal variable and the second column storing the values of the second nominal variable. The data can be stored in another way. In Example 15.2, we could have recorded the data in three columns, one column for each MBA major. The columns would contain the codes representing the undergraduate degree. Alternatively, we could have stored the data in four columns, one column for each undergraduate degree. The columns would contain the codes for the MBA majors. In either case, we have to count the number of each value and construct the cross-tabulation table using the counts. Both Excel and Minitab can calculate the chi-squared statistic and its p-value from the cross-tabulation table. We will illustrate this approach with the solution to the chapter-opening example.

General Social Surveys

Has Support for Capital Punishment for Murderers Changed since 2002?

© Bob Daemmerich/PhotoEdit

IDENTIFY

The problem objective is to compare public opinion in four different years. The variable is nominal since its values are Favor and Oppose represented by 1 and 2, respectively. The appropriate technique is the chi-squared test of a contingency table. The hypotheses are

H_0 : The two variables are independent
H_1 : The two variables are dependent

In this application, the two variables are year (2002, 2004, 2006, and 2008) and the answer to the question posed by the General Social Survey (Favor and Oppose).

Unlike Example 15.2, the data are not stored in two columns. To produce the statistical result we will need to count the number of Americans in favor and the number opposed in each of the four years. The following table was determined by counting the numbers of 1's and 2's for each year.

	Year			
	2002	2004	2006	2008
Favor	899	855	1,885	1,263
Oppose	409	402	930	639

EXCEL

	A	B	C	D	E	F	G
1	Contingency Table						
2							
3							
4			Year 2002	Year 2004	Year 2006	Year 2008	TOTAL
5		Favor	899	855	1885	1263	4902
6		Oppose	409	402	930	639	2380
7		TOTAL	1308	1257	2815	1902	7282
8							
9							
10		Chi-Squared Statistic		2.3517			
11		Degrees Of Freedom		3			
12		P-Value		0.5027			
13		Chi-Squared Critical		7.8147			

INTERPRET

The p-value is .5027. There is not enough evidence to infer that the two variables are independent. Thus, there is not enough evidence to conclude that support for capital punishment for murder varies from year to year.

Here is a summary of the factors that tell us when to apply the chi-squared test of a contingency table. Note that there are two problem objectives satisfied by this statistical procedure.

Factors That Identify the Chi-Squared Test of a Contingency Table

1. **Problem objectives**: Analyze the relationship between two variables and compare two or more populations
2. **Data type**: Nominal

EXERCISES

Developing an Understanding of Statistical Concepts

15.22 Conduct a test to determine whether the two classifications L and M are independent, using the data in the accompanying cross-classification table. (Use $\alpha = .05$.)

	M_1	M_2
L_1	28	68
L_2	56	36

15.23 Repeat Exercise 15.22 using the following table:

	M_1	M_2
L_1	14	34
L_2	28	18

15.24 Repeat Exercise 15.22 using the following table:

	M_1	M_2
L_1	7	17
L_2	14	9

15.25 Review the results of Exercises 15.22–15.24. What is the effect of decreasing the sample size?

15.26 Conduct a test to determine whether the two classifications R and C are independent, using the data in the accompanying cross-classification table. (Use $\alpha = .10$.)

	C_1	C_1	C_3
R_1	40	32	48
R_2	30	48	52

Applications

Use a 5% significance level unless specified otherwise.

15.27 The trustee of a company's pension plan has solicited the opinions of a sample of the company's employees about a proposed revision of the plan. A breakdown of the responses is shown in the accompanying table. Is there enough evidence to infer that the responses differ between the three groups of employees?

Responses	Blue-Collar Workers	White-Collar Workers	Managers
For	67	32	11
Against	63	18	9

15.28 The operations manager of a company that manufactures shirts wants to determine whether there are differences in the quality of workmanship among the three daily shifts. She randomly selects 600 recently made shirts and carefully inspects them. Each shirt is classified as either perfect or flawed, and the shift that produced it is also recorded. The accompanying table summarizes the number of shirts that fell into each cell. Do these data provide sufficient evidence to infer that there are differences in quality between the three shifts?

	Shift		
Shirt Condition	1	2	3
Perfect	240	191	139
Flawed	10	9	11

15.29 One of the issues that came up in a recent national election (and is likely to arise in many future elections) is how to deal with a sluggish economy. Specifically, should governments cut spending, raise taxes, inflate the economy (by printing more money) or do none of the above and let the deficit rise? And as with most other issues, politicians need to know which parts of the electorate support these options. Suppose that a random sample of 1,000 people was asked which option they support and their political affiliations. The possible responses to the question about political affiliation were Democrat, Republican, and Independent (which included a variety of political persuasions). The responses are summarized in the accompanying table. Do these results allow us to conclude at the 1% significance level that political affiliation affects support for the economic options?

Economic Options	Political Affiliation		
	Democrat	Republican	Independent
Cut spending	101	282	61
Raise taxes	38	67	25
Inflate the economy	131	88	31
Let deficit increase	61	90	25

15.30 Econetics Research Corporation, a well-known Montreal-based consulting firm, wants to test how it can influence the proportion of questionnaires returned from surveys. In the belief that the inclusion of an inducement to respond may be important, the firm sends out 1,000 questionnaires: Two hundred promise to send respondents a summary of the survey results, 300 indicate that 20 respondents (selected by lottery) will be awarded gifts, and 500 are accompanied by no inducements. Of these, 80 questionnaires promising a summary, 100 questionnaires offering gifts, and 120 questionnaires offering no inducements are returned. What can you conclude from these results?

Exercises 15.31–15.46 require the use of a computer and software. Use a 5% significance level unless specified otherwise. The answers to Exercises 15.31–15.38 may be calculated manually. See Appendix A for the sample statistics.

15.31 Xm02-04 (Example 2.4 revisited) A major North American city has four competing newspapers: the *Globe and Mail* (G&M), *Post*, *Sun*, and *Star*. To help design advertising campaigns, the advertising managers of the newspapers need to know which segments of the newspaper market are reading their papers. A survey was conducted to analyze the relationship between newspapers read and occupation. A sample of newspaper readers was asked to report which newspaper they read: *Globe and Mail* (1) *Post* (2), *Star* (3), *Sun* (4), and to indicate whether they were blue-collar workers (1), white-collar workers (2), or professionals (3). Can we infer that occupation and newspaper are related?

15.32 Xr15-32 An investor who can correctly forecast the direction and size of changes in foreign currency exchange rates is able to reap huge profits in the international currency markets. A knowledgeable reader of the *Wall Street Journal* (in particular, of the currency futures market quotations) can determine the direction of change in various exchange rates that is predicted by all investors, viewed collectively. Predictions from 216 investors, together with the subsequent actual directions of change, were recorded in the following way: Column 1: predicted change where 1 = positive and 2 = negative; column 2: actual change where 1 = positive and 2 = negative.

a. Can we infer at the 10% significance level that a relationship exists between the predicted and actual directions of change?
b. To what extent would you make use of these predictions in formulating your forecasts of future exchange rate changes?

15.33 Xr02-43 (Exercise 2.43 revisited) Is there brand loyalty among car owners in their purchases of gasoline? To help answer the question, a random sample of car owners was asked to record the brand of gasoline in their last two purchases: 1 = Exxon, 2 = Amoco, 3 = Texaco, 4 = Other. Can we conclude that there is brand loyalty in gasoline purchases?

15.34 Xr15-34 During the past decade, many cigarette smokers have attempted to quit. Unfortunately, nicotine is highly addictive. Smokers use a large number of different methods to help them quit. These include nicotine patches, hypnosis, and various forms of therapy. A researcher for the Addiction Research Council wanted to determine why some people quit while others attempted to quit but failed. He surveyed 1,000 people who planned to quit smoking. He determined their educational level and whether they continued to smoke 1 year later. Educational level was recorded in the following way:

1 = Did not finish high school
2 = High school graduate
3 = University or college graduate
4 = Completed a postgraduate degree

A continuing smoker was recorded as 1; a quitter was recorded as 2. Can we infer that the amount of education is a factor in determining whether a smoker will quit?

15.35 Xr15-35 Because television audiences of newscasts tend to be older (and because older people suffer from a variety of medical ailments), pharmaceutical companies' advertising often appears on national news on the three networks (ABC, CBS, and NBC). To determine how effective the ads are a survey was undertaken. Adults over 50 were asked about their primary sources of news. The responses are

1. ABC News	2. CBS News	3. NBC News
4. Newspapers	5. Radio	6. None of the above

Each person was also asked whether they suffer from heartburn, and if so, what remedy they take. The answers were recorded as follows:

1. Do not suffer from heartburn
2. Suffer from heartburn but take no remedy

3. Suffer from heartburn and take an over-the-counter remedy (e.g., Tums, Gavoscol)
4. Suffer from heartburn and take a prescription pill (e.g., Nexium)

Is there a relationship between an adult's source of news and his or her heartburn condition?

15.36 Xr02-42 (Exercise 2.42 revisited) The associate dean of a business school was looking for ways to improve the quality of the applicants to its MBA program. In particular, she wanted to know whether the undergraduate degree of applicants differed among her school and the three nearby universities with MBA programs. She sampled 100 applicants of her program and an equal number from each of the other universities. She recorded their undergraduate degrees (1 = BA, 2 = BEng, 3 = BBA, 4 = other) as well as universities (codes 1, 2, 3, and 4). Do these data provide sufficient evidence to infer that undergraduate degree and the university each person applied are related?

15.37 Xr15-37 The relationship between drug companies and medical researchers is under scrutiny because of possible conflict of interest. The issue that started the controversy was a 1995 case control study that suggested that the use of calcium-channel blockers to treat hypertension led to an increase risk of heart disease. This led to an intense debate both in technical journals and in the press. Researchers writing in the *New England Journal of Medicine* ("Conflict of Interest in the Debate over Calcium Channel Antagonists," January 8, 1998, p. 101) looked at the 70 reports that appeared during 1996–1997, classifying them as favorable, neutral, or critical toward the drugs. The researchers then contacted the authors of the reports and questioned them about financial ties to drug companies. The results were recorded in the following way:

Column 1: Results of the scientific study; 1 = favorable, 2 = neutral, 3 = critical
Column 2: 1 = financial ties to drug companies, 2 = no ties to drug companies

Do these data allow us to infer that the research findings for calcium-channel blockers are affected by whether the research is funded by drug companies?

15.38 Xr15-38 After a thorough analysis of the market, a publisher of business and economics statistics books has divided the market into three general approaches to teach applied statistics. These are (1) use of a computer and statistical software with no manual calculations, (2) traditional teaching of concepts and solution of problems by hand, and (3) mathematical approach with emphasis on derivations and proofs. The publisher wanted to know whether this market could be segmented on the basis of the educational background of the instructor. As a result, the statistics editor organized a survey that asked 195 professors of business and economics statistics to report their approach to teaching and which one of the following categories represents their highest degree:

1. Business (MBA or Ph.D. in business)
2. Economics
3. Mathematics or engineering
4. Other

a. Can the editor infer that there are differences in type of degree among the three teaching approaches? If so, how can the editor use this information?
b. Suppose that you work in the marketing department of a textbook publisher. Prepare a report for the editor that describes this analysis.

General Social Survey Exercises

15.39 GSS2002* GSS2004* GSS2006* GSS2008* The issue of gun control in the United States is often debated, particularly during elections. The question arises, What does the public think about the issue and does support vary from year to year? Test to determine whether there is enough evidence to conclude that support for gun laws (GUNLAW) varied from year to year.

15.40 GSS2002* GSS2004* GSS2006* GSS2008* Can we conclude that Americans' marital status (MARITAL) distribution has changed from year to year?

15.41 GSS2008* In the last two decades, an increasing proportion of women have entered the workforce. Determine whether there is enough evidence to conclude that men and women (SEX) differ in their work status (WRKSTA).

15.42 GSS2008* Is there sufficient evidence to infer that support for capital punishment (CAPPUN) is related to political affiliation (PARTYID3: 1 = Democrat, 2 = Republican, 3 = Independent)?

American National Election Survey Exercises

For each of the following variables, conduct a test to determine whether there are differences between the three political party affiliations (PARTY: 1 = Democrat, 2 = Republican, 3 = Independent).

15.43 ANES2008* Know where to vote (KNOW)

15.44 ANES2008* Read about campaign in newspaper (READ)

15.45 ANES2008* Have health insurance (HEALTH)

15.46 ANES2008* Have access to the Internet (ACCESS)

15.3 Summary of Tests on Nominal Data

At this point in the textbook, we've described four tests that are used when the data are nominal:

z-test of p (Section 12.3)
z-test of $p_1 - p_2$ (Section 13.5)
Chi-squared goodness-of-fit test (Section 15.1)
Chi-squared test of a contingency table (Section 15.2)

In the process of presenting these techniques, it was necessary to concentrate on one technique at a time and focus on the kinds of problems each addresses. However, this approach tends to conflict somewhat with our promised goal of emphasizing the "when" of statistical inference. In this section, we summarize the statistical tests on nominal data to ensure that you are capable of selecting the correct method.

There are two critical factors in identifying the technique used when the data are nominal. The first, of course, is the problem objective. The second is the number of categories that the nominal variable can assume. Table 15.1 provides a guide to help select the correct technique.

TABLE **15.1** Statistical Techniques for Nominal Data

PROBLEM OBJECTIVE	NUMBER OF CATEGORIES	STATISTICAL TECHNIQUE
Describe a population	2	z-test of p or the chi-squared goodness-of-fit test
Describe a population	More than 2	Chi-squared goodness-of-fit test
Compare two populations	2	z-test of $p_1 - p_2$ or chi-squared test of a contingency table
Compare two populations	More than 2	Chi-squared test of a contingency table
Compare two or more populations	2 or more	Chi-squared test of a contingency table
Analyze the relationship between two variables	2 or more	Chi-squared test of a contingency table

Notice that when we describe a population of nominal data with exactly two categories, we can use either of two techniques. We can employ the z-test of p or the chi-squared goodness-of-fit test. These two tests are equivalent because if there are only two categories, the multinomial experiment is actually a binomial experiment (one of the categorical outcomes is labeled *success*, and the other is labeled *failure*). Mathematical statisticians have established that if we square the value of z, the test statistic for the test of p, we produce the χ^2-statistic; that is, $z^2 = \chi^2$. Thus, if we want to conduct a two-tail test of a population proportion, we can employ either technique. However, the chi-squared goodness-of-fit test can test only to determine whether the hypothesized values of p_1 (which we can label p) and p_2 (which we call $1 - p$) are not equal to their specified values. Consequently, to perform a one-tail test of a population proportion, we must use the z-test of p. (This issue was discussed in Chapter 14 when we pointed out that we can use either the t-test of $\mu_1 - \mu_2$ or the analysis of variance to conduct a test to determine whether two population means differ.)

When we test for differences between two populations of nominal data with two categories, we can also use either of two techniques: the z-test of $p_1 - p_2$ (Case 1) or the chi-squared test of a contingency table. Once again, we can use either technique to perform a two-tail test about $p_1 - p_2$. (Squaring the value of the z-statistic yields the value of the χ^2-statistic.) However, one-tail tests must be conducted by the z-test of $p_1 - p_2$. The rest of the table is quite straightforward. Notice that when we want to compare two populations when there are more than two categories, we use the chi-squared test of a contingency table.

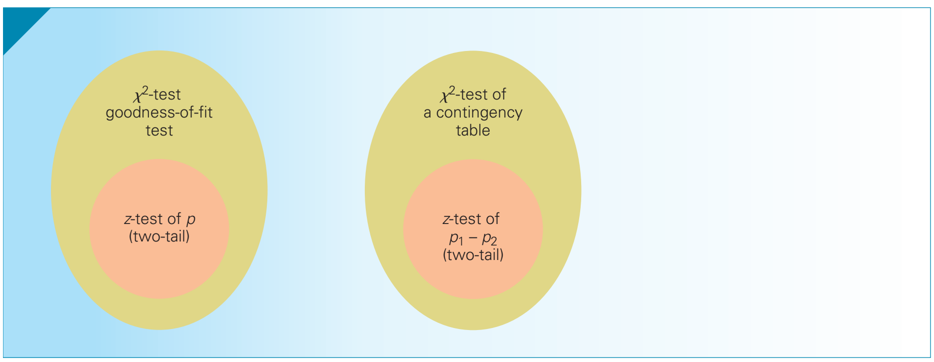
Figure 15.4 offers another summary of the tests that deal with nominal data introduced in this book. There are two groups of tests: those that test hypotheses about single populations and those that test either for differences or for independence. In the first set, we have the z-test of p, which can be replaced by the chi-squared test of a multinomial experiment. The latter test is employed when there are more than two categories.

To test for differences between two proportions, we apply the z-test of $p_1 - p_2$. Instead we can use the chi-squared test of a contingency table, which can be applied to a variety of other problems.

Developing an Understanding of Statistical Concepts

Table 15.1 and Figure 15.4 summarize how we deal with nominal data. We determine the frequency of each category and use these frequencies to compute test statistics. We can then compute proportions to calculate z-statistics or use the frequencies to calculate

FIGURE **15.4** Tests on Nominal Data



χ^2-statistics. Because squaring a standard normal random variable produces a chi-squared variable, we can employ either statistic to test for differences. As a consequence, when you encounter nominal data in the problems described in this book (and other introductory applied statistics books), the most logical starting point in selecting the appropriate technique will be either a z-statistic or a χ^2-statistic. However, you should know that there are other statistical procedures that can be applied to nominal data, techniques that are not included in this book.

15.4 (Optional) Chi-Squared Test for Normality

We can use the goodness-of-fit test presented in Section 15.1 in another way. We can test to determine whether data were drawn from any distribution. The most common application of this procedure is a test of normality.

In the examples and exercises shown in Section 15.1, the probabilities specified in the null hypothesis were derived from the question. In Example 15.1, the probabilities p_1, p_2, and p_3 were the market shares before the advertising campaign. To test for normality (or any other distribution), the probabilities must first be calculated using the hypothesized distribution. To illustrate, consider Example 12.1, where we tested the mean amount of discarded newspaper using the Student t distribution. The required condition for this procedure is that the data must be normally distributed. To determine whether the 148 observations in our sample were indeed taken from a normal distribution, we must calculate the theoretical probabilities assuming a normal distribution. To do so, we must first calculate the sample mean and standard deviation: $\bar{x} = 2.18$ and $s = .981$. Next, we find the probabilities of an arbitrary number of intervals. For example, we can find the probabilities of the following intervals:

Interval 1: $X \leq .709$

Interval 2: $.709 < X \leq 1.69$

Interval 3: $1.69 < X \leq 2.67$

Interval 4: $2.67 < X \leq 3.65$

Interval 5: $X > 3.65$

We will discuss the reasons for our choices of intervals later.

The probabilities are computed using the normal distribution and the values of $\bar{x}$ and s as estimators of μ and σ. We calculated the sample mean and standard deviation as $\bar{x} = 2.18$ and $s = .981$. Thus,

$$P(X \leq .709) = P\left(\frac{X-\mu}{\sigma} \leq \frac{.709 - 2.18}{.981}\right) = P(Z \leq -1.5) = .0668$$

$$P(.709 < X \leq 1.69) = P\left(\frac{.709 - 2.18}{.981} < \frac{X-\mu}{\sigma} \leq \frac{1.69 - 2.18}{.981}\right)$$

$$= P(-1.5 < Z \leq -.5) = .2417$$

$$P(1.69 < X \leq 2.67) = P\left(\frac{1.69 - 2.18}{.981} < \frac{X-\mu}{\sigma} \leq \frac{2.67 - 2.18}{.981}\right)$$

$$= P(-.5 < Z \leq .5) = .3829$$

$$P(2.67 < X \leq 3.65) = P\left(\frac{2.67 - 2.18}{.981} < \frac{X - \mu}{\sigma} \leq \frac{3.65 - 2.18}{.981}\right)$$

$$= P(.5 < Z \leq 1.5) = .2417$$

$$P(X > 3.65) = P\left(\frac{X - \mu}{\sigma} > \frac{3.65 - 2.18}{.981}\right) = P(Z > 1.5) = .0668$$

To test for normality is to test the following hypotheses:

H_0: $p_1 = .0668, p_2 = .2417, p_3 = .3829, p_4 = .2417, p_5 = .0668$

H_1: At least two proportions differ from their specified values

We complete the test as we did in Section 15.1, except that the number of degrees of freedom associated with the chi-squared statistic is the number of intervals minus 1 minus the number of parameters estimated, which in this illustration is two. (We estimated the population mean μ and the population standard deviation σ.) Thus, in this case, the number of degrees of freedom is $k - 1 - 2 = 5 - 1 - 2 = 2$.

The expected values are

$$e_1 = np_1 = 148(.0668) = 9.89$$
$$e_2 = np_2 = 148(.2417) = 35.78$$
$$e_3 = np_3 = 148(.3829) = 56.67$$
$$e_4 = np_4 = 148(.2417) = 35.78$$
$$e_5 = np_5 = 148(.0668) = 9.89$$

The observed values are determined manually by counting the number of values in each interval. Thus,

$$f_1 = 10$$
$$f_2 = 36$$
$$f_3 = 54$$
$$f_4 = 39$$
$$f_5 = 9$$

The chi-squared statistic is

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i} = \frac{(10 - 9.89)^2}{9.89} + \frac{(36 - 35.78)^2}{35.78} + \frac{(54 - 56.67)^2}{56.67} + \frac{(39 - 35.78)^2}{35.78} + \frac{(9 - 9.89)^2}{9.89}$$
$$= .50$$

The rejection region is

$$\chi^2 > \chi^2_{\alpha,k-3} = \chi^2_{.05,2} = 5.99$$

There is not enough evidence to conclude that these data are not normally distributed.

Class Intervals

In practice you can use any intervals you like. We chose the intervals we did to facilitate the calculation of the normal probabilities. The number of intervals was chosen to comply with the rule of five, which requires that all expected values be at least equal to 5. Because the number of degrees of freedom is $k - 3$, the minimum number of intervals is $k = 4$.

Using the Computer

EXCEL

	A	B	C	D
1	Chi-Squared Test of Normality			
2				
3		*Newspaper*		
4	Mean	2.18		
5	Standard deviation	0.981		
6	Observations	148		
7				
8	Intervals	Probability	Expected	Observed
9	(z <= −1.5)	0.0668	9.89	10
10	(−1.5 < z <= −0.5)	0.2417	35.78	36
11	(−0.5 < z <= 1.5)	0.3829	56.67	54
12	(0.5 < z <=1.5)	0.2417	35.78	39
13	(z > 1.5)	0.0668	9.89	9
14				
15				
16	chi-squared Stat	0.50		
17	df	2		
18	p-value	0.7792		
19	chi-squared Critical	5.9915		

We programmed Excel to calculate the value of the test statistic so that the expected values are at least 5 (where possible) and the minimum number of intervals is 4. Hence, if the number of observations is more than 220, the intervals and probabilities are

Interval	Probability
$Z \le -2$	.0228
$-2 < Z \le -1$	.1359
$-1 < Z \le 0$	.3413
$0 < Z \le 1$	.3413
$1 < Z \le 2$	.1359
$Z > 2$	.0228

If the sample size is less than or equal to 220 and greater than 80, the intervals are

Interval	Probability
$Z \le -1.5$	.0668
$-1.5 < Z \le -0.5$	.2417
$-0.5 < Z \le 0.5$	.3829
$0.5 < Z \le 1.5$	.2417
$Z > 1.5$	.0668

If the sample size is less than or equal to 80, we employ the minimum number of intervals, 4. When the sample size is less than 32, at least one expected value will be less than 5. The intervals are

(Continued)

Interval	Probability
$Z \leq -1$	.1587
$-1 < Z \leq 0$	.3413
$0 < Z \leq 1$	.3413
$Z > 1$	.1587

INSTRUCTIONS

1. Type or import the data into one column. (Open Xm12-01.)
2. Click **Add-Ins**, **Data Analysis Plus**, and **Chi-Squared Test of Normality**.
3. Specify the **Input Range** (A1:A149) and the value of α(.05).

MINITAB

Minitab does not conduct this procedure.

Interpreting the Results of a Chi-Squared Test for Normality

In the example above, we found that there was little evidence to conclude that the weight of discarded newspaper is not normally distributed. However, had we found evidence of nonnormality, this would not necessarily invalidate the *t*-test we conducted in Example 12.1. As we pointed out in Chapter 12, the *t*-test of a mean is a robust procedure, which means that only if the variable is extremely nonnormal and the sample size is small can we conclude that the technique is suspect. The problem here is that if the sample size is large and the variable is only slightly nonnormal, the chi-squared test for normality will, in many cases, conclude that the variable is not normally distributed. However, if the variable is even quite nonnormal and the sample size is large, the *t*-test will still be valid. Although there are situations in which we need to know whether a variable is nonnormal, we continue to advocate that the way to decide if the normality requirement for almost all statistical techniques applied to interval data is satisfied is to draw histograms and look for shapes that are far from bell shaped (e.g., highly skewed or bimodal). We will use this approach in Chapter 19 when we introduce nonparametric techniques that are used when interval data are nonnormal.

EXERCISES

15.47 Suppose that a random sample of 100 observations was drawn from a population. After calculating the mean and standard deviation, each observation was standardized and the number of observations in each of the following intervals was counted. Can we infer at the 5% significance level that the data were not drawn from a normal population?

Interval	Frequency
$Z \leq 1.5$	10
$-1.5 < Z \leq -0.5$	18
$-0.5 < Z \leq 0.5$	48
$0.5 < Z \leq 1.5$	16
$Z > 1.5$	8

15.48 A random sample of 50 observations yielded the following frequencies for the standardized intervals:

Interval	Frequency
$Z \le -1$	6
$-1 < Z \le 0$	27
$0 < Z \le 1$	14
$Z > 1$	3

Can we infer that the data are not normal? (Use $\alpha = .10$.)

The following exercises require the use of a computer and software.

15.49 Xr12-31 Refer to Exercise 12.31. Test at the 10% significance level to determine whether the amount of time spent working at part-time jobs is normally distributed. If there is evidence of nonnormality, is the t-test invalid?

15.50 Xr12-37 The t-test in Exercise 12.37 requires that the costs of prescriptions is normally distributed. Conduct a test with $\alpha = .05$ to determine whether the required condition is unsatisfied. If there is enough evidence to conclude that the requirement is not satisfied, does this indicate that the t-test is invalid?

15.51 Xr13-25 Exercise 13.25 required you to conduct a t-test of the difference between two means. Each sample's productivity data are required to be normally distributed. Is that required condition violated? Test with $\alpha = .05$.

15.52 Xr13-26 Exercise 13.26 asked you to conduct a t-test of the difference between two means (reaction times). Test to determine whether there is enough evidence to infer that the reaction times are not normally distributed. A 5% significance level is judged to be suitable.

15.53 Xr13-59 In Exercise 13.59, you performed a test of the mean matched pairs difference. The test result depends on the requirement that the differences are normally distributed. Test with a 10% significance level to determine whether the requirement is violated.

Chapter Summary

This chapter introduced three statistical techniques. The first is the chi-squared goodness-of-fit test, which is applied when the problem objective is to describe a single population of nominal data with two or more categories. The second is the chi-squared test of a contingency table. This test has two objectives: to analyze the relationship between two nominal variables and to compare two or more populations of nominal data. The last procedure is designed to test for normality.

IMPORTANT TERMS

Multinomial experiment 597
Chi-squared goodness-of-fit test 598
Expected frequency 598
Observed frequencies 599
Cross-classification table 604
Chi-squared test of a contingency table 604
Contingency table 607

SYMBOLS

Symbol	Pronounced	Represents
f_i	f sub i	Frequency of the ith category
e_i	e sub i	Expected value of the ith category
χ^2	Chi squared	Test statistic

FORMULA

Test statistic for all procedures

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

COMPUTER OUTPUT AND INSTRUCTIONS

Technique	Excel	Minitab
Chi-squared goodness-of-fit test	600	601
Chi-squared test of a contingency table (raw data)	609	609
Chi-squared test of a contingency table	609	609
Chi-squared test of normality	619	620

CHAPTER EXERCISES

Use a 5% significance level unless specified otherwise.

15.54 An organization dedicated to ensuring fairness in television game shows is investigating *Wheel of Fortune*. In this show, three contestants are required to solve puzzles by selecting letters. Each contestant gets to select the first letter and continues selecting until he or she chooses a letter that is not in the hidden word, phrase, or name. The order of contestants is random. However, contestant 1 gets to start game 1, contestant 2 starts game 2, and so on. The contestant who wins the most money is declared the winner, and he or she is given an opportunity to win a grand prize. Usually, more than three games are played per show, and as a result it appears that contestant 1 has an advantage: Contestant 1 will start two games, whereas contestant 3 will usually start only one game. To see whether this is the case, a random sample of 30 shows was taken, and the starting position of the winning contestant for each show was recorded. These are shown in the following table:

Starting position	1	2	3
Number of wins	14	10	6

Do the tabulated results allow us to conclude that the game is unfair?

15.55 It has been estimated that employee absenteeism costs North American companies more than $100 billion per year. As a first step in addressing the rising cost of absenteeism, the personnel department of a large corporation recorded the weekdays during which individuals in a sample of 362 absentees were away over the past several months. Do these data suggest that absenteeism is higher on some days of the week than on others?

Day of the week	Monday	Tuesday	Wednesday	Thursday	Friday
Number absent	87	62	71	68	74

15.56 Suppose that the personnel department in Exercise 15.55 continued its investigation by categorizing absentees according to the shift on which they worked, as shown in the accompanying table. Is there sufficient evidence at the 10% significance level of a relationship between the days on which employees are absent and the shift on which the employees work?

Shift	Monday	Tuesday	Wednesday	Thursday	Friday
Day	52	28	37	31	33
Evening	35	34	34	37	41

15.57 A management behavior analyst has been studying the relationship between male–female supervisory structures in the workplace and the level of employees' job satisfaction. The results of a recent survey are shown in the accompanying table. Is there sufficient evidence to infer that the level of job satisfaction depends on the boss–employee gender relationship?

	Boss/Employee			
Level of Satisfaction	Female/ Male	Female/ Female	Male/ Male	Male/ Female
Satisfied	21	25	54	71
Neutral	39	49	50	38
Dissatisfied	31	48	10	11

The following exercises require the use of a computer and software. The answers may be calculated manually. See Appendix A for the sample statistics. Use a 5% significance level unless specified otherwise.

15.58 Xr15-58 Stress is a serious medical problem that costs businesses and government billions of dollars annually. As a result, it is important to determine the causes and possible cures. It would be helpful to know whether the causes are universal or if they vary from country to country. In a survey, American and Canadian adults were asked to report their primary source of stress in their lives. The responses are

1 = Job, 2 = Finances, 3 = Health,
4 = Family life, 5 = Other

The data were recorded using the codes above plus 1 = American and 2 = Canadian. Do these data provide sufficient evidence to conclude that Americans and Canadians differ in their sources of stress?

15.59 Xr15-59 According to NBC News (March 11, 1994) more than 3,000 Americans quit smoking each day. (Unfortunately, more than 3,000 Americans start smoking each day.) Because nicotine is one of the most addictive drugs, quitting smoking is a difficult and frustrating task. It usually takes several tries before success is achieved. There are various methods, including cold turkey, nicotine patches, hypnosis, and group therapy sessions. In an experiment to determine how these methods differ, a random sample of smokers who have decided to quit is selected. Each smoker has chosen one of the methods listed above. After one year, the respondents report whether they have quit (1 = yes and 2 = no) and which method they used (1 = cold turkey, 2 = nicotine patch, 3 = hypnosis, 4 = group therapy sessions). Is there sufficient evidence to conclude that the four methods differ in their success?

15.60 Xr15-60 A newspaper publisher, trying to pinpoint his market's characteristics, wondered whether the way people read a newspaper is related to the reader's educational level. A survey asked adult readers which section of the paper they read first and asked them to report their highest educational level. These data were recorded (Column 1 = first section read where 1 = front page, 2 = sports, 3 = editorial, and 4 = other; and column 2 = educational level where 1 = did not complete high school, 2 = high school graduate, 3 = university or college graduate, and 4 = postgraduate degree). What do these data tell the publisher about how educational level affects the way adults read the newspaper?

15.61 Xr15-61 Every week, the Florida Lottery draws six numbers between 1 and 49. Lottery ticket buyers are naturally interested in whether certain numbers are drawn more frequently than others. To assist players, the *Sun-Sentinel* publishes the number of times each of the 49 numbers has been drawn in the past 52 weeks. The numbers and the frequency with which each occurred were recorded.

a. If the numbers are drawn from a uniform distribution, what is the expected frequency for each number?
b. Can we infer that the data were not generated from a uniform distribution?

In Section 15.4, we showed how to test for normality. However, we can use the same process to test for any other distribution.

15.62 Xr15-62 A scientist believes that the gender of a child is a binomial random variable with probability = .5 for a boy and .5 for a girl. To help test her belief, she randomly samples 100 families with five children. She records the number of boys. Can the scientist infer that the number of boys in families with five children is not a binomial random variable with $p = .5$? (*Hint*: Find the probability of $X = 0, 1, 2, 3, 4,$ and 5 from a binomial distribution with $n = 5$ and $p = .5$).

15.63 Xr15-63 Given the high cost of medical care, research that points the way to avoid illness is welcome. Previously performed research tells us that stress affects the immune system. Two scientists at Carnegie Mellon Hospital in Pittsburgh asked 114 healthy adults about their social circles; they were asked to list every group they had contact with at least once every 2 weeks—family, co-workers, neighbors, friends, and religious and community groups. Participants also reported negative life events over the past year, including the death of a friend or relative, divorce, or job-related problems. The participants were divided into four groups:

Group 1: Highly social and highly stressed
Group 2: Not highly social and highly stressed
Group 3: Highly social and not highly stressed
Group 4: Not highly social and not highly stressed

Each individual was classified in this way. In addition, whether each person contracted a cold over the next 12 weeks was recorded (1 = cold, 2 = no cold). Can we infer that there are differences between the four groups in terms of contracting a cold?

The following exercises employ data files associated with examples and exercises seen earlier in this book.

15.64 Xr12-91* Exercise 12.91 described the problem of a looming shortage of professors, possibly made worse by professors desiring to retire before the age of 65. A survey asked a random sample of professors whether they intended to retire before 65. The responses are "no" (1) "yes" (2). In addition, the survey asked to which faculty each professor belonged (1 = Arts, 2 = Science, 3 = Business, 4 = Engineering, 5 = other). Do these provide sufficient evidence to infer that whether a professor wishes to retire is related to the faculty?

15.65 Xr12-95* Refer to Exercise 12.95. Determine whether there is enough evidence to infer that there are differences in the choice of Christmas tree between the three age categories.

15.66 Xr12-97* Exercise 12.97 described a study to determine whether viewers (older than 50) of the network news had contacted their physician to ask about one of the prescription drugs advertised during the newscast. The responses (1 = no and 2 = yes) were recorded. Also recorded were which of the three networks they normally watch (1 = ABC, 2 = CBS, 3 = NBC). Can we conclude that there are differences in responses between the three network news shows?

15.67 Xr13-110* Exercise 13.110 described a survey of adults wherein, on the basis of several probing questions, each was classified as being or not being a member of the health-conscious group (belonging = 1, not belonging = 2) and whether he or she buys Special X (1 = no, 2 = yes). In addition, his or her educational attainment was recorded (1 = did not finish high school, 2 = finished high school, 3 = finished college or university, 4 = postgraduate degree).

a. Do the data allow the surveyor to conclude that there are differences in educational attainment between those who do and those who do not belong to the health-conscious group?

b. Can we infer that there is a relationship between the four educational groups and whether or not a person buys Special X?

15.68 Xm12-05* Example 12.5 described exit polls wherein people are asked whether they voted for the Democrat or Republican candidate for president. The surveyors also record gender (1 = female, 2 = male), educational attainment (1 = did not finish high school, 2 = completed high school, 3 = completed college or university, 4 = postgraduate degree), and income level (1 = less than $25,000, 2 = $25,000 to $49,999, 3 = $50,000 to $75,000), 4 = more than $75,000).

a. Is there sufficient evidence to infer that voting and gender are related?

b. Do the data allow the conclusion that voting and educational level are related?

c. Can we infer that voting and income are related?

APPLICATIONS in MARKETING

© Richard Grown/Taxi/ Getty Images

Market Segmentation

In Section 12.4 and Chapters 13 and 14, we described how marketing managers use statistical analyses to estimate the size of market segments and determine whether there are differences between segments.

The following exercises require the application of the chi-squared test of a contingency table to determine whether market segments differ with respect to some nominal variable.

15.69 Xr12-126* Exercise 12.126 described the market segments defined by JC Penney. Another question included in the questionnaire that classified the women surveyed asked whether each worked outside the home. The responses were

1. No
2. Part-time job
3. Full-time job

These data plus the classifications (1 = conservative, 2 = traditional, and 3 = contemporary) were recorded. Can we infer from these data that there are differences in employment status between the three market segments?

15.70 Xr12-126* Refer to Exercise 12.126. The women in the survey were also asked to define value by identifying what they considered to be the most important attribute of value. The responses are

1. Price
2. Quality
3. Fashion

The responses and the classifications of segments (1 = conservative, 2 = traditional, and 3 = contemporary) were recorded. Do these data allow us to infer that there are differences in the definition of value between the three market segments?

15.71 Xm12-06* Refer to Example 12.6. In segmenting the breakfast cereal market, a food manufacturer uses health and diet consciousness as the segmentation variable. Four segments are developed:

1. Concerned about eating healthy foods
2. Concerned primarily about weight
3. Concerned about health because of illness
4. Unconcerned

A survey was undertaken, and each person was asked how often they ate a healthy breakfast (defined as cereal with or without fruit). The responses are

1. Never
2. Seldom
3. Often
4. Always

The responses and the market segments of each respondent were recorded. Can we infer that there are differences in frequency of healthy breakfasts between the market segments?

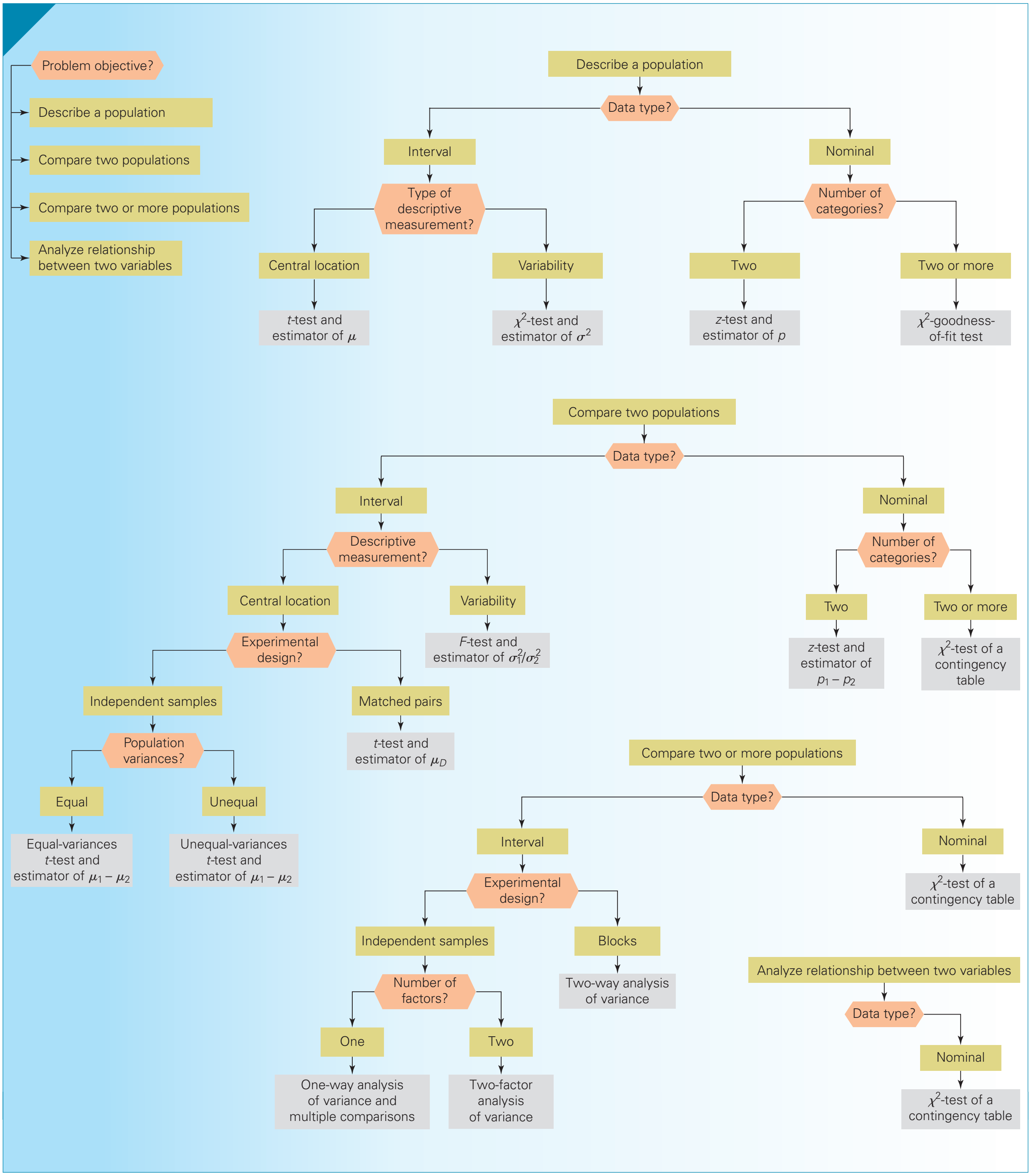
APPENDIX 15 / REVIEW OF CHAPTERS 12 TO 15

Here are the updated list of statistical techniques (Table A15.1) and the flowchart (Figure A15.1) for Chapters 12 to 15. Counting the two techniques of chi-squared tests introduced here (we do not include the chi-squared test for normality), we have covered 22 statistical methods.

TABLE **A15.1** Summary of Statistical Techniques in Chapters 12 to 15

t-test of μ
Estimator of μ (including estimator of $N\mu$)
χ^2-test of σ^2
Estimator of σ^2
z-test of p
Estimator of p (including estimator of Np)
Equal-variances t-test of $\mu_1 - \mu_2$
Equal-variances estimator of $\mu_1 - \mu_2$
Unequal-variances t-test of $\mu_1 - \mu_2$
Unequal-variances estimator of $\mu_1 - \mu_2$
t-test of μ_D
Estimator of μ_D
F-test of σ_1^2/σ_2^2
Estimator of σ_1^2/σ_2^2
z-test of $p_1 - p_2$ (Case 1)
z-test of $p_1 - p_2$ (Case 2)
Estimator of $p_1 - p_2$
One-way analysis of variance (including multiple comparisons)
Two-way (randomized blocks) analysis of variance
Two-factor analysis of variance
χ^2-goodness-of-fit test
χ^2-test of a contingency table

FIGURE **A15.1** Summary of Statistical Techniques in Chapters 12 to 15



EXERCISES

We remind you that we do not specify significance levels in the exercise that follow. Choose your own.

A15.1 XrA15-01 An analysis of the applicants of all MBA programs in North America reveals that the proportions of each type of undergraduate degree are as follows:

Undergraduate Degree	Proportion (%)
B.A. (1)	50
B.B.A. (2)	20
B.Sc. (3)	15
B.Eng. (4)	10
Other (5)	5

The director of Wilfrid Laurier University's (WLU's) MBA program recorded the undergraduate degree of the applicants for this year using the codes in parentheses. Do these data indicate that applicants to WLU's MBA program are different in terms of their undergraduate degrees from the population of MBA applicants?

A15.2 XrA15-02 The experiment to determine the effect of taking a preparatory course to improve SAT scores in Exercise A13.16 was criticized by other statisticians. They argued that the first test would provide a valuable learning experience that would produce a higher test score from the second exam even without the preparatory course. Consequently, another experiment was performed. Forty students wrote the SAT without taking any preparatory course. At the next scheduled exam (3 months later), these same students took the exam again (again with no preparatory course). The scores for both exams were recorded in columns 1 (first test scores) and 2 (second test scores). Can we infer that repeating the SAT produces higher exam scores even without the preparatory course?

A15.3 XrA15-03 How does dieting affect the brain? This question was addressed by researchers in Australia. The experiment used 40 middle-age women in Adelaide, Australia; half were on a diet and half were not (*National Post*, December 1, 2003). The mental arithmetic part of the experiment required the participants to add two three-digit numbers. The amount of time taken to solve the 48 problems was recorded. The participants were given another test that required them to repeat a string of five letters they had been told 10 seconds earlier. They were asked to repeat the test with five words told to them 10 seconds earlier. The data were recorded in the following way:

Column 1: Identification number
Column 2: 1 = dieting, 2 = not dieting
Column 3: Time to solve 48 problems (seconds)
Column 4: Repeat string of 5 letters (1 = no, 2 = yes)
Column 5: Repeat string of 5 words (1 = no, 2 = yes)

Is there sufficient evidence to infer that dieting adversely affects the brain?

A15.4 XrA15-04 A small but important part of a university library's budget is the amount collected in fines on overdue books. Last year, a library collected $75,652.75 in fine payments; however, the head librarian suspects that some employees are not bothering to collect the fines on overdue books. In an effort to learn more about the situation, she asked a sample of 400 students (out of a total student population of 50,000) how many books they had returned late to the library in the previous 12 months. They were also asked how many days overdue the books had been. The results indicated that the total number of days overdue ranged from 0 to 55 days. The number of days overdue was recorded.

a. Estimate with 95% confidence the average number of days overdue for all 50,000 students at the university.
b. If the fine is 25 cents per day, estimate the amount that should be collected annually. Should the librarian conclude that not all the fines were collected?

A15.5 XrA15-05 An apple juice manufacturer has developed a new product—a liquid concentrate that produces 1 liter of apple juice when mixed with water. The product has several attractive features. First, it is more convenient than bottled apple juice, which is the way apple juice is currently sold. Second, because the apple juice that is sold in cans is actually made from concentrate, the quality of the new product is at least as high as that of bottled apple juice. Third, the cost of the new product is slightly lower than that of bottled apple juice. The marketing manager has to decide how to market the new product. She can create advertising that emphasizes convenience, quality, or price. To facilitate a decision, she conducts an experiment in three different small cities. In one city, she launches the product with advertising stressing the convenience of the liquid concentrate (e.g., easy to carry from store to home and takes up less room in the freezer). In the second city, the advertisements emphasize the quality of the product ("average" shoppers are depicted discussing how good the apple juice tastes). Advertising that highlights the relatively low cost of the liquid concentrate is used in the third city. The number of packages sold weekly is recorded for the 20 weeks following the beginning of the campaign. The marketing manager wants to know whether differences in sales exist between the three advertising strategies. (We will assume that except for the type of advertising, the three cities are identical.)

A15.6 XrA15-06 Mutual funds are a popular way of investing in the stock market. A financial analyst wanted to determine the effect income had on ownership of mutual funds and whether the relationship had changed from four years earlier. She took a random sample of adults 25 years of age and older and asked each person whether he or she owned mutual funds (No = 1 and Yes = 2) and to report the annual household income. The categories are

1. Less than $25,000
2. $25,000 to $34,999
3. $35,000 to $49,999
4. $50,000 to $74,999
5. $75,000 to $100,000
6. More than $100,000

Can we infer from the data that household income and ownership of mutual funds are related? (Adapted from the *Statistical Abstract of the United States, 2006*, Table 1200.)

A15.7 XrA15-07 Refer to Exercise A15.5. Suppose that in addition to varying the marketing strategy, the manufacturer also decided to advertise in one of the two media that are available: television and newspapers. As a consequence, the experiment was repeated in the following way. Six different small cities were selected. In city 1, the marketing emphasized convenience, and all the advertising was conducted on television. In city 2, marketing also emphasized convenience, but all the advertising was conducted in the daily newspaper. Quality was emphasized in cities 3 and 4. City 3 learned about the product from television commercials, and city 4 saw newspaper advertising. Price was the marketing emphasis in cities 5 and 6. City 5 saw television commercials, and city 6 saw newspaper advertisements. In each city, the weekly sales for each of 10 weeks were recorded. What conclusions can be drawn from these data?

A15.8 XrA15-08 After a recent study, researchers reported on the effects of folic acid on the occurrence of spina bifida—a birth defect in which there is incomplete formation of the spine. A sample of 2,000 women who gave birth to children with spina bifida and who were planning another pregnancy was recruited. Before attempting to get pregnant again, half the sample was given regular doses of folic acid, and the other half was given a placebo. After 18 months, researchers recorded the result for each woman: 1 = birth to normal baby, 2 = birth to baby with spina bifida, 3 = not pregnant or no baby yet delivered. Can we infer that folic acid reduces the incidence of spina bifida in newborn babies?

A15.9 XrA15-09 Slow play of golfers is a serious problem for golf clubs. Slow play results in fewer rounds of golf and less profits for public course owners. To examine this problem, a random sample of British and American golf courses was selected. The amount of time taken (in minutes) was recorded for a random sample of British and American golfers. Can we conclude that British golfers play golf in less time than do American golfers? (Source: *Golf Magazine*, July 2001.)

A15.10 XrA15-10 The United States and Canada (among others) are countries in which a significant proportion of citizens are immigrants. Many arrive in North America with few assets but quickly adapt to a changed economic environment. The question often arises, How quickly do immigrants increase their standard of living? A study initiated by Statistics Canada surveyed three different types of families:

1. Immigrants who arrived before 1976
2. Immigrants who came to Canada after 1986
3. Canadian-born families

The survey measured family wealth, which includes houses, cars, income, and savings and recorded the results (in $1,000s). Can we infer that differences exist between the three groups? If so, what are those differences?

A15.11 XrA15-11 During the decade of the 1980s, professional baseball thrived in North America. However, in the 1990s attendance dropped, and the number of television viewers also decreased. To examine the popularity of baseball relative to other sports, surveys were performed. In 1985 and again in 1992, a Harris Poll asked a random sample of 500 people to name their favorite sport. The results, which were published in the *Wall Street Journal* (July 6, 1993), were recorded in the following way: favorite sport (1 = professional football, 2 = baseball, 3 = professional basketball, 4 = college basketball, 5 = college football, 6 = golf, 7 = auto racing, 8 = tennis, and 9 = other); year (1 = 1985, 2 = 1992). Do these results indicate that North Americans changed their favorite sport between 1985 and 1992?

A15.12 XrA15-12 In an attempt to learn more about traffic congestion in a large North American city, the number of cars passing through intersections was determined (*National Post*, October 18, 2006). The number of cars was counted in 5-minute samples throughout several days. The counts for one busy intersection were recorded. Estimate with 95% confidence the mean number of cars in 5 minutes. Use the result to estimate the counts for a 24-hour day.

A15.13 XrA15-13 Organizations that sponsor various leisure activities need to know the number of people who wish to participate. Bureaucrats need to know the number because many organizations apply for government grants to pay the costs. The U.S. National Endowment for the Arts conducts surveys of American adults to acquire this type of information. One part of the survey asked a random sample of adults whether they participated in exercise programs. The responses (1 = yes and 2 = no) were recorded. A recent census reveals that there are 205.9 million adults in the United States. Estimate with 95% confidence the number of American adults who participate in exercise programs. (Adapted from the *Statistical Abstract of the United States, 2006*, Table 1227.)

A15.14 XrA15-14 Low back pain is a common medical problem that sometimes results in disability and absence from work. Any method of treatment that decreases absence would be welcome by individuals and insurance companies. A randomized control study (published in *Annals of Internal Medicine*, January 2004) was undertaken to determine whether an alternate form of treatment is effective. The study examined 134 workers who were absent from work because of low back pain. Half the sample was assigned to graded activity, a physical exercise program designed to stimulate rapid return to work. The other half was assigned to the usual care, which involves mostly rest. For each worker, the number of days absent from work because of low back pain in the following 6 months was recorded. Do these data provide sufficient evidence to infer that the graded activity is effective?

A15.15 XrA15-15 Clinical depression is a serious and sometimes debilitating disease. It is often treated by antidepressants such as Prozac and Zoloft. Recent studies may indicate another possible remedy. Researchers took a random sample of people who are clinically depressed and divided them into three groups. The first group was treated with antidepressants and light therapy, the second was treated with a placebo and light therapy, and the third group treated with a placebo. Whether the patient showed improvement (code = 1) or not (code = 2) and the group number were recorded. Can we infer that there are differences between the three groups?

A15.16 How well do airlines keep to their schedules? To help answer this question, an economist conducted a survey of 780 takeoffs in the United States and determined that 77.4% of them departed on time (defined as a departure that is within 15 minutes of its scheduled time). There were 7,140,596 flight departures in the United States in 2005. Estimate with 95% confidence the total number of on-time departures.

General Social Survey Exercises

A15.17 GSS2006 GSS2008* During 2008, the United States was in the throes of a deep recession. The unemployment rate rose sharply. How did this affect job tenure (the amount of time a worker has been with his or her current employer)? Is there sufficient evidence to conclude that job tenure changed between 2006 (YEARSJOB) and 2008 (CUREMPYR)?

A15.18 GSS2008* Capital punishment for murderers exists in most U.S. states. However, a few states ban this form of punishment. Politicians often need to know which members of the public support and which oppose. Can we conclude from the data that there is a difference between Democrats, Republicans, and Independents (PARTYID: 0, 1 = Democrat, 2, 3, 4 = Independent, 5, 6 = Republican) in terms of support for capital punishment (CAPPUN)?

A15.19 GSS2008* Are married couples postponing bearing children? One way to measure this is to determine how old people are when their first child is born. Estimate with 95% confidence the average age of Americans when their first child is born (AGEKDBRN).

A15.20 GSS2008* In Chapter 2, we used a graphical technique and data from the American National Election

Survey to attempt to determine whether men and women differ in their political affiliation. Use a suitable statistical inference technique to determine whether there is sufficient evidence to infer that men and women (SEX) differ in their political affiliations (PARTYID).

A15.21 Do teenage and adult children living with their parents contribute to household income by holding down full- or part-time jobs? And is it more likely that they do so for affluent than for less affluent families? To answer the question, test to determine whether the data allow us to conclude that there are differences in the number of family members earning money (EARNRS) between the four classes (CLASS).

A15.22 GSS2002* GSS2004* GSS2006* A generally accepted method of finding whether Americans have improved financially over a multiyear period is to calculate inflation-adjusted incomes. Using the General Social Survey data, can we infer that American inflation-adjusted incomes (CONRINC) varied from year to year in 2002, 2004, and 2006 ?

A15.23 GSS2008* Does the race of an individual affect whether he or she is likely to be self-employed? Can we conclude that differences in whether an individual works for him- or herself (WRKSLF: 1 = Self-employed, 2 = Someone else) exists between the races (RACE)?

A15.24 GSS2008* Does being unemployed for any period of time affect an individual's political persuasion? Using the GSS 2008 data, determine whether there is enough evidence to infer that Americans who have been unemployed in the last 10 years (UNEMP) have different party affiliations (PARTYID: 0, 1 = Democrat, 2, 3, 4 = Independent, 5, 6 = Republican) than those who have not been unemployed.

American National Election Survey Exercises

A15.25 ANES2008* In recent years, the proportion of eligible voters in the United States who actually vote for president has hovered around 50%. Turning out the vote is considered a critical function for most political voters. Are there differences between Liberals and Conservatives in their intention to vote? Conduct a test to determine whether there is sufficient evidence to infer that liberals and conservatives (LIBCON: 1, 2, 3 = liberal, 5, 6, 7 = conservative) differ in their intention to vote (DEFINITE).

A15.26 ANES2004* ANES2008* The economy in 2004 was strong, with growth in the economy and unemployment low. By 2008, the U.S. economy was in recession. Can we conclude that employment status (EMPLOY) has changed between 2004 and 2008?

A15.27 ANES2008* Do the data provide sufficient evidence to conclude that Americans who consider themselves strong Democrat or Republicans (STRENGTH: 1 = Strong, 5 = Not very strong) have more education (EDUC) than those who do not?

CASE A15.1 Which Diets Work?

DATA CA15-01

Every year, millions of people start new diets. There is a bewildering array of diets to choose from. The question for many people is, which ones work? Researchers at Tufts University in Boston made an attempt to point dieters in the right direction. Four diets were used:

1. Atkins low-carbohydrate diet
2. Zone high-protein, moderate-carbohydrate diet
3. Weight Watchers diet
4. Dr. Ornish's low-fat diet

The study recruited 160 overweight people and randomly assigned 40 to each diet. The average weight before dieting was 220 pounds, and all needed to lose between 30 and 80 pounds. All volunteers agreed to follow their diets for 2 months. No exercise or regular meetings were required. The following variables were recorded for each dieter using the format shown here:

Column 1: Identification number
Column 2: Diet
Column 3: Percent weight loss
Column 4: Percent low-density lipoprotein (LDL)—"bad" cholesterol—decrease
Column 5: Percent high-density lipoprotein (HDL)—"good" cholesterol—increase
Column 6: Quit after 2 months? 1 = yes, 2 = no
Column 7: Quit after 1 year? 1 = yes, 2 = no

Is there enough evidence to conclude that there are differences between the diets with respect to

a. percent weight loss?
b. percent LDL decrease?
c. percent HDL increase?
d. proportion quitting within 2 months?
e. proportion quitting after 1 year?

© AFP/Getty Images

16

SIMPLE LINEAR REGRESSION AND CORRELATION

16.1 Model

16.2 Estimating the Coefficients

16.3 Error Variable: Required Conditions

16.4 Assessing the Model

16.5 Using the Regression Equation

16.6 Regression Diagnostics—I

Appendix 16 Review of Chapters 12 to 16

Education and Income: How Are They Related?

DATA
GSS2008*

If you're taking this course, you're probably a student in an undergraduate or graduate business or economics program. Your plan is to graduate, get a good job, and draw a high salary. You have probably assumed that more education equals better job equals higher income. Is this true? Fortunately, the General Social Survey recorded two variables that will help determine whether education and income are related and, if so, what the value of an additional year of education might be.

© Vicki Beaver

On page 663, we will provide our answer.

INTRODUCTION

Regression analysis is used to predict the value of one variable on the basis of other variables. This technique may be the most commonly used statistical procedure because, as you can easily appreciate, almost all companies and government institutions forecast variables such as product demand, interest rates, inflation rates, prices of raw materials, and labor costs.

The technique involves developing a mathematical equation or model that describes the relationship between the variable to be forecast, which is called the **dependent variable**, and variables that the statistics practitioner believes are related to the dependent variable. The dependent variable is denoted Y, whereas the related variables are called **independent variables** and are denoted $X_1, X_2, \ldots, X_k$ (where k is the number of independent variables).

If we are interested only in determining whether a relationship exists, we employ correlation analysis, a technique that we have already introduced. In Chapter 3, we presented the graphical method to describe the association between two interval variables—the scatter diagram. We introduced the coefficient of correlation and covariance in Chapter 4.

Because regression analysis involves many new techniques and concepts, we divided the presentation into three chapters. In this chapter, we present techniques that allow us to determine the relationship between only two variables. In Chapter 17, we expand our discussion to more than two variables; in Chapter 18, we discuss how to build regression models.

Here are three illustrations of the use of regression analysis.

Illustration 1 The product manager in charge of a particular brand of children's breakfast cereal would like to predict the demand for the cereal during the next year. To use regression analysis, she and her staff list the following variables as likely to affect sales:

Price of the product
Number of children 5 to 12 years of age (the target market)
Price of competitors' products
Effectiveness of advertising (as measured by advertising exposure)
Annual sales this year
Annual sales in previous years

Illustration 2 A gold speculator is considering a major purchase of gold bullion. He would like to forecast the price of gold 2 years from now (his planning horizon), using regression analysis. In preparation, he produces the following list of independent variables:

Interest rates
Inflation rate
Price of oil
Demand for gold jewelry
Demand for industrial and commercial gold
Dow Jones Industrial Average

Illustration 3 A real estate agent wants to predict the selling price of houses more accurately. She believes that the following variables affect the price of a house:

Size of the house (number of square feet)
Number of bedrooms

Frontage of the lot
Condition
Location

In each of these illustrations, the primary motive for using regression analysis is forecasting. Nonetheless, analyzing the relationship among variables can also be quite useful in managerial decision making. For instance, in the first application, the product manager may want to know how price is related to product demand so that a decision about a prospective change in pricing can be made.

Regardless of why regression analysis is performed, the next step in the technique is to develop a mathematical equation or model that accurately describes the nature of the relationship that exists between the dependent variable and the independent variables. This stage—which is only a small part of the total process—is described in the next section. In the ensuing sections of this chapter (and in Chapter 17), we will spend considerable time assessing and testing how well the model fits the actual data. Only when we're satisfied with the model do we use it to estimate and forecast.

16.1 MODEL

The job of developing a mathematical equation can be quite complex, because we need to have some idea about the nature of the relationship between each of the independent variables and the dependent variable. The number of different mathematical models that could be proposed is virtually infinite. Here is an example from Chapter 4.

$$\begin{aligned}\text{Profit} = &\ (\text{Price per unit} - \text{variable cost per unit})\\ &\times \text{Number of units sold} - \text{Fixed costs}\end{aligned}$$

You may encounter the next example in a finance course:

$$F = P(1 + i)^n$$

where

F = Future value of an investment
P = principle or present value
i = interest rate per period
n = number of periods

These are all examples of **deterministic models**, so named because such equations allow us to determine the value of the dependent variable (on the left side of the equation) from the values of the independent variables. In many practical applications of interest to us, deterministic models are unrealistic. For example, is it reasonable to believe that we can determine the selling price of a house solely on the basis of its size? Unquestionably, the size of a house affects its price, but many other variables (some of which may not be measurable) also influence price. What must be included in most practical models is a method to represent the randomness that is part of a real-life process. Such a model is called a **probabilistic model**.

To create a probabilistic model, we start with a deterministic model that approximates the relationship we want to model. We then add a term that measures the random error of the deterministic component.

Suppose that in illustration 3, the real estate agent knows that the cost of building a new house is about \$100 per square foot and that most lots sell for about \$100,000. The approximate selling price would be

$$y = 100{,}000 + 100x$$

where y = selling price and x = size of the house in square feet. A house of 2,000 square feet would therefore be estimated to sell for

$$y = 100{,}000 + 100(2{,}000) = 300{,}000$$

We know, however, that the selling price is not likely to be exactly \$300,000. Prices may actually range from \$200,000 to \$400,000. In other words, the deterministic model is not really suitable. To represent this situation properly, we should use the probabilistic model

$$y = 100{,}000 + 100x + \varepsilon$$

where ε (the Greek letter epsilon) represents the **error variable**—the difference between the actual selling price and the estimated price based on the size of the house. The error thus accounts for all the variables, measurable and immeasurable, that are not part of the model. The value of ε will vary from one sale to the next, even if x remains constant. In other words, houses of exactly the same size will sell for different prices because of differences in location and number of bedrooms and bathrooms, as well as other variables.

In the three chapters devoted to regression analysis, we will present only probabilistic models. In this chapter, we describe only the straight-line model with one independent variable. This model is called the **first-order linear model**—sometimes called the **simple linear regression model**.*

First-Order Linear Model

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where

y = dependent variable

x = independent variable

β_0 = y-intercept

β_1 = slope of the line (defined as rise/run)

ε = error variable

The problem objective addressed by the model is to analyze the relationship between two variables, x and y, both of which must be interval. To define the relationship between x and y, we need to know the value of the coefficients β_0 and β_1. However, these coefficients are population parameters, which are almost always unknown. In the next section, we discuss how these parameters are estimated.

*We use the term *linear* in two ways. The "linear" in linear regression refers to the form of the model wherein the terms form a linear combination of the coefficients β_0 and β_1. Thus, for example, the model $y = \beta_0 + \beta_1 x^2 + \varepsilon$ is a linear combination whereas $y = \beta_0 + \beta_1^2 x + \varepsilon$ is not. The simple linear regression model $y = \beta_0 + \beta_1 x + \varepsilon$ describes a straight-line or linear relationship between the dependent variable and one independent variable. In this book, we use the linear regression technique only. Hence, when we use the word *linear* we will be referring to the straight-line relationship between the variables.

16.2 ESTIMATING THE COEFFICIENTS

We estimate the parameters β_0 and β_1 in a way similar to the methods used to estimate all the other parameters discussed in this book. We draw a random sample from the population of interest and calculate the sample statistics we need. However, because β_0 and β_1 represent the coefficients of a straight line, their estimators are based on drawing a straight line through the sample data. The straight line that we wish to use to estimate β_0 and β_1 is the "best" straight line—best in the sense that it comes closest to the sample data points. This best straight line, called the *least squares line*, is derived from calculus and is represented by the following equation:

$$\hat{y} = b_0 + b_1 x$$

Here b_0 is the y-intercept, b_1 is the slope, and $\hat{y}$ is the predicted or fitted value of y. In Chapter 4, we introduced the **least squares method**, which produces a straight line that minimizes the sum of the squared differences between the points and the line. The coefficients b_0 and b_1 are calculated so that the sum of squared deviations

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

is minimized. In other words, the values of $\hat{y}$ on average come closest to the observed values of y. The calculus derivation is available in Keller's website appendix, Deriving the Normal Equations, which shows how the following formulas, first shown in Chapter 4, were produced.

Least Squares Line Coefficients

$$b_1 = \frac{s_{xy}}{s_x^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

where

$$s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n - 1}$$

$$s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

In Chapter 4, we provided shortcut formulas for the sample variance (page 110) and the sample covariance (page 127). Combining them provides a shortcut method to manually calculate the slope coefficient.

Shortcut Formula for b_1

$$b_1 = \frac{s_{xy}}{s_x^2}$$

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right]$$

$$s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$$

Statisticians have shown that b_0 and b_1 are unbiased estimators of β_0 and β_1, respectively.

Although the calculations are straightforward, we would rarely compute the regression line manually because the work is time consuming. However, we illustrate the manual calculations for a very small sample.

EXAMPLE 16.1 Annual Bonus and Years of Experience

DATA Xm16-01

The annual bonuses ($1,000s) of six employees with different years of experience were recorded as follows. We wish to determine the straight-line relationship between annual bonus and years of experience.

Years of experience *x*	1	2	3	4	5	6
Annual bonus *y*	6	1	9	5	17	12

SOLUTION

To apply the shortcut formula, we need to compute four summations. Using a calculator, we find

$$\sum_{i=1}^{n} x_i = 21$$

$$\sum_{i=1}^{n} y_i = 50$$

$$\sum_{i=1}^{n} x_i y_i = 212$$

$$\sum_{i=1}^{n} x_i^2 = 91$$

The covariance and the variance of x can now be computed:

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right] = \frac{1}{6-1}\left[212 - \frac{(21)(50)}{6}\right] = 7.4$$

$$s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right] = \frac{1}{6-1}\left[91 - \frac{(21)^2}{6}\right] = 3.5$$

The sample slope coefficient is calculated next:

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{7.4}{3.5} = 2.114$$

The y-intercept is computed as follows:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{21}{6} = 3.5$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{50}{6} = 8.333$$

$$b_0 = \bar{y} - b_1\bar{x} = 8.333 - (2.114)(3.5) = .934$$

Thus, the least squares line is

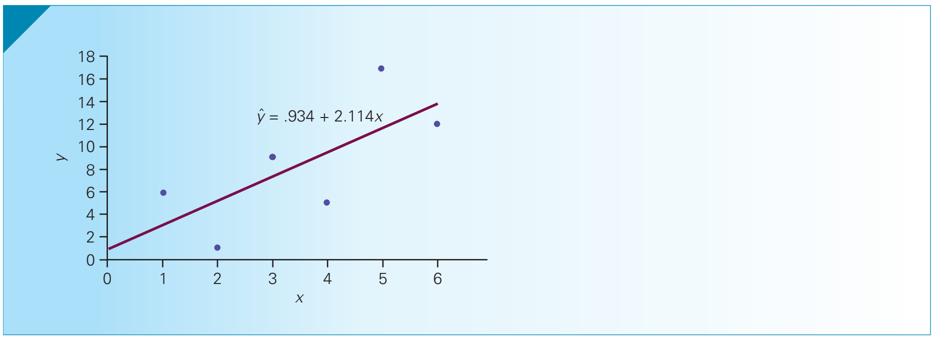
$$\hat{y} = .934 + 2.114x$$

Figure 16.1 depicts the least squares (or regression) line. As you can see, the line fits the data reasonably well. We can measure how well by computing the value of the minimized sum of squared deviations. The deviations between the actual data points and the line are called **residuals**, denoted e_i; that is,

$$e_i = y_i - \hat{y}_i$$

The residuals are observations of the error variable. Consequently, the minimized sum of squared deviations is called the **sum of squares for error**, denoted SSE.

FIGURE **16.1** Scatter Diagram with Regression Line for Example 16.1



The calculation of the residuals in this example is shown in Figure 16.2. Notice that we compute $\hat{y}_i$ by substituting x_i into the formula of the regression line. The residuals are the differences between the observed values of y_i and the fitted or predicted values of $\hat{y}_i$. Table 16.1 describes these calculations.

Thus, SSE = 81.104. No other straight line will produce a sum of squared deviations as small as 81.104. In that sense, the regression line fits the data best. The sum of squares for error is an important statistic because it is the basis for other statistics that assess how well the linear model fits the data. We will introduce these statistics in Section 16.4.

FIGURE **16.2** Calculation of Residuals in Example 16.1

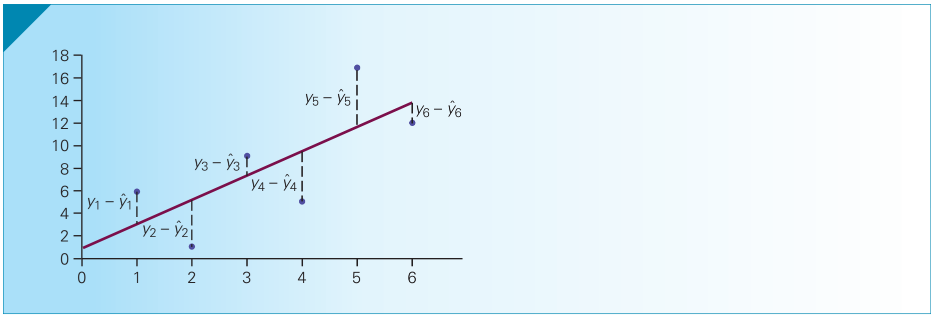


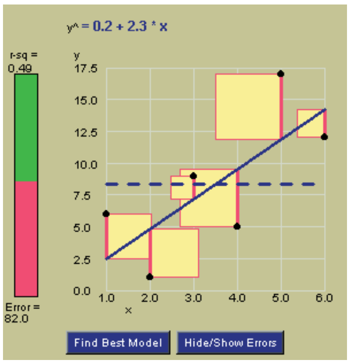
TABLE **16.1** Calculation of Residuals in Example 16.1

x_i	y_i	$\hat{y}_i = .934 + 2.114x_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	6	3.048	2.952	8.714
2	1	5.162	−4.162	17.322
3	9	7.276	1.724	2.972
4	5	9.390	−4.390	19.272
5	17	11.504	5.496	30.206
6	12	13.618	−1.618	2.618
			$\sum(y_i - \hat{y}_i)^2 =$	81.104

SEEING STATISTICS

applet 18 Fitting the Regression Line

This applet allows you to experiment with the data in Example 16.1. Click or drag the mouse in the graph to change the slope of the line. The errors are measured by the red lines. The squares represent the squared errors. (You can hide or show them by clicking on the **Hide/Show Errors button**.) The error meter on the left keeps track of your progress. The amount of the error that turns green is the proportion of the squared error you eliminate by finding a better regression line. The sum of squared errors is shown at the bottom. The coefficient of correlation squared (which is the coefficient of determination, explained in Section 16.4) is shown at the top. Change the slope until the sum of squares for error as indicated in the error meter is minimized. If you need help, click the **Find Best Model** button.



Applet Exercises

Change the slope (if necessary) so that the line is horizontal.

17.1 What is the slope of this line?

17.2 What is the y-intercept?

17.3 The y-intercept is equal to $\bar{y}$. What does this tell you about predicting the value of y?

17.4 Drag the mouse to change the slope to 1. What is the sum of squared errors?

17.5 Drag the mouse to change the slope to .5. What is the sum of squared errors?

17.6 Experiment with different lines. What point is common to all the lines?

EXAMPLE 16.2 Odometer Reading and Prices of Used Toyota Camrys, Part 1

DATA
Xm16-02*

Car dealers across North America use the so-called Blue Book to help them determine the value of used cars that their customers trade in when purchasing new cars. The book, which is published monthly, lists the trade-in values for all basic models of cars. It provides alternative values for each car model according to its condition and optional features. The values are determined on the basis of the average paid at recent used-car auctions, the source of supply for many used-car dealers. However, the Blue Book does not indicate the value determined by the odometer reading, despite the fact that a critical factor for used-car buyers is how far the car has been driven. To examine this issue, a used-car dealer randomly selected 100 3-year old Toyota Camrys that were sold at auction during the past month. Each car was in top condition and equipped with all the features that come standard with this car. The dealer recorded the price ($1,000) and the number of miles (thousands) on the odometer. Some of these data are listed here. The dealer wants to find the regression line.

Car	Price ($1,000)	Odometer (1,000 mi)
1	14.6	37.4
2	14.1	44.8
3	14.0	45.8
⋮	⋮	⋮
98	14.5	33.2
99	14.7	39.2
100	14.3	36.4

SOLUTION

IDENTIFY

Notice that the problem objective is to analyze the relationship between two interval variables. Because we believe that the odometer reading affects the selling price, we identify the former as the independent variable, which we label x, and the latter as the dependent variable, which we label y.

COMPUTE

MANUALLY

From the data set, we find

$$\sum_{i=1}^{n} x_i = 3{,}601.1$$

$$\sum_{i=1}^{n} y_i = 1{,}484.1$$

$$\sum_{i=1}^{n} x_i y_i = 53{,}155.93$$

$$\sum_{i=1}^{n} x_i^2 = 133{,}986.59$$

Next we calculate the covariance and the variance of the independent variable x:

$$s_{xy} = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}\right]$$

$$= \frac{1}{100-1}\left[53{,}155.93 - \frac{(3{,}601.1)(1{,}484.1)}{100}\right] = -2.909$$

$$s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right]$$

$$= \frac{1}{100-1}\left[133{,}986.59 - \frac{(3{,}601.1)^2}{100}\right] = 43.509$$

The sample slope coefficient is calculated next:

$$b_1 = \frac{s_{xy}}{s_x^2} = \frac{-2.909}{43.509} = -.0669$$

The y-intercept is computed as follows:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{3{,}601.1}{100} = 36.011$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{1{,}484.1}{100} = 14.841$$

$$b_0 = \bar{y} - b_1\bar{x} = 14.841 - (-.0669)(36.011) = 17.250$$

The sample regression line is

$$\hat{y} = 17.250 - 0.0669x$$

EXCEL

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.8052				
5	R Square	0.6483				
6	Adjusted R Square	0.6447				
7	Standard Error	0.3265				
8	Observations	100				
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression	1	19.26	19.26	180.64	5.75E-24
13	Residual	98	10.45	0.11		
14	Total	99	29.70			
15						
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
17	Intercept	17.25	0.182	94.73	3.57E-98	
18	Odometer	–0.0669	0.0050	–13.44	5.75E-24	

INSTRUCTIONS

1. Type or import data into two columns*, one storing the dependent variable and the other the independent variable. (Open Xm16-02.)
2. Click **Data**, **Data Analysis,** and **Regression**.
3. Specify the **Input *Y* Range** (A1:A101) and the **Input *X* Range** (B1:B101).

To draw the scatter diagram follow the instructions provided in Chapter 3 on page 76.

MINITAB

Regression Analysis: Price versus Odometer

The regression equation is
Price = 17.2 - 0.0669 Odometer

Predictor	Coef	SE Coef	T	P
Constant	17.2487	0.1821	94.73	0.000
Odometer	–0.066861	0.004975	–13.44	0.000

S = 0.326489   R-Sq = 64.8%   R-Sq(adj) = 64.5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	19.256	19.256	180.64	0.000
Residual Error	98	10.446	0.107		
Total	99	29.702			

INSTRUCTIONS

1. Type or import the data into two columns. (Open Xm16-02.)
2. Click **Stat**, **Regression**, and **Regression . . . .**
3. Type the name of the dependent variable in the **Response** box (Price) and the name of the independent variable in the **Predictors** box (Odometer).

To draw the scatter diagram click **Stat**, **Regression,** and **Fitted Line Plot**. Alternatively, follow the instructions provide in Chapter 3.

The printouts include more statistics than we need right now. However, we will be discussing the rest of the printouts later.

INTERPRET

The slope coefficient b_1 is -0.0669, which means that for each additional 1,000 miles on the odometer, the price decreases by an average of \$.0669 thousand. Expressed more simply, the slope tells us that for each additional mile on the odometer, the price decreases on average by \$.0669 or 6.69 cents.

The intercept is $b_0 = 17.250$. Technically, the intercept is the point at which the regression line and the y-axis intersect. This means that when $x = 0$ (i.e., the car was not driven at all) the selling price is \$17.250 thousand or \$17,250. We might be tempted to

*If one or both columns contain a blank (representing missing data) the row must be deleted.

interpret this number as the price of cars that have not been driven. However, in this case, the intercept is probably meaningless. Because our sample did not include any cars with zero miles on the odometer, we have no basis for interpreting b_0. As a general rule, we cannot determine the value of $\hat{y}$ for a value of x that is far outside the range of the sample values of x. In this example, the smallest and largest values of x are 19.1 and 49.2, respectively. Because $x = 0$ is not in this interval, we cannot safely interpret the value of $\hat{y}$ when $x = 0$.

It is important to bear in mind that the interpretation of the coefficients pertains only to the sample, which consists of 100 observations. To infer information about the population, we need statistical inference techniques, which are described subsequently.

In the sections that follow, we will return to this problem and the computer output to introduce other statistics associated with regression analysis.

EXERCISES

16.1 The term *regression* was originally used in 1885 by Sir Francis Galton in his analysis of the relationship between the heights of children and parents. He formulated the "law of universal regression," which specifies that "each peculiarity in a man is shared by his kinsmen, but on average in a less degree." (Evidently, people spoke this way in 1885.) In 1903, two statisticians, K. Pearson and A. Lee, took a random sample of 1,078 father–son pairs to examine Galton's law ("On the Laws of Inheritance in Man, I. Inheritance of Physical Characteristics," *Biometrika* 2:457–462). Their sample regression line was

$$\text{Son's height} = 33.73 + .516 \times \text{Father's height}$$

a. Interpret the coefficients.
b. What does the regression line tell you about the heights of sons of tall fathers?
c. What does the regression line tell you about the heights of sons of short fathers?

16.2 Xr16-02 Attempting to analyze the relationship between advertising and sales, the owner of a furniture store recorded the monthly advertising budget ($ thousands) and the sales ($ millions) for a sample of 12 months. The data are listed here.

Advertising	23	46	60	54	28	33
Sales	9.6	11.3	12.8	9.8	8.9	12.5

Advertising	25	31	36	88	90	99
Sales	12.0	11.4	12.6	13.7	14.4	15.9

a. Draw a scatter diagram. Does it appear that advertising and sales are linearly related?
b. Calculate the least squares line and interpret the coefficients.

16.3 Xr16-03 To determine how the number of housing starts is affected by mortgage rates an economist recorded the average mortgage rate and the number of housing starts in a large county for the past 10 years. These data are listed here.

Rate	8.5	7.8	7.6	7.5	8.0
Starts	115	111	185	201	206

Rate	8.4	8.8	8.9	8.5	8.0
Starts	167	155	117	133	150

a. Determine the regression line.
b. What do the coefficients of the regression line tell you about the relationship between mortgage rates and housing starts?

16.4 Xr16-04 Critics of television often refer to the detrimental effects that all the violence shown on television has on children. However, there may be another problem. It may be that watching television also reduces the amount of physical exercise, causing weight gains. A sample of 15 10-year-old children was taken. The number of pounds each child was overweight was recorded (a negative number indicates the child is underweight). In addition, the number of hours of television viewing per week was also recorded. These data are listed here.

Television	42	34	25	35	37	38	31	33
Overweight	18	6	0	−1	13	14	7	7

Television	19	29	38	28	29	36	18
Overweight	−9	8	8	5	3	14	−7

a. Draw the scatter diagram.
b. Calculate the sample regression line and describe what the coefficients tell you about the relationship between the two variables.

16.5 Xr16-05 To help determine how many beers to stock the concession manager at Yankee Stadium wanted to know how the temperature affected beer sales. Accordingly, she took a sample of 10 games and recorded the number of beers sold and the temperature in the middle of the game.

Temperature	80	68	78	79	87
Number of beers	20,533	1,439	13,829	21,286	30,985

Temperature	74	86	92	77	84
Number of beers	17,187	30,240	37,596	9,610	28,742

a. Compute the coefficients of the regression line.
b. Interpret the coefficients.

The exercises that follow were created to allow you to see how regression analysis is used to solve realistic problems. As a result, most feature a large number of observations. We anticipate that most students will solve these problems using a computer and statistical software. However, for students without these resources, we have computed the means, variances, and covariances that will permit them to complete the calculations manually. (See Appendix A.)

16.6 Xr16-06* In television's early years, most commercials were 60 seconds long. Now, however, commercials can be any length. The objective of commercials remains the same—to have as many viewers as possible remember the product in a favorable way and eventually buy it. In an experiment to determine how the length of a commercial is related to people's memory of it, 60 randomly selected people were asked to watch a 1-hour television program. In the middle of the show, a commercial advertising a brand of toothpaste appeared. Some viewers watched a commercial that lasted for 20 seconds, others watched one that lasted for 24 seconds, 28 seconds, . . . , 60 seconds. The essential content of the commercials was the same. After the show, each person was given a test to measure how much he or she remembered about the product. The commercial times and test scores (on a 30-point test) were recorded.

a. Draw a scatter diagram of the data to determine whether a linear model appears to be appropriate.
b. Determine the least squares line.
c. Interpret the coefficients.

16.7 Xr16-07 Florida condominiums are popular winter retreats for many North Americans. In recent years, the prices have steadily increased. A real estate agent wanted to know why prices of similar-sized apartments in the same building vary. A possible answer lies in the floor. It may be that the higher the floor, the greater the sale price of the apartment. He recorded the price (in $1,000s) of 1,200 sq. ft. condominiums in several buildings in the same location that have sold recently and the floor number of the condominium.

a. Determine the regression line.
b. What do the coefficients tell you about the relationship between the two variables?

16.8 Xr16-08 In 2010, the United States conducted a census of the entire country. The census is completed by mail. To help ensure that the questions are understood, a random sample of Americans take the questionnaire before it is sent out. As part of their analysis, they record the amount of time and ages of the sample. Use the least squares method to analyze the relationship between the amount of time taken to complete the questionnaire and the age of the individual answering the questions. What do the coefficients tell you about the relationship between the two variables?

APPLICATIONS in HUMAN RESOURCES MANAGEMENT

© Brand X Pictures/Getty Images

Retaining Workers

Human resource managers are responsible for a variety of tasks within organizations. As we pointed out in the introduction in Chapter 1, personnel or human resource managers are involved with recruiting new workers, determining which applicants are most suitable to hire, and helping with various aspects of monitoring the workforce, including absenteeism and worker turnover. For many firms, worker turnover is a costly problem. First, there is the cost of recruiting and attracting qualified workers. The firm must advertise vacant positions and make certain that applicants are judged properly. Second, the cost

(Continued)

of training hirees can be high, particularly in technical areas. Third, new employees are often not as productive and efficient as experienced employees. Consequently, it is in the interests of the firm to attract and keep the best workers. Any information that the personnel manager can obtain is likely to be useful.

16.9 Xr16-09 The human resource manager of a telemarketing firm is concerned about the rapid turnover of the firm's telemarketers. It appears that many telemarketers do not work very long before quitting. There may be a number of reasons, including relatively low pay, personal unsuitability for the work, and the low probability of advancement. Because of the high cost of hiring and training new workers, the manager decided to examine the factors that influence workers to quit. He reviewed the work history of a random sample of workers who have quit in the last year and recorded the number of weeks on the job before quitting and the age of each worker when originally hired.

a. Use regression analysis to describe how the work period and age are related.
b. Briefly discuss what the coefficients tell you.

16.10 Xr16-10 Besides their known long-term effects, do cigarettes also cause short-term illnesses such as colds? To help answer this question, a sample of smokers was drawn. Each person was asked to report the average number of cigarettes smoked per day and the number of days absent from work due to colds last year.

a. Determine the regression line.
b. What do the coefficients tell you about the relationship between smoking cigarettes and sick days because of colds?

16.11 Xr16-11 Fire damage in the United States amounts to billions of dollars, much of it insured. The time taken to arrive at the fire is critical. This raises the question, Should insurance companies lower premiums if the home to be insured is close to a fire station? To help make a decision, a study was undertaken wherein a number of fires were investigated. The distance to the nearest fire station (in miles) and the percentage of fire damage were recorded. Determine the least squares line and interpret the coefficients.

16.12 Xr16-12* A real estate agent specializing in commercial real estate wanted a more precise method of judging the likely selling price (in $1,000s) of apartment buildings. As a first effort, she recorded the price of a number of apartment buildings sold recently and the number of square feet (in 1,000s) in the building.

a. Calculate the regression line.
b. What do the coefficients tell you about the relationship between price and square footage?

16.13 Xr16-13 Millions of boats are registered in the United States. As is the case with automobiles, there is an active used-boat market. Many of the boats purchased require bank financing, and, as a result, it is important for financial institutions to be capable of accurately estimating the price of boats. One variable that affects the price is the number of hours the engine has been run. To determine the effect of the hours on the price, a financial analyst recorded the price (in $1,000s) of a sample of 2007 24-foot Sea Ray cruisers (one of the most popular boats) and the number of hours they had been run. Determine the least squares line and explain what the coefficients tell you.

16.14 Xr03-54 (Exercise 3.54 revisited) In an attempt to determine the factors that affect the amount of energy used, 200 households were analyzed. In each, the number of occupants and the amount of electricity used were measured. Determine the regression line and interpret the results.

16.15 Xr16-15 An economist for the federal government is attempting to produce a better measure of poverty than is currently in use. To help acquire information, she recorded the annual household income (in $1,000s) and the amount of money spent on food during one week for a random sample of households. Determine the regression line and interpret the coefficients.

16.16 Xr16-16* An economist wanted to investigate the relationship between office rents (the dependent variable) and vacancy rates. Accordingly, he took a random sample of monthly office rents and the percentage of vacant office space in 30 different cities.

a. Determine the regression line.
b. Interpret the coefficients.

16.17 Xr03-56 (Exercise 3.56 revisited) One general belief held by observers of the business world is that taller men earn more money than shorter men. In a University of Pittsburgh study, 250 MBA graduates, all about 30 years old, were polled and asked to report their height (in inches) and their annual income (to the nearest $1,000).

a. Determine the regression line.
b. What do the coefficients tell you?

APPLICATIONS in **HUMAN RESOURCES MANAGEMENT**


© Romilly Lockyer/The Image Bank/Getty Images

Testing Job Applicants

The recruitment process at many firms involves tests to determine the suitability of candidates. The tests may be written to determine whether the applicant has sufficient knowledge in his or her area of expertise to perform well on the job. There may be oral tests to determine whether the applicant's personality matches the needs of the job. Manual or technical skills can be tested through a variety of physical tests. The test results contribute to the decision to hire. In some cases, the test result is the only criterion to hire. Consequently, it is vital to ensure that the test is a reliable predictor of job performance. If the tests are poor predictors, they should be discontinued. Statistical analyses allow personnel managers to examine the link between the test results and job performance.

16.18 Xr16-18 Although a large number of tasks in the computer industry are robotic, many operations require human workers. Some jobs require a great deal of dexterity to properly position components into place. A large North American computer maker routinely tests applicants for these jobs by giving a dexterity test that involves a number of intricate finger and hand movements. The tests are scored on a 100-point scale. Only those who have scored above 70 are hired. To determine whether the tests are valid predictors of job performance, the personnel manager drew a random sample of 45 workers who were hired 2 months ago. He recorded their test scores and the percentage of nondefective computers they produced in the last week. Determine the regression line and interpret the coefficients.

16.3 ERROR VARIABLE: REQUIRED CONDITIONS

In the previous section, we used the least squares method to estimate the coefficients of the linear regression model. A critical part of this model is the error variable ε. In the next section, we will present an inferential method that determines whether there is a relationship between the dependent and independent variables. Later we will show how we use the regression equation to estimate and predict. For these methods to be valid, however, four requirements involving the probability distribution of the error variable must be satisfied.

Required Conditions for the Error Variable

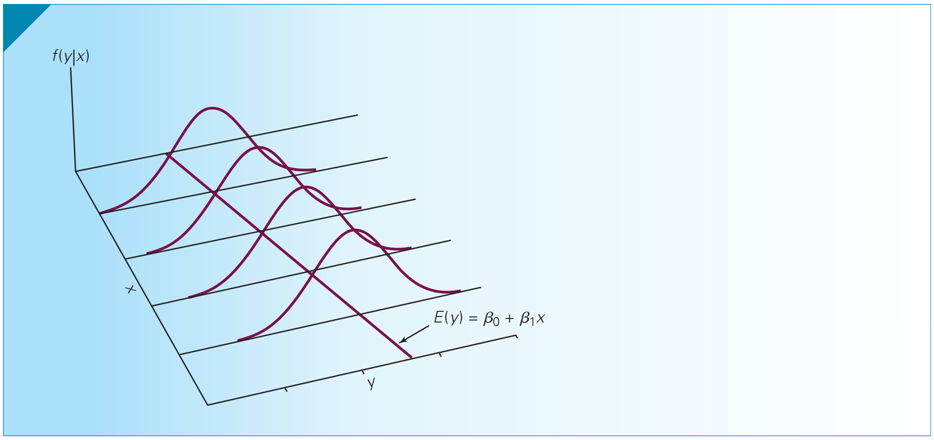
1. The probability distribution of ε is normal.
2. The mean of the distribution is 0; that is, $E(\varepsilon) = 0$.
3. The standard deviation of ε is σ_ε, which is a constant regardless of the value of x.
4. The value of ε associated with any particular value of y is independent of ε associated with any other value of y.

Requirements 1, 2, and 3 can be interpreted in another way: For each value of x, y is a normally distributed random variable whose mean is

$$E(y) = \beta_0 + \beta_1 x$$

and whose standard deviation is σ_ε. Notice that the mean depends on x. The standard deviation, however, is not influenced by x because it is a constant over all values of x. Figure 16.3 depicts this interpretation. Notice that for each value of x, $E(y)$ changes, but the shape of the distribution of y remains the same. In other words, for each x, y is normally distributed with the same standard deviation.

FIGURE **16.3** Distribution of y Given x



In Section 16.6, we will discuss how departures from these required conditions affect the regression analysis and how they are identified.

Observational and Experimental Data

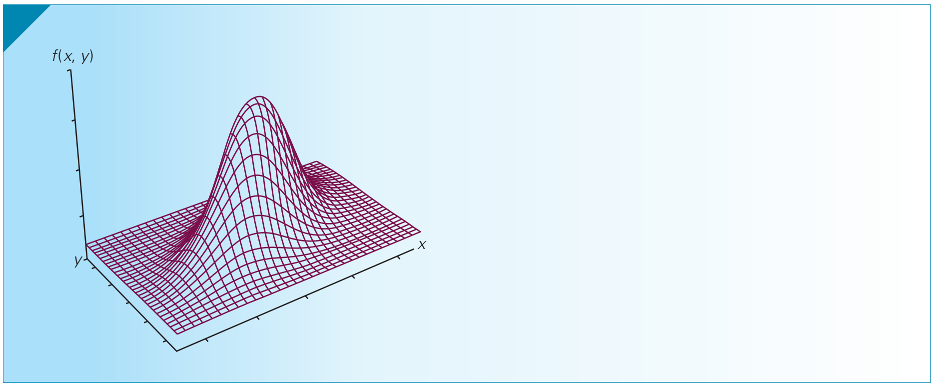
In Chapter 5 and again in Chapter 13, we described the difference between observational and experimental data. We pointed out that statistics practitioners often design controlled experiments to enable them to interpret the results of their analyses more clearly than would be the case after conducting an observational study. Example 16.2 is an illustration of observational data. In that example, we merely observed the odometer reading and auction selling price of 100 randomly selected cars.

If you examine Exercise 16.6, you will see experimental data gathered through a controlled experiment. To determine the effect of the length of a television commercial on its viewers' memories of the product advertised, the statistics practitioner arranged for 60 television viewers to watch a commercial of differing lengths and then tested their memories of that commercial. Each viewer was randomly assigned a commercial length. The values of x ranged from 20 to 60 and were set by the statistics practitioner as part of the experiment. For each value of x, the distribution of the memory test scores is assumed to be normally distributed with a constant variance.

We can summarize the difference between the experiment described in Example 16.2 and the one described in Exercise 16.6. In Example 16.2, both the odometer reading and the auction selling price are random variables. We hypothesize that for each possible odometer reading, there is a theoretical population of auction selling prices that are normally distributed with a mean that is a linear function of the odometer reading and a variance that is constant. In Exercise 16.6, the length of the commercial is not a random variable but a series of values selected by the statistics practitioner. For each commercial length, the memory test scores are required to be normally distributed with a constant variance.

Regression analysis can be applied to data generated from either observational or controlled experiments. In both cases, our objective is to determine how the independent variable is related to the dependent variable. However, observational data can be analyzed in another way. When the data are observational, both variables are random variables. We need not specify that one variable is independent and the other is dependent. We can simply determine *whether* the two variables are related. The equivalent of the required conditions described in the previous box is that the two variables are bivariate normally distributed. (Recall that in Section 7.2 we introduced the bivariate distribution, which describes the joint probability of two variables.) A bivariate normal distribution is described in Figure 16.4. As you can see, it is a three-dimensional bell-shaped curve. The dimensions are the variables x, y, and the joint density function $f(x,y)$.

FIGURE **16.4** Bivariate Normal Distribution



In Section 16.4, we will discuss the statistical technique that is used when both x and y are random variables and they are bivariate normally distributed. In Chapter 19, we will introduce a procedure applied when the normality requirement is not satisfied.

EXERCISES

16.19 Describe what the required conditions mean in Exercise 16.6. If the conditions are satisfied, what can you say about the distribution of memory test scores?

16.20 What are the required conditions for Exercise 16.8? Do these seem reasonable?

16.21 Assuming that the required conditions are satisfied in Exercise 16.13, what does this tell you about the distribution of used boat prices?

16.4 Assessing the Model

The least squares method produces the best straight line. However, there may, in fact, be no relationship or perhaps a nonlinear relationship between the two variables. If so, a straight-line model is likely to be impractical. Consequently, it is important for us to assess how well the linear model fits the data. If the fit is poor, we should discard the linear model and seek another one.

Several methods are used to evaluate the model. In this section, we present two statistics and one test procedure to determine whether a linear model should be employed. They are the **standard error of estimate**, the t-test of the slope, and the coefficient of determination. All these methods are based on the sum of squares for error.

Sum of Squares for Error

The least squares method determines the coefficients that minimize the sum of squared deviations between the points and the line defined by the coefficients. Recall from Section 16.2 that the minimized sum of squared deviations is called the *sum of squares for error*, denoted SSE. In that section, we demonstrated the direct method of calculating SSE. For each value of x, we compute the value of $\hat{y}$. In other words, for $i = 1$ to n, we compute

$$\hat{y}_i = b_0 + b_1 x_i$$

For each point, we then compute the difference between the actual value of y and the value calculated at the line, which is the residual. We square each residual and sum the squared values. Table 16.1 on page 640 shows these calculations for Example 16.1. To calculate SSE manually requires a great deal of arithmetic. Fortunately, there is a shortcut method available that uses the sample variances and the covariance.

Shortcut Calculation of SSE

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = (n-1)\left(s_y^2 - \frac{s_{xy}^2}{s_x^2}\right)$$

where s_y^2 is the sample variance of the dependent variable.

Standard Error of Estimate

In Section 16.3, we pointed out that the error variable ε is normally distributed with mean 0 and standard deviation σ_ε. If σ_ε is large, some of the errors will be large, which implies that the model's fit is poor. If σ_ε is small, the errors tend to be close to the mean (which is 0); as a result, the model fits well. Hence, we could use σ_ε to measure the suitability of using a linear model. Unfortunately, σ_ε is a population parameter and, like most other parameters, is unknown. We can, however, estimate σ_ε from the data. The estimate is based on SSE. The unbiased estimator of the variance of the error variable σ_ε^2 is

$$s_\varepsilon^2 = \frac{\text{SSE}}{n-2}$$

The square root of s_ε^2 is called the *standard error of estimate*.

Standard Error of Estimate

$$s_\varepsilon = \sqrt{\frac{\text{SSE}}{n-2}}$$

EXAMPLE 16.3

Odometer Reading and Prices of Used Toyota Camrys—Part 2

Find the standard error of estimate for Example 16.2 and describe what it tells you about the model's fit.

SOLUTION

COMPUTE

MANUALLY

To compute the standard error of estimate, we must compute SSE, which is calculated from the sample variances and the covariance. We have already determined the covariance and the variance of x: -2.909 and 43.509, respectively. The sample variance of y (applying the shortcut method) is

$$s_y^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right]$$

$$= \frac{1}{100-1}\left[22{,}055.23 - \frac{(1{,}484.1)^2}{100}\right]$$

$$= .300$$

$$\text{SSE} = (n-1)\left(s_y^2 - \frac{s_{xy}^2}{s_x^2}\right)$$

$$= (100-1)\left(.300 - \frac{[-2.909]^2}{43.509}\right)$$

$$= 10.445$$

The standard error of estimate follows:

$$s_\varepsilon = \sqrt{\frac{\text{SSE}}{n-2}} = \sqrt{\frac{10.445}{98}} = .3265$$

EXCEL

	A	B
7	Standard Error	0.3265

This part of the Excel printout was copied from the complete printout on page 642.

MINITAB

S = 0.326489

This part of the Minitab printout was copied from the complete printout on page 643.

INTERPRET

The smallest value that s_ε can assume is 0, which occurs when SSE = 0, that is, when all the points fall on the regression line. Thus, when s_ε is small, the fit is excellent, and the linear model is likely to be an effective analytical and forecasting tool. If s_ε is large, the model is a poor one, and the statistics practitioner should improve it or discard it.

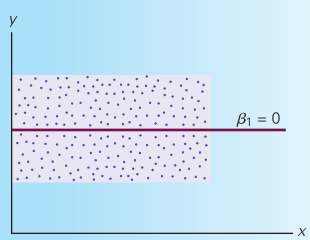
We judge the value of s_ε by comparing it to the values of the dependent variable y or more specifically to the sample mean $\bar{y}$. In this example, because $s_\varepsilon = .3265$ and $\bar{y} = 14.841$, it does appear that the standard error of estimate is small. However, because there is no predefined upper limit on s_ε, it is often difficult to assess the model in this way. In general, the standard error of estimate cannot be used as an absolute measure of the model's utility.

Nonetheless, s_ε is useful in comparing models. If the statistics practitioner has several models from which to choose, the one with the smallest value of s_ε should generally be the one used. As you'll see, s_ε is also an important statistic in other procedures associated with regression analysis.

Testing the Slope

To understand this method of assessing the linear model, consider the consequences of applying the regression technique to two variables that are not at all linearly related. If we could observe the entire population and draw the regression line, we would observe the scatter diagram shown in Figure 16.5. The line is horizontal, which means that no matter what value of x is used, we would estimate the same value for $\hat{y}$; thus, y is not linearly related to x. Recall that a horizontal straight line has a slope of 0, that is, $\beta_1 = 0$.

FIGURE **16.5** Scatter Diagram of Entire Population with $\beta_1 = 0$



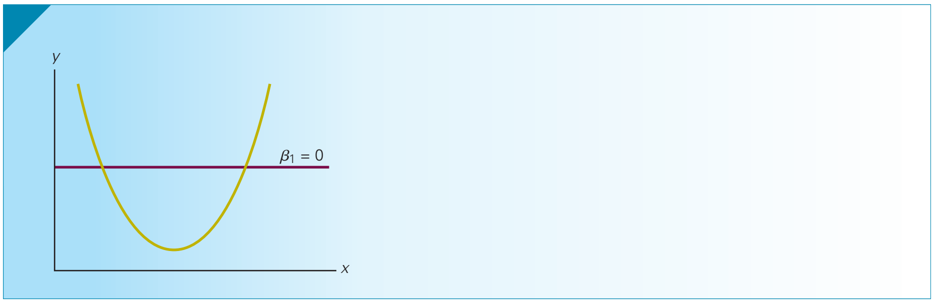
Because we rarely examine complete populations, the parameters are unknown. However, we can draw inferences about the population slope β_1 from the sample slope b_1.

The process of testing hypotheses about β_1 is identical to the process of testing any other parameter. We begin with the hypotheses. The null hypothesis specifies that there is no linear relationship, which means that the slope is 0. Thus, we specify

$$H_0: \quad \beta_1 = 0$$

It must be noted that if the null hypothesis is true, it does not necessarily mean that no relationship exists. For example, a quadratic relationship described in Figure 16.6 may exist where $\beta_1 = 0$.

FIGURE **16.6** Quadratic Relationship



We can conduct one- or two-tail tests of β_1. Most often, we perform a two-tail test to determine whether there is sufficient evidence to infer that a linear relationship exists.* We test the alternative hypothesis

$$H_1: \quad \beta_1 \neq 0$$

Estimator and Sampling Distribution

In Section 16.2, we pointed out that b_1 is an unbiased estimator of β_1; that is,

$$E(b_1) = \beta_1$$

The estimated standard error of b_1 is

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)s_x^2}}$$

where s_ε is the standard error of estimate and s_x^2 is the sample variance of the independent variable. If the required conditions outlined in Section 16.3 are satisfied, the sampling distribution of the t-statistic

$$t = \frac{b_1 - \beta_1}{s_{b_1}}$$

is Student t with degrees of freedom $\nu = n - 2$. Notice that the standard error of b_1 decreases when the sample size increases (which makes b_1 a consistent estimator of β_1) or the variance of the independent variable increases.

Thus, the test statistic and confidence interval estimator are as follows.

Test Statistic for β_1

$$t = \frac{b_1 - \beta_1}{s_{b_1}} \qquad \nu = n - 2$$

*If the alternative hypothesis is true it may be that a linear relationship exists or that a nonlinear relationship exists, but that the relationship can be approximated by a straight line.

Confidence Interval Estimator of β_1

$$b_1 \pm t_{\alpha/2} s_{b_1} \quad \nu = n - 2$$

EXAMPLE 16.4 Are Odometer Reading and Price of Used Toyota Camrys Related?

Test to determine whether there is enough evidence in Example 16.2 to infer that there is a linear relationship between the auction price and the odometer reading for all 3-year-old Toyota Camrys. Use a 5% significance level.

SOLUTION

We test the hypotheses

$$H_0: \quad \beta_1 = 0$$

$$H_1: \quad \beta_1 \neq 0$$

If the null hypothesis is true, no linear relationship exists. If the alternative hypothesis is true, some linear relationship exists.

COMPUTE

MANUALLY

To compute the value of the test statistic, we need b_1 and s_{b_1}. In Example 16.2, we found

$$b_1 = -.0669$$

and

$$s_x^2 = 43.509$$

Thus,

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)s_x^2}} = \frac{.3265}{\sqrt{(99)(43.509)}} = .00497$$

The value of the test statistic is

$$t = \frac{b_1 - \beta_1}{s_{b_1}} = \frac{-.0669 - 0}{.00497} = -13.46$$

The rejection region is

$$t < -t_{\alpha/2,\nu} = -t_{.025,98} \approx -1.984 \quad \text{or} \quad t > t_{\alpha/2,\nu} = t_{.025,98} \approx 1.984$$

EXCEL

	A	B	C	D	E
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*
17	Intercept	17.25	0.182	94.73	3.57E-98
18	Odometer	−0.0669	0.0050	−13.44	5.75E-24

MINITAB

Predictor	Coef	SE Coef	T	P
Constant	17.2487	0.1821	94.73	0.000
Odometer	-0.066861	0.004975	-13.44	0.000

INTERPRET

The value of the test statistic is $t = -13.44$, with a p-value of 0. (Excel uses scientific notation, which in this case is 5.75×10^{-24}, which is approximately 0.) There is overwhelming evidence to infer that a linear relationship exists. What this means is that the odometer reading may affect the auction selling price of the cars. (See the subsection on cause-and-effect relationship on page 659.)

As was the case when we interpreted the y-intercept, the conclusion we draw here is valid only over the range of the values of the independent variable. We can infer that there is a relationship between odometer reading and auction price for the 3-year-old Toyota Camrys whose odometer readings lie between 19.1 (thousand) and 49.2 (thousand) miles (the minimum and maximum values of x in the sample). Because we have no observations outside this range, we do not know how, or even whether, the two variables are related.

Notice that the printout includes a test for β_0. However, as we pointed out before, interpreting the value of the y-intercept can lead to erroneous, if not ridiculous, conclusions. Consequently, we generally ignore the test of β_0.

We can also acquire information about the relationship by estimating the slope coefficient. In this example, the 95% confidence interval estimate (approximating $t_{.025}$ with 98 degrees of freedom with $t_{.025}$ with 100 degrees of freedom) is

$$b_1 \pm t_{\alpha/2} s_{b_1} = -.0669 \pm 1.984(.00497) = -.0669 \pm .0099$$

We estimate that the slope coefficient lies between $-.0768$ and $-.0570$.

One-Tail Tests

If we wish to test for positive or negative linear relationships, we conduct one-tail tests. To illustrate, suppose that in Example 16.2 we wanted to know whether there is evidence of a negative linear relationship between odometer reading and auction selling price. We would specify the hypotheses as

$$H_0: \quad \beta_1 = 0$$

$$H_1: \quad \beta_1 < 0$$

The value of the test statistic would be exactly as computed previously (Example 16.4). However, in this case the p-value would be the two-tail p-value divided by 2; using Excel's p-value, this would be $(5.75 \times 10^{-24})/2 = 2.875 \times 10^{-24}$, which is still approximately 0.

Coefficient of Determination

The test of β_1 addresses only the question of whether there is enough evidence to infer that a linear relationship exists. In many cases, however, it is also useful to measure the strength of that linear relationship, particularly when we want to compare several different models. The statistic that performs this function is the **coefficient of determination**, which is denoted R^2. Statistics practitioners often refer to this statistic as the "R-square." Recall that we introduced the coefficient of determination in Chapter 4, where we pointed out that this statistic is a measure of the amount of variation in the dependent variable that is explained by the variation in the independent variable. However, we did not describe why we interpret the R-square in this way.

Coefficient of Determination

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2}$$

With a little algebra, statisticians can show that

$$R^2 = 1 - \frac{\text{SSE}}{\sum (y_i - \bar{y})^2}$$

We'll return to Example 16.1 to learn more about how to interpret the coefficient of determination. In Chapter 14, we partitioned the total sum of squares into two sources of variation. We do so here as well. We begin by adding and subtracting $\hat{y}_i$ from the deviation between y_i from the mean $\bar{y}$; that is,

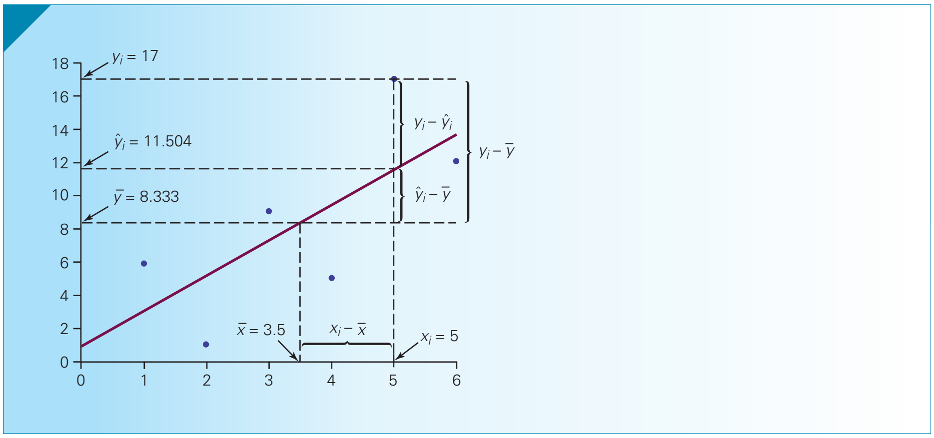
$$(y_i - \bar{y}) = (y_i - \bar{y}) + \hat{y}_i - \hat{y}_i$$

We observe that by rearranging the terms, the deviation between y_i and $\bar{y}$ can be decomposed into two parts; that is,

$$(y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

This equation is represented graphically (for $i = 5$) in Figure 16.7.

FIGURE **16.7** Partitioning the Deviation for $i = 5$



Now we ask why the values of y are different from one another. From Figure 16.7, we see that part of the difference between y_i and $\bar{y}$ is the difference between $\hat{y}_i$ and $\bar{y}$, which is accounted for by the difference between x_i and $\bar{x}$. In other words, some of the variation in y is explained by the changes to x. The other part of the difference between y_i and $\bar{y}$, however, is accounted for by the difference between y_i and $\hat{y}_i$. This difference is the residual, which represents variables not otherwise represented by the model. As a result, we say that this part of the difference is *unexplained* by the variation in x.

If we now square both sides of the equation, sum over all sample points, and perform some algebra, we produce

$$\sum(y_i - \bar{y})^2 = \sum(y_i - \hat{y}_i)^2 + \sum(\hat{y}_i - \bar{y})^2$$

The quantity on the left side of this equation is a measure of the variation in the dependent variable y. The first quantity on the right side of the equation is SSE, and the second term is denoted SSR, for sum of squares for regression. We can rewrite the equation as

$$\text{Variation in } y = \text{SSE} + \text{SSR}$$

As we did in the analysis of variance, we partition the variation of y into two parts: SSE, which measures the amount of variation in y that remains unexplained; and SSR, which measures the amount of variation in y that is explained by the variation in the independent variable x. We can incorporate this analysis into the definition of R^2.

Coefficient of Determination

$$R^2 = 1 - \frac{\text{SSE}}{\sum(y_i - \bar{y})^2} = \frac{\sum(y_i - \bar{y})^2 - \text{SSE}}{\sum(y_i - \bar{y})^2} = \frac{\text{Explained variation}}{\text{Variation in } y}$$

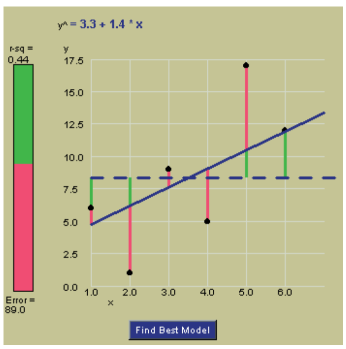
It follows that R^2 measures the proportion of the variation in y that can be explained by the variation in x.

SEEING STATISTICS

applet 19 Analysis of Regression Deviations

This applet provides another way to understand the coefficient of determination.

Move the regression line to reduce the sum of squared errors. The vertical line from each point to the horizontal line depicts the deviation from the mean. In regression this is divided into two parts—the green part, which is the deviation that is eliminated by using the regression line, and the red part, which is the deviation remaining. Note that for some points the deviations become larger.



Applet Exercises

Change the slope (if necessary) so that the line is horizontal.

19.1 How much of the variation in y is explained by the variation in x? Why is this so?

Move the line so that it goes through the sixth point ($x = 6$).

19.2 What is the value of R^2?

19.3 How much of the variation between y_6 and $\bar{y}$ is explained by the variation between x_6 and $\bar{x}$? Why is this so?

Produce the least squares line. (Click the **Find Best Model** button.)

19.4 How much of the variation in y is explained by the variation in x?

EXAMPLE 16.5

Measuring the Strength of the Linear Relationship between Odometer Reading and Price of Used Toyota Camrys

Find the coefficient of determination for Example 16.2 and describe what this statistic tells you about the regression model.

SOLUTION

COMPUTE

MANUALLY

We have already calculated all the necessary components of this statistic. In Example 16.2 we found

$$s_{xy} = -2.909$$
$$s_x^2 = 43.509$$

and from Example 16.3

$$s_y^2 = .300$$

Thus,

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = \frac{(-2.909)^2}{(43.509)(.300)} = .6483$$

EXCEL

	A	B
5	R Square	0.6483

MINITAB

R-Sq = 64.8%

Both Minitab and Excel print a second R^2 statistic called the *coefficient of determination adjusted for degrees of freedom*. We will define and describe this statistic in Chapter 17.

INTERPRET

We found that R^2 is equal to .6483. This statistic tells us that 64.83% of the variation in the auction selling prices is explained by the variation in the odometer readings. The remaining 35.17% is unexplained. Unlike the value of a test statistic, the coefficient of determination does not have a critical value that enables us to draw conclusions. In general, the higher the value of R^2, the better the model fits the data. From the t-test of β_1 we already know that there is evidence of a linear relationship. The coefficient of determination merely supplies us with a measure of the strength of that relationship. As you will discover in the next chapter, when we improve the model, the value of R^2 increases.

Other Parts of the Computer Printout

The last part of the printout shown on pages 642 and 643 relates to our discussion of the interpretation of the value of R^2, when its meaning is derived from the partitioning of the variation in y. The values of SSR and SSE are shown in an analysis of variance table similar to the tables introduced in Chapter 14. The general form of the table is shown in Table 16.2. The F-test performed in the ANOVA table will be explained in Chapter 17.

TABLE **16.2** **General Form of the ANOVA Table in the Simple Linear Regression Model**

SOURCE	d.f.	SUMS OF SQUARES	MEAN SQUARES	*F*-STATISTIC
Regression	1	SSR	MSR = SSR/1	F = MSR/MSE
Error	$n - 2$	SSE	MSE = SSE/($n - 2$)	
Total	$n - 1$	Variation in y		

Note: Excel uses the word "Residual" to refer to the second source of variation, which we called "Error."

Developing an Understanding of Statistical Concepts

Once again, we encounter the concept of explained variation. We first discussed the concept in Chapter 13 when we introduced the matched pairs experiment, where the experiment was designed to reduce the variation among experimental units. This concept was extended in the analysis of variance, where we partitioned the total variation into two or more sources (depending on the experimental design). And now in regression analysis, we use the concept to measure how the dependent variable is related to the independent variable. We partition the variation of the dependent variable into the sources: the variation explained by the variation in the independent variable and the unexplained variation. The greater the explained variation, the better the model is. We often refer to the coefficient of determination as a measure of the explanatory power of the model.

Cause-and-Effect Relationship

A common mistake is made by many students when they attempt to interpret the results of a regression analysis when there is evidence of a linear relationship. They imply that changes in the independent variable cause changes in the dependent variable. It must be emphasized that we cannot infer a causal relationship from statistics alone. Any inference about the cause of the changes in the dependent variable must be justified by a reasonable theoretical relationship. For example, statistical tests established that the more one smoked, the greater the probability of developing lung cancer. However, this analysis did not prove that smoking causes lung cancer. It only demonstrated that smoking and lung cancer were somehow related. Only when medical investigations established the connection were scientists able to confidently declare that smoking causes lung cancer.

As another illustration, consider Example 16.2 where we showed that the odometer reading is linearly related to the auction price. Although it seems reasonable to conclude that decreasing the odometer reading would cause the auction price to rise, the conclusion may not be entirely true. It is theoretically possible that the price is determined by the overall condition of the car and that the condition generally worsens when the car is driven longer. Another analysis would be needed to establish the veracity of this conclusion.

Be cautious about the use of the terms *explained variation* and *explanatory power of the model*. Do not interpret the word *explained* to mean *caused*. We say that the coefficient of determination measures the amount of variation in y that is explained (not caused) by the variation in x. Thus, regression analysis can only show that a statistical relationship exists. We cannot infer that one variable causes another.

Recall that we first pointed this out in Chapter 3 using the following sentence:

Correlation is not causation.

Testing the Coefficient of Correlation

When we introduced the coefficient of correlation (also called the **Pearson coefficient of correlation**) in Chapter 4, we observed that it is used to measure the strength of association between two variables. However, the coefficient of correlation can be useful in another way. We can use it to test for a linear relationship between two variables.

When we are interested in determining *how* the independent variable is related to the dependent variable, we estimate and test the linear regression model. The t-test of the slope presented previously allows us to determine whether a linear relationship actually exists. As we pointed out in Section 16.3, the statistical test requires that for each value of x, there exists a population of values of y that are normally distributed with a constant variance. This condition is required whether the data are experimental or observational.

In many circumstances, we're interested in determining only *whether* a linear relationship exists and not the form of the relationship. When the data are observational and the two variables are bivariate normally distributed (See Section 16.3.) we can calculate the coefficient of correlation and use it to test for linear association.

As we noted in Chapter 4, the population coefficient of correlation is denoted ρ (the Greek letter *rho*). Because ρ is a population parameter (which is almost always unknown), we must estimate its value from the sample data. Recall that the sample coefficient of correlation is defined as follows.

Sample Coefficient of Correlation

$$r = \frac{s_{xy}}{s_x s_y}$$

When there is no linear relationship between the two variables, $\rho = 0$. To determine whether we can infer that ρ is 0, we test the hypotheses

H_0: $\rho = 0$

H_1: $\rho \neq 0$

The test statistic is defined in the following way.

Test Statistic for Testing $\rho = 0$

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

which is Student t distributed with $\nu = n - 2$ degrees of freedom provided that the variables are bivariate normally distributed.

EXAMPLE 16.6

Are Odometer Reading and Price of Used Toyota Camrys Linearly Related? Testing the Coefficient of Correlation

Conduct the t-test of the coefficient of correlation to determine whether odometer reading and auction selling price are linearly related in Example 16.2. Assume that the two variables are bivariate normally distributed.

SOLUTION

COMPUTE

MANUALLY

The hypotheses to be tested are

$$H_0: \quad \rho = 0$$

$$H_1: \quad \rho \neq 0$$

In Example 16.2, we found $s_{xy} = -2.909$ and $s_x^2 = 43.509$. In Example 16.5, we determined that $s_y^2 = .300$. Thus,

$$s_x = \sqrt{43.509} = 6.596$$

$$s_y = \sqrt{.300} = .5477$$

The coefficient of correlation is

$$r = \frac{s_{xy}}{s_x s_y} = \frac{-2.909}{(6.596)(.5477)} = -.8052$$

The value of the test statistic is

$$t = r\sqrt{\frac{n-2}{1-r^2}} = -.8052\sqrt{\frac{100-2}{1-(-.8052)^2}} = -13.44$$

Notice that this is the same value we produced in the t-test of the slope in Example 16.4. Because both sampling distributions are Student t with 98 degrees of freedom, the p-value and conclusion are also identical.

EXCEL

	A	B
1	**Correlation**	
2		
3	*Price and Odometer*	
4	Pearson Coefficient of Correlation	–0.8052
5	t Stat	–13.44
6	df	98
7	P(T<=t) one tail	0
8	t Critical one tail	1.6606
9	P(T<=t) two tail	0
10	t Critical two tail	1.9845

INSTRUCTIONS

1. Type or import the data into two adjacent columns*. (Open Xm16-02.)
2. Click **Add-ins, Data Analysis Plus**, and **Correlation (Pearson).**
3. Specify the **Variable 1 Input Range** (A1:A101), **Variable 2 Input Range** (B1:B101), and α (.05).

MINITAB

Correlations: Odometer, Price

Pearson correlation of Price and Odometer = –0.805
P-Value = 0.000

INSTRUCTIONS

1. Type or import the data into two adjacent columns. (Open Xm16-02.)
2. Click **Stat, Basic Statistics**, and **Correlation**.
3. Type the names of the variables in the **Variables** box (Odometer Price).

Notice that the t-test of ρ and the t-test of β_1 in Example 16.4 produced identical results. This should not be surprising because both tests are conducted to determine whether there is evidence of a linear relationship. The decision about which test to use is based on the type of experiment and the information we seek from the statistical analysis. If we're interested in discovering the relationship between two variables, or if we've conducted an experiment where we controlled the values of the independent variable (as in Exercise 16.6), the t-test of β_1 should be applied. If we're interested only in determining *whether* two random variables that are bivariate normally distributed are linearly related, the t-test of ρ should be applied.

As is the case with the t-test of the slope, we can also conduct one-tail tests. We can test for a positive or a negative linear relationship.

*If one or both columns contain a blank (representing missing data) the row must be deleted.

Education and Income: How Are They Related?

© Vicki Beaver

IDENTIFY

The problem objective is to analyze the relationship between two interval variables. Because we want to know how education affects income the independent variable is education (EDUC) and the dependent variable is income (INCOME).

COMPUTE

EXCEL

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.3790				
5	R Square	0.1436				
6	Adjusted R Square	0.1429				
7	Standard Error	35,972				
8	Observations	1189				
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression	1	257,561,051,309	257,561,051,309	199.04	6.702E-42
13	Residual	1187	1,535,986,496,000	1,294,007,158		
14	Total	1188	1,793,547,547,309			
15						
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
17	Intercept	–28926	5117	–5.65	1.971E-08	
18	EDUC	5111	362	14.11	6.702E-42	

MINITAB

Regression Analysis: INCOME versus EDUC

```
The regression equation is
Income = –28926 + 5111 EDUC
1189 cases used, 834 cases contain missing values

Predictor    Coef     SE Coef     T       P
Constant    –28926    5117      –5.65   0.000
EDUC         5110.7   362.2     14.11   0.000


S =  35972.3   R-Sq = 14.4%   R-Sq(adj) = 14.3%


Analysis of Variance

Source          DF      SS            MS           F       P
Regression       1   2.57561E+11   2.57561E+11   199.04   0.000
Residual Error 1187  1.53599E+12   1294007158
Total          1188  1.79355E+12
```

INTERPRET

The regression equation is $\hat{y} = -28926 + 5111x$. The slope coefficient tells us that on average for each additional year of education income increases by \$5,111. We test to determine whether there is evidence of a linear relationship.

$$H_0: \beta_1 = 0$$
$$H_1: \beta_1 \neq 0$$

The test statistic is $t = 14.11$ and the p-value is 6.702 X 10^{-42}, which is virtually 0. The coefficient of determination is $R^2 = .1436$, which means that 14.36% of the variation in income is explained by the variation in education and the remaining 85.64% is not explained.

Violation of the Required Condition

When the normality requirement is unsatisfied, we can use a nonparametric technique—the Spearman rank correlation coefficient (Chapter 19*)—to replace the t-test of ρ.

*Instructors who wish to teach the use of the Spearman rank correlation coefficient here can use Keller's website Appendix Spearman Rank Correlation Coefficient and Test.

EXERCISES

Use a 5% significance level for all tests of hypotheses.

16.22 You have been given the following data:

x	1	3	4	6	9	8	10
y	1	8	15	33	75	70	95

a. Draw the scatter diagram. Does it appear that x and y are related? If so, how?
b. Test to determine whether there is evidence of a linear relationship.

16.23 Suppose that you have the following data:

x	3	5	2	6	1	4
y	25	110	9	250	3	71

a. Draw the scatter diagram. Does it appear that x and y are related? If so, how?
b. Test to determine whether there is evidence of a linear relationship.

16.24 Refer to Exercise 16.2.
a. Determine the standard error of estimate.
b. Is there evidence of a linear relationship between advertising and sales?
c. Estimate β_1 with 95% confidence.
d. Compute the coefficient of determination and interpret this value.
e. Briefly summarize what you have learned in parts (a) through (d).

16.25 Calculate the coefficient of determination and conduct a test to determine whether a linear relationship exists between housing starts and mortgage interest in Exercise 16.3.

16.26 Is there evidence of a linear relationship between the number of hours of television viewing and how overweight the child is in Exercise 16.4?

16.27 Determine whether there is evidence of a negative linear relationship between temperature and the number of beers sold at Yankee Stadium in Exercise 16.5.

Exercises 16.28–16.53 require the use of a computer and software. The answers to Exercises 16.28 to 16.44 may be calculated manually. See Appendix A for the sample statistics.

16.28 Refer to Exercise 16.6.
a. What is the standard error of estimate? Interpret its value.
b. Describe how well the memory test scores and length of television commercial are linearly related.
c. Are the memory test scores and length of commercial linearly related? Test using a 5% significance level.
d. Estimate the slope coefficient with 90% confidence.

16.29 Refer to Exercise 16.7. Apply the three methods of assessing the model to determine how well the linear model fits.

16.30 Is there enough evidence to infer that age and the amount of time needed to complete the questionnaire are linearly related in Exercise 16.8?

16.31 Refer to Exercise 16.9. Use two statistics to measure the strength of the linear association. What do these statistics tell you?

16.32 Is there evidence of a linear relationship between number of cigarettes smoked and number of sick days in Exercise 16.10?

16.33 Refer to Exercise 16.11.
a. Test to determine whether there is evidence of a linear relationship between distance to the nearest fire station and percentage of damage.
b. Estimate the slope coefficient with 95% confidence.
c. Determine the coefficient of determination. What does this statistic tell you about the relationship?

16.34 Refer to Exercise 16.12.
a. Determine the standard error of estimate, and describe what this statistic tells you about the regression line.
b. Can we conclude that the size and price of the apartment building are linearly related?
c. Determine the coefficient of determination and discuss what its value tells you about the two variables.

16.35 Is there enough evidence to infer that as the number of hours of engine use increases, the price decreases in Exercise 16.13?

16.36 Assess fit of the regression line in Exercise 16.14.

16.37 Refer to Exercise 16.15.
a. Determine the coefficient of determination and describe what it tells you.
b. Conduct a test to determine whether there is evidence of a linear relationship between household income and food budget.

16.38 Can we infer that office rents and vacancy rates are linearly related in Exercise 16.16?

16.39 Are height and income in Exercise 16.17 positively linearly related?

16.40 Refer to Exercise 16.18.
a. Compute the coefficient of determination and describe what it tells you.
b. Can we infer that aptitude test scores and percentages of nondefectives are linearly related?

16.41 Repeat Exercise 16.13 using the t-test of the coefficient of correlation to determine whether there is a negative linear relationship between the number of hours of engine use and the selling price of the used boats.

16.42 Repeat Exercise 16.6 using the t-test of the coefficient of correlation. Is this result identical to the one you produced in Exercise 16.6?

16.43 Are food budget and household income in Exercise 16.15 linearly related? Employ the t-test of the coefficient of correlation to answer the question.

16.44 Refer to Exercise 16.10. Use the t-test of the coefficient of correlation to determine whether there is evidence of a positive linear relationship between number of cigarettes smoked and the number of sick days.

AMERICAN NATIONAL ELECTION SURVEY EXERCISES

16.45 ANES2008* Do more educated people spend more time watching or reading news on the Internet? Conduct a regression analysis to determine whether there is enough statistical evidence to conclude that the more education (EDUC) one has the more one watches or reads news on the Internet (TIME1)?

16.46 ANES2008* In the Chapter 16 opening example, we analyzed the relationship between income and education using the 2008 General Social Survey of 2008. Conduct a similar analysis using the 2008 American National Election Survey.

16.47 ANES2008* National news on television features commercials describing pharmaceutical drugs that treat ailments that plague older people. Apparently, the major networks believe that older people tend to watch national newscasts. Is there sufficient evidence to conclude age (AGE) and number of days watching national news on television (DAYS1) are positively related?

16.48 ANES2008* In most presidential elections in the United States, the voter turnout is quite low, often in the neighborhood of 50%. Political workers would like to be able to predict who is likely to vote. Thus, it is important to know which variables are related to intention to vote. One candidate is age. Is there sufficient evidence to infer that age (AGE) and intention to vote (DEFINITE) are linearly related?

16.49 ANES2008* Do more affluent people get their news from radio? Answer the question by conducting an analysis of the relationship between income (INCOME) and time listening to news on the radio (TIME4).

General Social Survey Exercises

16.50 GSS2008* Does income affect people's positions on the question, Should the government reduce income differences between rich and poor (EQWLTH)? Answer the question by testing the relationship between income (INCOME) and EQWLTH.

16.51 GSS2008* Conduct an analysis of the relationship between income (INCOME) and age (AGE). Estimate with 95% confidence the average increase in income for each additional year of age.

16.52 GSS2008* Is there sufficient evidence to conclude that more educated people (EDUC) watch less television (TVHOURS)?

16.53 GSS2006* Use the 2006 survey data to determine whether more education (EDUC) leads to higher income (INCOME).

16.5 Using the Regression Equation

Using the techniques in Section 16.4, we can assess how well the linear model fits the data. If the model fits satisfactorily, we can use it to forecast and estimate values of the dependent variable. To illustrate, suppose that in Example 16.2, the used-car dealer wanted to predict the selling price of a 3-year-old Toyota Camry with 40 (thousand) miles on the odometer. Using the regression equation, with $x = 40$, we get

$$\hat{y} = 17.250 - .0669x = 17.250 - 0.0669(40) = 14.574$$

We call this value the **point prediction**, and $\hat{y}$ is the point estimate or predicted value for y when $x = 40$. Thus, the dealer would predict that the car would sell for $14,574.

By itself, however, the point prediction does not provide any information about how closely the value will match the true selling price. To discover that information, we must use an interval. In fact, we can use one of two intervals: the prediction interval of a particular value of y or the confidence interval estimator of the expected value of y.

Predicting the Particular Value of y for a Given x

The first confidence interval we present is used whenever we want to predict a one-time occurrence for a particular value of the dependent variable when the independent variable is a given value x_g. This interval, often called the **prediction interval**, is calculated in the usual way (point estimator $\pm$ bound on the error of estimation). Here the point estimate for y is $\hat{y}$, and the bound on the error of estimation is shown below.

Prediction Interval

$$\hat{y} \pm t_{\alpha/2,n-2}s_\varepsilon\sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n - 1)s_x^2}}$$

where x_g is the given value of x and $\hat{y} = b_0 + b_1x_g$

Estimating the Expected Value of y for a Given x

The conditions described in Section 16.3 imply that, for a given value of x, there is a population of values of y whose mean is

$$E(y) = \beta_0 + \beta_1 x$$

To estimate the mean of y or long-run average value of y we would use the following interval referred to simply as the confidence interval. Again, the point estimator is $\hat{y}$, but the bound on the error of estimation is different from the prediction interval shown below.

Confidence Interval Estimator of the Expected Value of *y*

$$\hat{y} \pm t_{\alpha/2,n-2} s_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$$

Unlike the formula for the prediction interval, this formula does not include the 1 under the square-root sign. As a result, the **confidence interval estimate of the expected value of *y*** will be narrower than the prediction interval for the same given value of x and confidence level. This is because there is less error in estimating a mean value as opposed to predicting an individual value.

EXAMPLE 16.7

Predicting the Price and Estimating the Mean Price of Used Toyota Camrys

a. A used-car dealer is about to bid on a 3-year-old Toyota Camry equipped with all the standard features and with 40,000 ($x_g = 40$) miles on the odometer. To help him decide how much to bid, he needs to predict the selling price.

b. The used-car dealer mentioned in part (a) has an opportunity to bid on a lot of cars offered by a rental company. The rental company has 250 Toyota Camrys all equipped with standard features. All the cars in this lot have about 40,000 ($x_g = 40$) miles on their odometers. The dealer would like an estimate of the selling price of all the cars in the lot.

SOLUTION

IDENTIFY

a. The dealer would like to predict the selling price of a single car. Thus, he must employ the prediction interval

$$\hat{y} \pm t_{\alpha/2,n-2} s_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$$

b. The dealer wants to determine the mean price of a large lot of cars, so he needs to calculate the confidence interval estimator of the expected value:

$$\hat{y} \pm t_{\alpha/2,n-2}s_\varepsilon\sqrt{\frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$$

Technically, this formula is used for infinitely large populations. However, we can interpret our problem as attempting to determine the average selling price of all Toyota Camrys equipped as described above, all with 40,000 miles on the odometer. The crucial factor in part (b) is the need to estimate the mean price of a number of cars. We arbitrarily select a 95% confidence level.

COMPUTE

MANUALLY

From previous calculations, we have the following:

$$\hat{y} = 17.250 - .0669(40) = 14.574$$

$$s_\varepsilon = .3265$$

$$s_x^2 = 43.509$$

$$\bar{x} = 36.011$$

From Table 4 in Appendix B, we find

$$t_{\alpha/2} = t_{.025,98} \approx t_{.025,100} = 1.984$$

a. The 95% prediction interval is

$$\hat{y} \pm t_{\alpha/2,n-2}s_\varepsilon\sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$$

$$= 14.574 \pm 1.984 \times .3265\sqrt{1 + \frac{1}{100} + \frac{(40 - 36.011)^2}{(100 - 1)(43.509)}}$$

$$= 14.574 \pm .652$$

The lower and upper limits of the prediction interval are \$13,922 and \$15,226, respectively.

b. The 95% confidence interval estimator of the mean price is

$$\hat{y} \pm t_{\alpha/2,n-2}s_\varepsilon\sqrt{\frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$$

$$= 14.574 \pm 1.984 \times .3265\sqrt{\frac{1}{100} + \frac{(40 - 36.011)^2}{(100 - 1)(43.509)}}$$

$$= 14.574 \pm .076$$

The lower and upper limits of the confidence interval estimate of the expected value are \$14,498 and 14,650, respectively.

EXCEL

	A	B	C
1	**Prediction Interval**		
2			
3			Price
4			
5	Predicted value		14.574
6			
7	Prediction Interval		
8	Lower limit		13.922
9	Upper limit		15.227
10			
11	Interval Estimate of Expected Value		
12	Lower limit		14.498
13	Upper limit		14.650

INSTRUCTIONS

1. Type or import the data into two columns*. (Open Xm16-02.)
2. Type the given value of x into any cell. We suggest the next available row in the column containing the independent variable.
3. Click **Add-Ins, Data Analysis Plus,** and **Prediction Interval**.
4. Specify the **Input Y Range** (A1:A101), the **Input X Range** (B1:B101), the **Given X Range** (B102), and the **Confidence Level** (.95).

MINITAB

```
Predicted Values for New Observations

New
Obs    Fit      SE Fit   95% CI              95% PI
  1    14.5743  0.0382   (14.4985, 14.6501)  (13.9220, 15.2266)


Values of Predictors for New Observations

New
Obs    Odometer
  1       40.0
```

The output includes the predicted value $\hat{y}$ (**Fit**), the standard deviation of $\hat{y}$ (**SE Fit**), the 95% confidence interval estimate of the expected value of y (**CI**), and the 95% prediction interval (**PI**).

INSTRUCTIONS

1. Proceed through the three steps of regression analysis described on page 642. Do not click **OK**. Click **Options . . . .**
2. Specify the given value of x in the **Prediction intervals for new observations** box (40).
3. Specify the confidence level (.95).

INTERPRET

We predict that one car will sell for between $13,925 and $15,226. The average selling price of the population of 3-year-old Toyota Camrys is estimated to lie between $14,498 and $14,650. Because predicting the selling price of one car is more difficult than estimating the mean selling price of all similar cars, the prediction interval is wider than the interval estimate of the expected value.

*If one or both columns contain a blank (representing missing data) the row must be deleted.

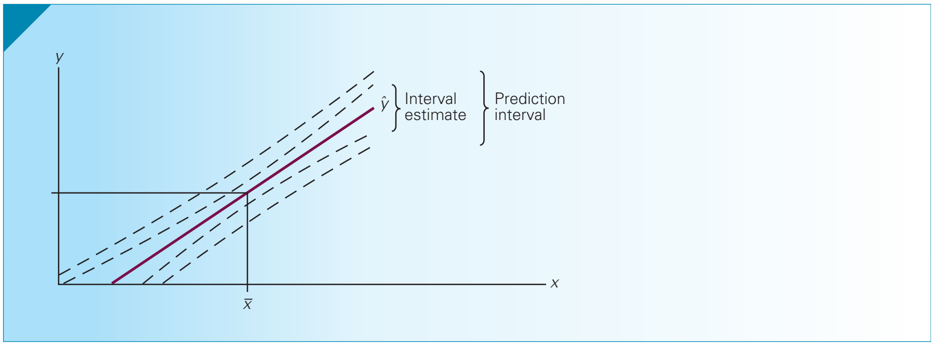
Effect of the Given Value of *x* on the Intervals

Calculating the two intervals for various values of x results in the graph in Figure 16.8. Notice that both intervals are represented by curved lines. This is because the farther the given value of x is from $\bar{x}$, the greater the estimated error becomes. This part of the estimated error is measured by

$$\frac{(x_g - \bar{x})^2}{(n-1)s_x^2}$$

which appears in both the prediction interval and the interval estimate of the expected value.

FIGURE **16.8** Interval Estimates and Prediction Intervals



EXERCISES

16.54 Briefly describe the difference between predicting a value of y and estimating the expected value of y.

16.55 Use the regression equation in Exercise 16.2 to predict with 90% confidence the sales when the advertising budget is $90,000.

16.56 Estimate with 90% confidence the mean monthly number of housing starts when the mortgage interest rate is 8% in Exercise 16.3.

16.57 Refer to Exercise 16.4.
a. Predict with 90% confidence the number of pounds overweight for a child who watches 30 hours of television per week.
b. Estimate with 90% confidence the mean number of pounds overweight for children who watch 30 hours of television per week.

16.58 Refer to Exercise 16.5. Predict with 90% confidence the number of beers to be sold when the temperature is 80 degrees.

Exercises 16.59–16.80 require the use of a computer and software. The answers to Exercises 16.59–16.72 may be calculated manually. See Appendix A for the sample statistics.

16.59 Refer to Exercise 16.6.
a. Predict with 95% confidence the memory test score of a viewer who watches a 36-second commercial.
b. Estimate with 95% confidence the mean memory test score of people who watch 36-second commercials.

16.60 Refer to Exercise 16.7.
a. Predict with 95% confidence the selling price of a 1,200 sq. ft. condominium on the 25th floor.
b. Estimate with 99% confidence the average selling price of a 1,200 sq. ft. condominium on the 12th floor.

16.61 Refer to Exercise 16.8. Estimate with 90% confidence the mean amount of time for 50-year old Americans to complete the census.

16.62 Refer to Exercise 16.9. The company has just hired a 25-year-old telemarketer. Predict with 95% confidence how long he will stay with the company.

16.63 Refer to Exercise 16.10. Predict with 95% confidence the number of sick days for individuals who smoke on average 30 cigarettes per day.

16.64 Refer to Exercise 16.11.

a. Predict with 95% confidence the percentage loss resulting from fire for a house that is 5 miles away from the nearest fire station.

b. Estimate with 95% confidence the average percentage loss resulting from fire for houses that are 2 miles away from the nearest fire station.

16.65 Refer to Exercise 16.12. Estimate with 95% confidence the mean price of 50,000 sq. ft. apartment buildings.

16.66 Refer to Exercise 16.13. Predict with 99% confidence the price of a 1999 24-foot Sea Ray cruiser with 500 hours of engine use.

16.67 Refer to Exercise 16.14. Estimate with 90% confidence the mean electricity consumption for households with 5 occupants.

16.68 Refer to Exercise 16.15. Predict the food budget of a family whose household income is $50,000. Use a 90% confidence level.

16.69 Refer to Exercise 16.16. Predict with 95% confidence the monthly office rent in a city when the vacancy rate is 10%.

16.70 Refer to Exercise 16.17

a. Estimate with 95% confidence the mean annual income of 6-foot-tall men.

b. Suppose that an individual is 5 feet 6 inches tall. Predict with 95% confidence his annual income.

16.71 Refer to Exercise 16.18. Estimate with 95% confidence the mean percentage of defectives for workers who score 75 on the dexterity test.

16.72 Refer to Exercise 16.18. Predict with 90% confidence the percentage of defectives for a worker who scored 80 on the dexterity test.

American National Election Survey Exercise

16.73 ANES2008* Refer to Exercise 16.45. Predict with 90% confidence the amount of time spent watching or reading news on the Internet by a person with 15 years of education

16.74 ANES2008* Refer to Exercise 16.46. Estimate with 95% confidence the income of average of people who have 10 years of education.

16.75 ANES2008* Refer to Exercise 16.47. Estimate with 99% confidence the mean number of days watching national news on television by 50-year-old people.

16.76 ANES2008* Refer to Exercise 16.49. Predict with 95% confidence the amount of time listening to news on the radio by individuals who earn $50,000 annually.

General Social Survey Exercises

16.77 GSS2008* Refer to Exercise 16.51. Predict the annual income of someone who is 45 years old.

16.78 GSS2008* Refer to Exercise 16.52. Estimate with 90% confidence the average number of hours of television watching per day for people with 12 years of education.

16.79 GSS2006* Refer to Exercise 16.53. Use the General Social Survey of 2006 to predict with 99% confidence the annual income of someone with 17 years of education.

16.80 GSS2008* Refer to Exercise 16.52. Predict with 90% confidence the number of hours of television watching per day for someone with 8 years of education.

16.6 Regression Diagnostics—I

In Section 16.3, we described the required conditions for the validity of regression analysis. Simply put, the error variable must be normally distributed with a constant variance, and the errors must be independent of each other. In this section, we show how to diagnose violations. In addition, we discuss how to deal with observations that are unusually large or small. Such observations must be investigated to determine whether an error was made in recording them.

Residual Analysis

Most departures from required conditions can be diagnosed by examining the residuals, which we discussed in Section 16.4. Most computer packages allow you to output the values of the residuals and apply various graphical and statistical techniques to this variable.

We can also compute the standardized residuals. We standardize residuals in the same way we standardize all variables, by subtracting the mean and dividing by the standard deviation. The mean of the residuals is 0, and because the standard deviation σ_ε is unknown, we must estimate its value. The simplest estimate is the standard error of estimate s_ε. Thus,

$$\text{Standardized residuals for point } i = \frac{e_i}{s_\varepsilon}$$

EXCEL

Excel calculates the standardized residuals by dividing the residuals by the standard deviation of the residuals. (The difference between the standard error of estimate and the standard deviation of the residuals is that in the formula of the former the denominator is $n - 2$, whereas in the formula for the latter, the denominator is $n - 1$.)

Part of the printout (we show only the first five and last five values) for Example 16.2 follows.

	A	B	C	D
1	RESIDUAL OUTPUT			
2				
3	*Observation*	*Predicted Price*	*Residuals*	*Standard Residuals*
4	1	14.748	−0.148	−0.456
5	2	14.253	−0.153	−0.472
6	3	14.186	−0.186	−0.574
7	4	15.183	0.417	1.285
8	5	15.129	0.471	1.449
9				
10				
11				
12	95	15.149	−0.049	−0.152
13	96	14.828	−0.028	−0.087
14	97	14.962	−0.362	−1.115
15	98	15.029	−0.529	−1.628
16	99	14.628	0.072	0.222
17	100	14.815	−0.515	−1.585

INSTRUCTIONS

Proceed with the three steps of regression analysis described on page 642. Before clicking **OK**, select **Residuals** and **Standardized Residuals**. The predicted values, residuals, and standardized residuals will be printed.

We can also standardize by computing the standard deviation of each residual. Statisticians have determined that the standard deviation of the residual for observation i is defined as follows.

Standard Deviation of the *i*th Residual

$$s_{e_i} = s_\varepsilon \sqrt{1 - h_i}$$

where

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{(n - 1)s_x^2}$$

The quantity h_i should look familiar; it was used in the formula for the prediction interval and confidence interval estimate of the expected value of y in Section 16.6. Minitab computes this version of the standardized residuals. Part of the printout (we show only the first five and last five values) for Example 16.2 is shown below.

MINITAB

Obs	Odometer	Price	Fit	SE Fit	Residual	St Resid
1	37.4	14.6000	14.7481	0.0334	–0.1481	–0.46
2	44.8	14.1000	14.2534	0.0546	–0.1534	–0.48
3	45.8	14.0000	14.1865	0.0586	–0.1865	–0.58
4	30.9	15.6000	15.1827	0.0414	0.4173	1.29
5	31.7	15.6000	15.1292	0.0391	0.4708	1.45
96	36.2	14.8000	14.8284	0.0327	–0.0284	–0.09
97	34.2	14.6000	14.9621	0.0339	–0.3621	–1.12
98	33.2	14.5000	15.0289	0.0355	–0.5289	–1.63
99	39.2	14.7000	14.6278	0.0363	0.0722	0.22
100	36.4	14.3000	14.8150	0.0327	–0.5150	–1.59

INSTRUCTIONS

Proceed with the three steps of regression analysis as described on page 642. After specifying the **Response** and **Predictors**, click **Results. . .**, and **In addition, the full table of fits and residuals**.

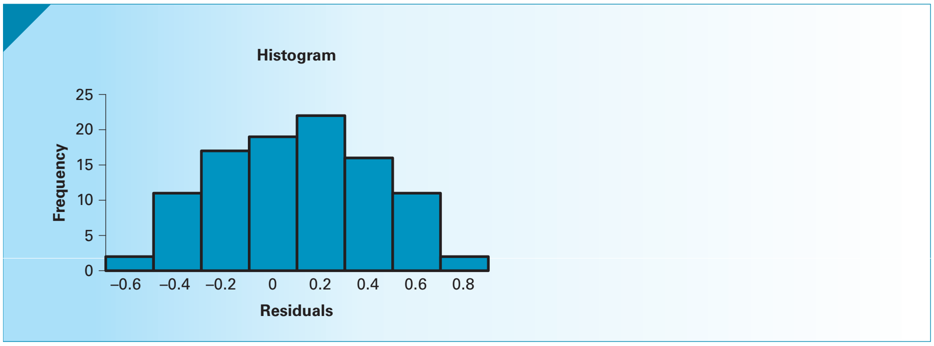
The predicted values, residuals, and standardized residuals will be printed.

An analysis of the residuals will allow us to determine whether the error variable is nonnormal, whether the error variance is constant, and whether the errors are independent. We begin with nonnormality.

Nonnormality

As we've done throughout this book, we check for normality by drawing the histogram of the residuals. Figure 16.9 is Excel's version (Minitab's is similar). As you can see, the histogram is bell shaped, leading us to believe that the error is normally distributed.

FIGURE 16.9 Histogram of Residuals for Example 16.2



Heteroscedasticity

The variance of the error variable σ_ε^2 is required to be constant. When this requirement is violated, the condition is called **heteroscedasticity**. (You can impress friends and relatives by using this term. If you can't pronounce it, try **homoscedasticity**, which refers to the condition where the requirement is satisfied.) One method of diagnosing heteroscedasticity is to plot the residuals against the predicted values of y. We then look for a change in the spread of the plotted points.* Figure 16.10 describes such a situation. Notice that in this illustration, σ_ε^2 appears to be small when $\hat{y}$ is small and large when $\hat{y}$ is large. Of course, many other patterns could be used to depict this problem.

FIGURE 16.10 Plot of Residuals Depicting Heteroscedasticity

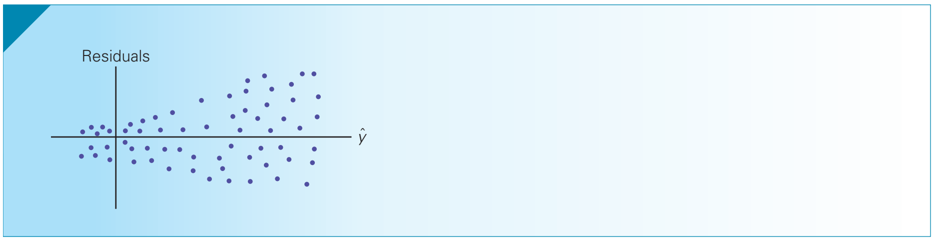
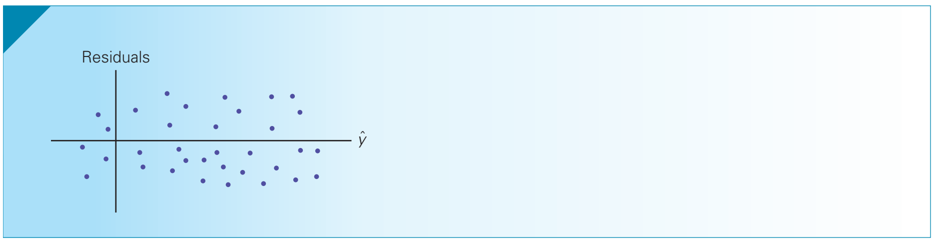


Figure 16.11 illustrates a case in which σ_ε^2 is constant. As a result, there is no apparent change in the variation of the residuals.

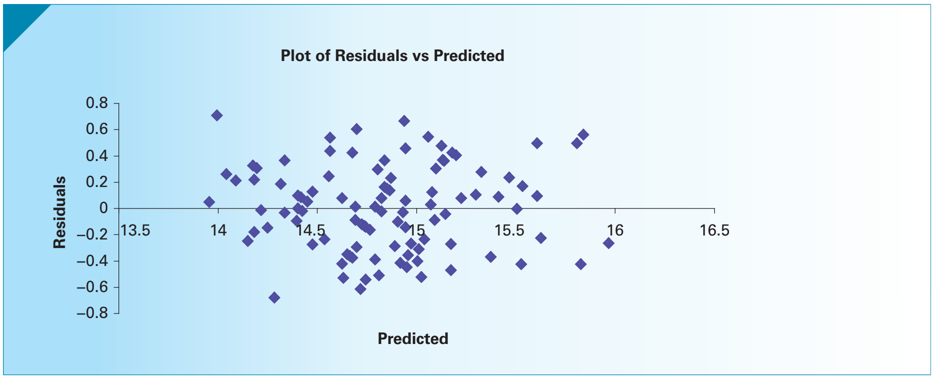
FIGURE 16.11 Plot of Residuals Depicting Homoscedasticity



Excel's plot of the residuals versus the predicted values of y for Example 16.2 is shown in Figure 16.12. There is no sign of heteroscedasticity.

*Keller's website Appendix Szroeter's Test describes a test for heteroscedasticity.

FIGURE **16.12** Plot of Predicted Values versus Residuals for Example 16.2



Nonindependence of the Error Variable

In Chapter 3, we briefly described the difference between cross-sectional and time-series data. Cross-sectional data are observations made at approximately the same time, whereas a time series is a set of observations taken at successive points of time. The data in Example 16.2 are cross-sectional because all of the prices and odometer readings were taken at about the same time. If we were to observe the auction price of cars every week for, say, a year, that would constitute a time series.

Condition 4 states that the values of the error variable are independent. When the data are time series, the errors often are correlated. Error terms that are correlated over time are said to be **autocorrelated** or **serially correlated**. For example, suppose that, in an analysis of the relationship between annual gross profits and some independent variable, we observe the gross profits for the years 1991 to 2010. The observed values of y are denoted $y_1, y_2, \ldots y_{20}$, where y_1 is the gross profit for 1991, y_2 is the gross profit for 1992, and so on. If we label the residuals $e_1, e_2, \ldots, e_{20}$, then—if the independence requirement is satisfied—there should be no relationship among the residuals. However, if the residuals are related it is likely that autocorrelation exists.

We can often detect autocorrelation by graphing the residuals against the time periods. If a pattern emerges, it is likely that the independence requirement is violated. Figures 16.13 (alternating positive and negative residuals) and 16.14 (increasing residuals) exhibit patterns indicating autocorrelation. (Notice that we joined the points to make it easier to see the patterns.) Figure 16.15 shows no pattern (the residuals appear to be randomly distributed over the time periods) and thus likely represent the occurrence of independent errors.

In Chapter 17, we introduce the Durbin-Watson test, which is another statistical test to determine whether one form of autocorrelation is present.

FIGURE **16.13** Plot of Residuals versus Time Indicating Autocorrelation (Alternating)

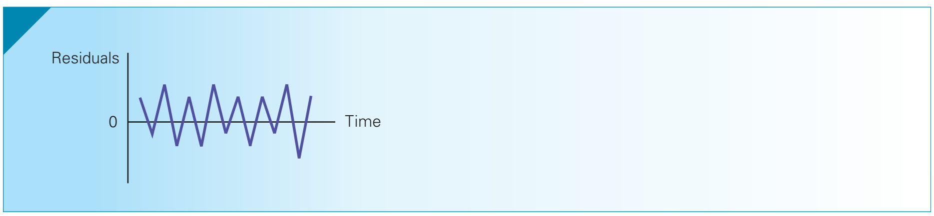
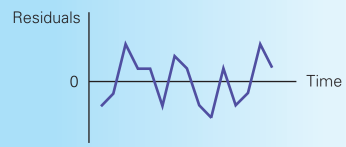


FIGURE **16.14** Plot of Residuals versus Time Indicating Autocorrelation (Increasing)

FIGURE **16.15** Plot of Residuals versus Time Indicating Independence



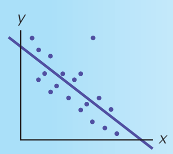
Outliers

An **outlier** is an observation that is unusually small or unusually large. To illustrate, consider Example 16.2, where the range of odometer readings was 19.1 to 49.2 thousand miles. If we had observed a value of 5,000 miles, we would identify that point as an outlier. We need to investigate several possibilities.

1. There was an error in recording the value. To detect an error, we would check the point or points in question. In Example 16.2, we could check the car's odometer to determine whether a mistake was made. If so, we would correct it before proceeding with the regression analysis.
2. The point should not have been included in the sample. Occasionally, measurements are taken from experimental units that do not belong with the sample. We can check to ensure that the car with the 5,000-mile odometer reading was actually 3 years old. We should also investigate the possibility that the odometer was rolled back. In either case, the outlier should be discarded.
3. The observation was simply an unusually large or small value that belongs to the sample and that was recorded properly. In this case, we would do nothing to the outlier. It would be judged to be valid.

Outliers can be identified from the scatter diagram. Figure 16.16 depicts a scatter diagram with one outlier. The statistics practitioner should check to determine whether the measurement was recorded accurately and whether the experimental unit should be included in the sample.

FIGURE **16.16** Scatter Diagram with One Outlier



The standardized residuals also can be helpful in identifying outliers. Large absolute values of the standardized residuals should be thoroughly investigated. Minitab automatically reports standardized residuals that are less than −2 and greater than 2.

Influential Observations

Occasionally, in a regression analysis, one or more observations have a large influence on the statistics. Figure 16.17 describes such an observation and the resulting least squares line. If the point had not been included, the least squares line in Figure 16.18 would have been produced. Obviously, one point has had an enormous influence on the results. Influential points can be identified by the scatter diagram. The point may be an outlier and as such must be investigated thoroughly. Minitab also identities influential observations.

FIGURE **16.17** **Scatter Diagram with One Influential Observation**

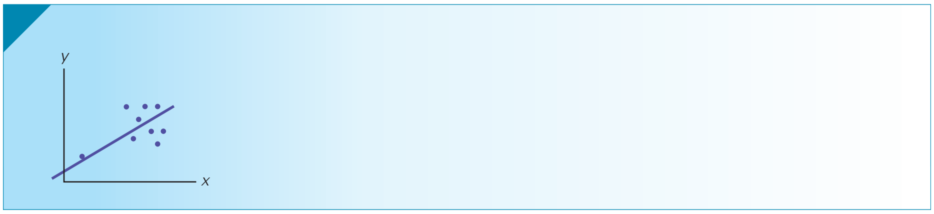
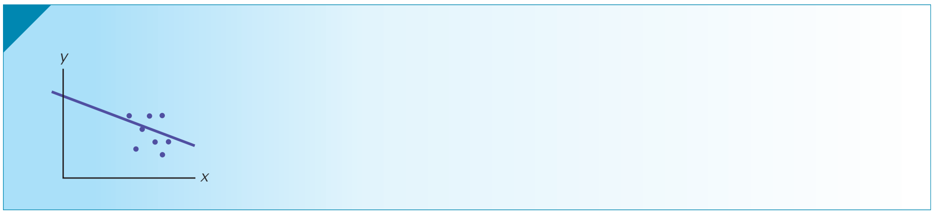


FIGURE **16.18** **Scatter Diagram without the Influential Observation**



Procedure for Regression Diagnostics

The order of the material presented in this chapter is dictated by pedagogical requirements. Consequently, we presented the least squares method of assessing the model's fit, predicting and estimating using the regression equation, coefficient of correlation, and finally, the regression diagnostics. In a practical application, the regression diagnostics would be conducted earlier in the process. It is appropriate to investigate violations of the required conditions when the model is assessed and before using the regression equation to predict and estimate. The following steps describe the entire process. (In Chapter 18, we will discuss model building, for which the following steps represent only a part of the entire procedure.)

1. Develop a model that has a theoretical basis; that is, for the dependent variable in question, find an independent variable that you believe is linearly related to it.
2. Gather data for the two variables. Ideally, conduct a controlled experiment. If that is not possible, collect observational data.

3. Draw the scatter diagram to determine whether a linear model appears to be appropriate. Identify possible outliers.

4. Determine the regression equation.

5. Calculate the residuals and check the required conditions:

   Is the error variable nonnormal?
   Is the variance constant?
   Are the errors independent?
   Check the outliers and influential observations.

6. Assess the model's fit.

   Compute the standard error of estimate.
   Test to determine whether there is a linear relationship. (Test β_1 or ρ.)
   Compute the coefficient of determination.

7. If the model fits the data, use the regression equation to predict a particular value of the dependent variable or estimate its mean (or both).

EXERCISES

16.81 You are given the following six points:

x	−5	−2	0	3	4	7
y	15	9	7	6	4	1

a. Determine the regression equation.
b. Use the regression equation to determine the predicted values of y.
c. Use the predicted and actual values of y to calculate the residuals.
d. Compute the standardized residuals.
e. Identify possible outliers.

16.82 Refer to Exercise 16.2. Calculate the residuals and the predicted values of y.

16.83 Calculate the residuals and predicted values of y in Exercise 16.3.

16.84 Refer to Exercise 16.4.
a. Calculate the residuals
b. Calculate the predicted values of y.
c. Plot the residuals (on the vertical axis) and the predicted values of y.

16.85 Calculate and plot the residuals and predicted values of y for Exercise 16.5.

The following exercises require the use of a computer and software.

16.86 Refer to Exercise 16.6.
a. Determine the residuals and the standardized residuals.
b. Draw the histogram of the residuals. Does it appear that the errors are normally distributed? Explain.
c. Identify possible outliers.
d. Plot the residuals versus the predicted values of y. Does it appear that heteroscedasticity is a problem? Explain.

16.87 Refer to Exercise 16.7.
a. Does it appear that the errors are normally distributed? Explain.
b. Does it appear that heteroscedasticity is a problem? Explain.

16.88 Are the required conditions satisfied in Exercise 16.8?

16.89 Refer to Exercise 16.9.
a. Determine the residuals and the standardized residuals.
b. Draw the histogram of the residuals. Does it appear that the errors are normally distributed? Explain.
c. Identify possible outliers.
d. Plot the residuals versus the predicted values of y. Does it appear that heteroscedasticity is a problem? Explain.

16.90 Refer to Exercise 16.10. Are the required conditions satisfied?

16.91 Refer to Exercise 16.11.

a. Determine the residuals and the standardized residuals.

b. Draw the histogram of the residuals. Does it appear that the errors are normally distributed? Explain.

c. Identify possible outliers.

d. Plot the residuals versus the predicted values of y. Does it appear that heteroscedasticity is a problem? Explain.

16.92 Check the required conditions for Exercise 16.12.

16.93 Refer to Exercise 16.13. Are the required conditions satisfied?

16.94 Refer to Exercise 16.14.

a. Determine the residuals and the standardized residuals.

b. Draw the histogram of the residuals. Does it appear that the errors are normally distributed? Explain.

c. Identify possible outliers.

d. Plot the residuals versus the predicted values of y. Does it appear that heteroscedasticity is a problem? Explain.

16.95 Are the required conditions satisfied for Exercise 16.15?

16.96 Check to ensure that the required conditions for Exercise 16.16 are satisfied.

16.97 Are the required conditions satisfied for Exercise 16.17?

16.98 Perform a complete diagnostic analysis for Exercise 16.18 to determine whether the required conditions are satisfied.

CHAPTER SUMMARY

Simple linear regression and correlation are techniques for analyzing the relationship between two interval variables. Regression analysis assumes that the two variables are linearly related. The least squares method produces estimates of the intercept and the slope of the regression line. Considerable effort is expended in assessing how well the linear model fits the data. We calculate the standard error of estimate, which is an estimate of the standard deviation of the error variable. We test the slope to determine whether there is sufficient evidence of a linear relationship. The strength of the linear association is measured by the coefficient of determination. When the model provides a good fit, we can use it to predict the particular value and to estimate the expected value of the dependent variable. We can also use the Pearson correlation coefficient to measure and test the relationship between two bivariate normally distributed variables. We completed this chapter with a discussion of how to diagnose violations of the required conditions.

IMPORTANT TERMS

Regression analysis 634
Dependent variable 634
Independent variable 634
Deterministic model 635
Probabilistic model 635
Error variable 636
First-order linear model 636
Simple linear regression model 636
Least squares method 637
Residuals 639
Sum of squares for error 639
Standard error of estimate 650
Coefficient of determination 656
Confidence interval estimate of the expected value of y 667
Pearson coefficient of correlation 660
Point prediction 666
Prediction interval 666
Heteroscedasticity 674
Homoscedasticity 674
Autocorrelation 675
Serial correlation 675

SYMBOLS

Symbol	Pronounced	Represents
β_0	Beta sub zero or beta zero	y-intercept coefficient
β_1	Beta sub one or beta one	Slope coefficient
ε	Epsilon	Error variable
$\hat{y}$	y hat	Fitted or calculated value of y
b_0	b sub zero or b zero	Sample y-intercept coefficient
b_1	b sub one or b one	Sample slope coefficient
σ_ε	Sigma sub epsilon or sigma epsilon	Standard deviation of error variable
s_ε	s sub epsilon or s epsilon	Standard error of estimate
s_{b_1}	s sub b sub one or s b one	Standard error of b_1
R^2	R squared	Coefficient of determination
x_g	x sub g or x g	Given value of x
ρ	Rho	Pearson coefficient of correlation
r		Sample coefficient of correlation
e_i	e sub i or e i	Residual of ith point

FORMULAS

Sample slope

$$b_1 = \frac{s_{xy}}{s_x^2}$$

Sample y-intercept

$$b_0 = \bar{y} - b_1\bar{x}$$

Sum of squares for error

$$\text{SSE} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Standard error of estimate

$$s_\varepsilon = \sqrt{\frac{\text{SSE}}{n-2}}$$

Test statistic for the slope

$$t = \frac{b_1 - \beta_1}{s_{b_1}}$$

Standard error of b_1

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{(n-1)s_x^2}}$$

Coefficient of determination

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = 1 - \frac{\text{SSE}}{\sum(y_i - \bar{y})^2}$$

Prediction interval

$$\hat{y} \pm t_{\alpha/2,n-2}s_\varepsilon\sqrt{1 + \frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$$

Confidence interval estimator of the expected value of y

$$\hat{y} \pm t_{\alpha/2,n-2}s_\varepsilon\sqrt{\frac{1}{n} + \frac{(x_g - \bar{x})^2}{(n-1)s_x^2}}$$

Sample coefficient of correlation

$$r = \frac{s_{xy}}{s_x s_y}$$

Test statistic for testing $\rho = 0$

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

COMPUTER OUTPUT AND INSTRUCTIONS

Technique	Excel	Minitab
Regression	642	643
Correlation	662	662
Prediction interval	669	669
Regression diagnostics	672	673

CHAPTER EXERCISES

The following exercises require the use of a computer and software. The answers to some of the questions may be calculated manually. See Appendix A for the sample statistics. Conduct all tests of hypotheses at the 5% significance level.

16.99 Xr16-99 The manager of Colonial Furniture has been reviewing weekly advertising expenditures. During the past 6 months, all advertisements for the store have appeared in the local newspaper. The number of ads per week has varied from one to seven. The store's sales staff has been tracking the number of customers who enter the store each week. The number of ads and the number of customers per week for the past 26 weeks were recorded.

a. Determine the sample regression line.
b. Interpret the coefficients.
c. Can the manager infer that the larger the number of ads, the larger the number of customers?
d. Find and interpret the coefficient of determination.
e. In your opinion, is it a worthwhile exercise to use the regression equation to predict the number of customers who will enter the store, given that Colonial intends to advertise five times in the newspaper? If so, find a 95% prediction interval. If not, explain why not.

16.100 Xr16-100 The president of a company that manufactures car seats has been concerned about the number and cost of machine breakdowns. The problem is that the machines are old and becoming quite unreliable. However, the cost of replacing them is quite high, and the president is not certain that the cost can be made up in today's slow economy. To help make a decision about replacement, he gathered data about last month's costs for repairs and the ages (in months) of the plant's 20 welding machines.

a. Find the sample regression line.
b. Interpret the coefficients.
c. Determine the coefficient of determination, and discuss what this statistic tells you.
d. Conduct a test to determine whether the age of a machine and its monthly cost of repair are linearly related.
e. Is the fit of the simple linear model good enough to allow the president to predict the monthly repair cost of a welding machine that is 120 months old? If so, find a 95% prediction interval. If not, explain why not.

16.101 Xr16-101 An agronomist wanted to investigate the factors that determine crop yield. Accordingly, she undertook an experiment wherein a farm was divided into 30 1-acre plots. The amount of fertilizer applied to each plot was varied. Corn was then planted, and the amount of corn harvested at the end of the season was recorded.

a. Find the sample regression line and interpret the coefficients.
b. Can the agronomist conclude that there is a linear relationship between the amount of fertilizer and the crop yield?
c. Find the coefficient of determination and interpret its value.
d. Does the simple linear model appear to be a useful tool in predicting crop yield from the amount of fertilizer applied? If so, produce a 95% prediction interval of the crop yield when 300 pounds of fertilizer are applied. If not, explain why not.

16.102 Xr16-102 Every year, the United States Federal Trade Commission rates cigarette brands according to their levels of tar and nicotine, substances that are hazardous to smokers' health. In addition, the commission includes the amount of carbon monoxide, which is a by-product of burning tobacco that seriously affects the heart. A random sample of 25 brands was taken.

a. Are the levels of tar and nicotine linearly related?
b. Are the levels of nicotine and carbon monoxide linearly related?

16.103 Xr16-103 Some critics of television complain that the amount of violence shown on television contributes to violence in our society. Others point out that television also contributes to the high level of obesity among children. We may have to add financial problems to the list. A sociologist theorized that people who watch television frequently are exposed to many commercials, which in turn leads them to buy more, finally resulting in increasing debt. To test this belief, a sample of 430 families was drawn. For each, the total debt and the number of hours the television is turned on per week were recorded. Perform a statistical procedure to help test the theory.

16.104 Xr16-104 The analysis the human resources manager performed in Exercise 16.18 indicated that the dexterity test is not a predictor of job performance. However, before discontinuing the test he decided that the problem is that the statistical analysis was flawed because it examined the relationship between test score and job performance only for those who scored well in the test. (Recall that only those who scored above 70 were hired; applicants who achieved scores below 70 were not hired.) The manager decided to perform another statistical analysis. A sample of 50 job applicants who scored above 50 were hired; as before, the workers' performance was measured. The test scores and percentages of nondefective computers produced were recorded. On the basis of these data, should the manager discontinue the dexterity tests?

16.105 Xr16-105 Mutual funds minimize risks by diversifying the investments they make. There are mutual funds that specialize in particular types of investments. For example, the TD Precious Metal Mutual Fund buys shares in gold-mining companies. The value of this mutual fund depends on a number of factors related to the companies in which the fund invests as well as on the price of gold. To investigate the relationship between the value of the fund and the price of gold, an MBA student gathered the daily fund price and the daily price of gold for a 28-day period. Can we infer from these data that there is a positive linear relationship between the value of the fund and the price of gold? (The author is grateful to Jim Wheat for writing this exercise.)

16.106 Xr03-59 (Exercise 3.59 revisited) A very large contribution to profits for a movie theater is the sale of popcorn, soft drinks, and candy. A movie theater manager speculated that the longer the time between showings of a movie, the greater the sales of concessions. To acquire more information, the manager conducted an experiment. For a month, he varied the amount of time between movie showings and calculated the sales. Can the manager conclude that when the times between movies increase so do sales?

16.107 Xr16-107* A computer dating service typically asks for various pieces of information such as height, weight, and income. One such service requested the length of index fingers. The only plausible reason for this request is to act as a proxy on height. Women have often complained that men lie about their heights. If there is a strong relationship between heights and index fingers, the information can be used to "correct" false claims about heights. To test the relationship between the two variables, researchers gathered the heights and lengths of index fingers (in centimeters) of 121 students.

a. Graph the relationship between the two variables.
b. Is there sufficient evidence to infer that height and length of index fingers are linearly related?
c. Predict with 95% confidence the height of someone whose index finger is 6.5 cm long. Is this prediction likely to be useful? Explain. (The author would like to thank Howard Waner for supplying the problem and data.)

The following exercises employ data files associated with two previous exercises.

16.108 Xr12-31* In addition to the data recorded for Exercises 12.31 and 13.153, we recorded the grade point average of the students who held down part-time jobs. Determine whether there is evidence of a linear relationship between the hours spent at part-time jobs and the grade point averages.

16.109 Xr13-19* Exercise 13.19 described a survey that asked people between 18 and 34 years of age and 35 to 50 years of age how much time they spent listening to FM radio each day. Also recorded were the amounts spent on music throughout the year. Can we infer that a linear relationship exists between listening times and amounts spent on music?

CASE 16.1 Insurance Compensation for Lost Revenues‡


© Spencer Grant/PhotoEdit

DATA C16-01

In July 1990, a rock-and-roll museum opened in Atlanta, Georgia. The museum was located in a large city block containing a variety of stores. In late July 1992, a fire that started in one of these stores burned the entire block, including the museum. Fortunately, the museum had taken out insurance to cover the cost of rebuilding as well as lost revenue. As a general rule, insurance companies base their payment on how well the company performed in the past. However, the owners of the museum argued that the revenues were increasing, and hence they were entitled to more money under their insurance plan. The argument was based on the revenues and attendance figures of an amusement park, featuring rides and other similar attractions that had opened nearby. The amusement park opened in December 1991. The two entertainment facilities were operating jointly during the last 4 weeks of 1991 and the first 28 weeks of 1992 (the point at which the fire destroyed the museum). In April 1995, the museum reopened with considerably more features than the original one.

The attendance figures for both facilities for December 1991 to October 1995 are listed in columns 1 (museum) and 2 (amusement park). During the period when the museum was closed, the data show zero attendance.

The owners of the museum argued that the weekly attendance from the 29th week of 1992 to the 16th week of 1995 should be estimated using the most current data (17th to 42nd week of 1995). The insurance company argued that the estimates should be based on the 4 weeks of 1991 and the 28 weeks of 1992, when both facilities were operating and before the museum reopened with more features than the original museum.

a. Estimate the coefficients of the simple regression model based on the insurance company's argument. In other words, use the attendance figures for the last 4 weeks in 1991 and the next 28 weeks in 1992 to estimate the coefficients. Then use the model to calculate point predictions for the museum's weekly attendance figures when the museum was closed. Calculate the predicted total attendance.

b. Repeat part (a) using the museum's argument—that is, use the attendance figures after the reopening in 1995 to estimate the regression coefficients and use the equation to predict the weekly attendance when the museum was closed. Calculate the total attendance that was lost because of the fire.

c. Write a report to the insurance company discussing this analysis and include your recommendation about how much the insurance company should award the museum?

[†]The case and the data are real. The names have been changed to preserve anonymity. The author wishes to thank Dr. Kevin Leonard for supplying the problem and the data.

CASE 16.2 Predicting University Grades from High School Grades[§]

© Andresr/Shutterstock

DATA C16-02

Ontario high school students must complete a minimum of six Ontario Academic Credits (OACs) to gain admission to a university in the province. Most students take more than six OACs because universities take the average of the best six in deciding which students to admit. Most programs at universities require high school students to select certain courses. For example, science programs require two of chemistry, biology, and physics. Students applying to engineering must complete at least two mathematics OACs as well as physics. In recent years, one business program began an examination of all aspects of its program, including the criteria used to admit students. Students are required to take English and calculus OACs, and the minimum high school average is about 85%. Strangely enough, even though students are required to complete English and calculus, the marks in these subjects are not included in the average unless they are in the top six courses in a student's transcript. To examine the issue, the registrar took a random sample of students who recently graduated with the BBA (bachelor of business administration degree). He recorded the university GPA (range 0 to 12), the high school average based on the best six courses, and the high school average using English and calculus and the four next best marks.

a. Is there a relationship between university grades and high school average using the best six OACs?

b. Is there a relationship between university grades and high school average using the best four OACs plus calculus and English?

c. Write a report to the university's academic vice president describing your statistical analysis and your recommendations.

[§]The author is grateful to Leslie Grauer for her help in gathering the data for this case.

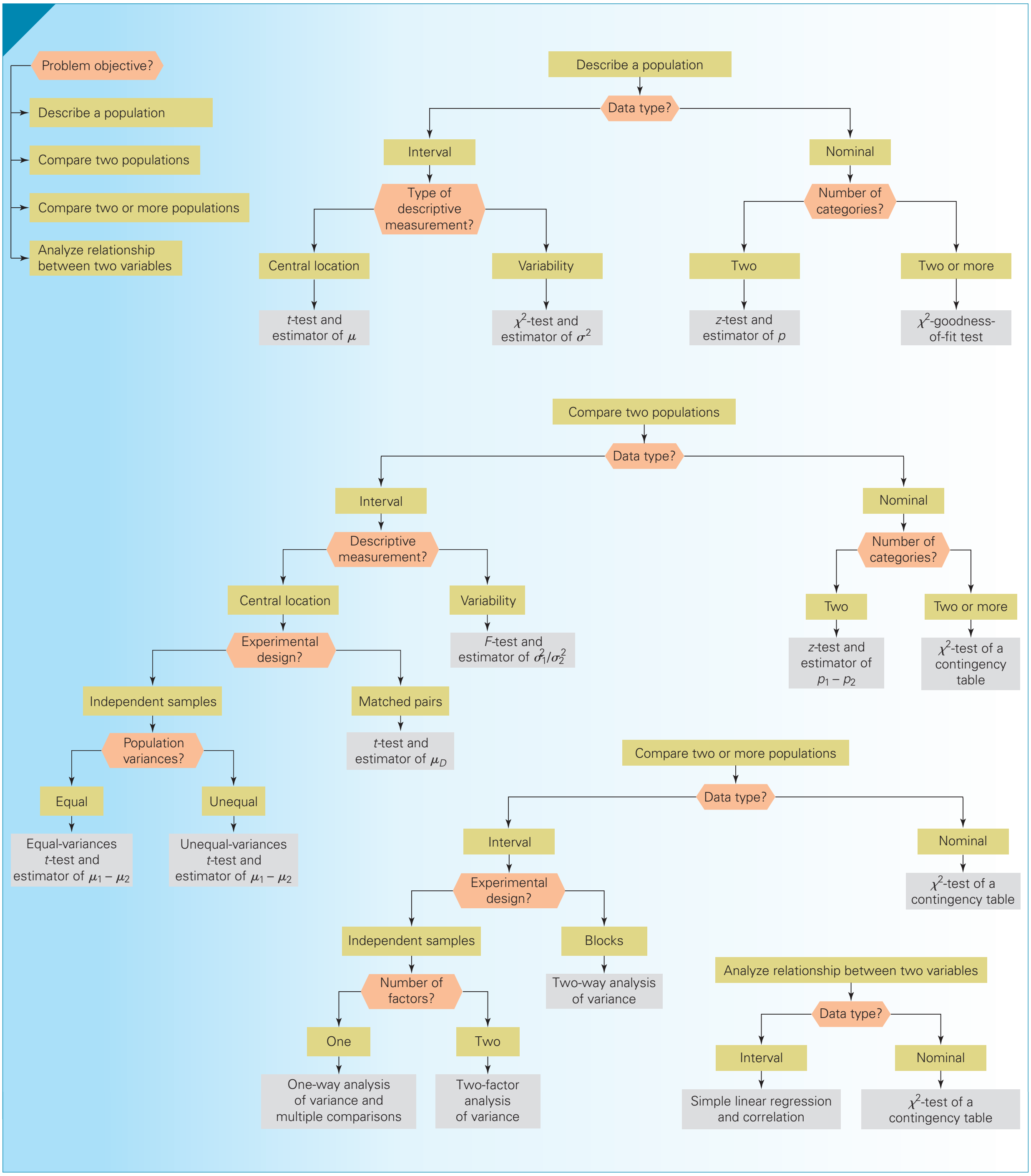
APPENDIX 16 / REVIEW OF CHAPTERS 12 TO 16

We have now presented two dozen inferential techniques. Undoubtedly, the task of choosing the appropriate technique is growing more difficult. Table A16.1 lists all the statistical inference methods covered since Chapter 12. Figure A16.1 is a flowchart to help you choose the correct technique.

TABLE **A16.1** Summary of Statistical Techniques in Chapters 12 to 16

t-test of μ
Estimator of μ (including estimator of $N\mu$)
χ^2-test of σ^2
Estimator of σ^2
z-test of p
Estimator of p (including estimator of Np)
Equal-variances t-test of $\mu_1 - \mu_2$
Equal-variances estimator of $\mu_1 - \mu_2$
Unequal-variances t-test of $\mu_1 - \mu_2$
Unequal-variances estimator of $\mu_1 - \mu_2$
t-test of μ_D
Estimator of μ_D
F-test of σ_1^2/σ_2^2
Estimator of σ_1^2/σ_2^2
z-test of $p_1 - p_2$ (Case 1)
z-test of $p_1 - p_2$ (Case 2)
Estimator of $p_1 - p_2$
One-way analysis of variance (including multiple comparisons)
Two-way (randomized blocks) analysis of variance
Two-factor analysis of variance
χ^2-goodness-of-fit test
χ^2-test of a contingency table
Simple linear regression and correlation (including t-tests of β_1 and ρ, and prediction and confidence intervals)

FIGURE A16.1 Flowchart of Techniques in Chapters 12 to 16



EXERCISES

A16.1 XrA16-01 In the last decade, society in general and the judicial system in particular have altered their opinions on the seriousness of drunken driving. In most jurisdictions, driving an automobile with a blood alcohol level in excess of .08 is a felony. Because of a number of factors, it is difficult to provide guidelines for when it is safe for someone who has consumed alcohol to drive a car. In an experiment to examine the relationship between blood alcohol level and the weight of a drinker, 50 men of varying weights were each given three beers to drink, and 1 hour later their blood alcohol levels were measured. If we assume that the two variables are normally distributed, can we conclude that blood alcohol level and weight are related?

A16.2 XrA16-02 An article in the journal *Appetite* (December 2003) described an experiment to determine the effect that breakfast meals have on school children. A sample of 29 children was tested on four successive days, having a different breakfast each day. The breakfast meals were

1. Cereal (Cheerios)
2. Cereal (Shreddies)
3. A glucose drink
4. No breakfast

The order of breakfast meals was randomly assigned. A computerized test of working memory was conducted prior to breakfast and again 2 hours later. The decrease in scores was recorded. Do these data allow us to infer that there are differences in the decrease depending on the type of breakfast?

A16.3 Do cell phones cause cancer? This is a multibillion-dollar question. Currently, dozens of lawsuits are pending that claim cell phone use has caused cancer. To help shed light on the issue, several scientific research projects have been undertaken. One such project was conducted by Danish researchers (*Source*: Journal of the National Cancer Institute, 2001). The 13-year study examined 420,000 Danish cell phone users. The scientists determined the number of Danes who would be expected to contract various forms of cancer. The expected number and the actual number of cell phone users who developed each type of cancer are listed here.

Cancer	Expected Number	Actual Number
Brain and nervous system	143	135
Salivary glands	9	7
Leukemia	80	77
Pharynx	52	32
Esophagus	57	42
Eye	12	8
Thyroid	13	13

a. Can we infer from these data that there is a relationship between cell phone use and cancer?

b. Discuss the results, including whether the data are observational or experimental. Provide several interpretations of the statistics. In particular, indicate whether you can infer that cell phone use causes cancer.

A16.4 XrA16-04 A new antiflu vaccine designed to reduce the duration of symptoms has been developed. However, the effect of the drug varies from person to person. To examine the effect of age on the effectiveness of the drug, a sample of 140 flu sufferers was drawn. Each person reported how long the symptoms of the flu persisted and his or her age. Do these data provide sufficient evidence to infer that the older the patient, the longer it takes for the symptoms to disappear?

A16.5 XrA16-05 Several years ago we heard about the "Mommy Track," the phenomenon of women being underpaid in the corporate world because of what is seen as their divided loyalties between home and office. There may also be a "Daddy Differential," which refers to the situation where men whose wives stay at home earn more than men whose wives work. It is argued that the differential occurs because bosses reward their male employees if they come from "traditional families." Linda Stroh of Loyola University of Chicago studied a random sample of 348 male managers employed by 20 *Fortune 500* companies. Each manager reported whether his wife stayed at home to care for their children or worked outside the home, and his annual income. The incomes (in thousands of dollars) were recorded. The incomes of the managers whose wives stay at home are stored in column 1. Column 2 contains the incomes of managers whose wives work outside the home.

a. Can we conclude that men whose wives stay at home earn more than men whose wives work outside the home?
b. If your answer in part (a) is affirmative, does this establish a case for discrimination? Can you think of another cause-and-effect scenario? Explain.

A16.6 XrA16-06 There are enormous differences between health-care systems in the United States and Canada. In a study to examine one dimension of these differences, 300 heart attack victims in each country were randomly selected. (Results of the study conducted by Dr. Daniel Mark of Duke University Medical Center, Dr. David Naylor of Sunnybrook Hospital in Toronto, and Dr. Paul Armstrong of the University of Alberta were published in the *Toronto Sun*, October 27, 1994.) Each patient was asked the following questions regarding the effect of his or her treatment:

1. How many days did it take you to return to work?
2. Do you still have chest pain? (This question was asked 1 month, 6 months, and 12 months after the patients' heart attacks.)

The responses were recorded in the following way:

Column 1: Code representing nationality: 1 = U.S.; 2 = Canada
Column 2: Responses to question 1
Column 3: Responses to question 2–1 month after heart attack: 2 = yes; 1 = no
Column 4: Responses to question 2–6 months after heart attack: 2 = yes; 1 = no
Column 5: Responses to question 2–12 months after heart attack: 2 = yes; 1 = no

Can we conclude that recovery is faster in the United States?

A16.7 XrA16-07 Betting on the results of National Football League games is a popular North American activity. In some states and provinces, it is legal to do so provided that wagers are made through a government-authorized betting organization. In the province of Ontario, Pro-Line serves that function. Bettors can choose any team on which to wager, and Pro-Line sets the odds, which determine the winning payoffs. It is also possible to bet that in any game a tie will be the result. (A tie is defined as a game in which the winning margin is 3 or fewer points. A win occurs when the winning margin is greater than 3.) To assist bettors, Pro-Line lists the favorite for each game and predicts the point spread between the two teams. To judge how well Pro-Line predicts outcomes, the Creative Statistics Company tracked the results of a recent season. It recorded whether a team was favored by (1) 3 or fewer points, (2) 3.5 to 7 points, (3) 7.5 to 11 points, or (4) 11.5 or more points. It also recorded whether the favored team (1) won, (2) lost, or (3) tied. These data are recorded in columns 1 (Pro-Line's predictions) and 2 (game results). Can we conclude that Pro-Line's forecasts are useful for bettors?

A16.8 XrA16-08 As all baseball fans know, first base is the only base that the base runner may overrun. At both second and third base, the runner may be tagged out if he runs past the base. Consequently, on close plays at second and third base, the runner will slide, enabling him to stop at the base. In recent years, however, several players have chosen to slide headfirst when approaching first base, claiming that this is faster than simply running over the base. In an experiment to test this claim, 25 players on one National League team were recruited. Each player ran to first base with and without sliding, and the times to reach the base were recorded. Can we conclude that sliding is slower than not sliding?

A16.9 XrA16-09 How does mental outlook affect a person's health? The answer to this question may allow physicians to care more effectively for their patients. In an experiment to examine the relationship between attitude and physical health, Dr. Daniel Mark, a heart specialist at Duke University, studied 1,719 men and women who had recently undergone a heart catheterization, a procedure that checks for clogged arteries. Patients undergo this procedure when heart disease results in chest pain. All of the patients in the experiment were in about the same condition. In interviews, 14% of the patients doubted that they would recover sufficiently to resume their daily routines. Dr. Mark identified these individuals as pessimists; the others were (by default) optimists. After one year, Dr. Mark recorded how many patients were still alive. The data are stored in columns 1 (1 = optimist, 2 = pessimist) and 2 (2 = alive, 1 = dead). Do these data allow us to infer that pessimists are less likely to survive than optimists with similar physical ailments?

A16.10 XrA16-10 Physicians have been recommending more exercise for their patients, particularly those who are overweight. One benefit of regular exercise appears to be a reduction in cholesterol, a substance associated with heart disease. To study the relationship more carefully, a physician took a random sample of 50 patients who do not exercise and measured their cholesterol levels. He then started them on regular exercise programs. After 4 months, he asked each patient how many minutes per week (on average) he or she exercised; he also measured their cholesterol levels. Column 1 = weekly exercise in minutes, column 2 = cholesterol level before exercise program, and column 3 = cholesterol level after exercise program.

a. Do these data allow us to infer that the amount of exercise and the reduction in cholesterol levels are related?
b. Produce a 95% interval of the amount of cholesterol reduction for someone who exercises for 100 minutes per week.
c. Produce a 95% interval for the average cholesterol reduction for people who exercise for 120 minutes per week.

A16.11 XrA16-11 An economist working for a state university wanted to acquire information about salaries in publicly funded and private colleges and universities. She conducted a survey of 623 public-university faculty members and 592 private-university faculty members asking each to report his or her rank (instructor = 1, assistant professor = 2, associate professor = 3, and professor = 4) and current salary ($1,000). (Adapted from the American Association of University Professors, *AAUP Annual Report on the Economic Status of the Profession.)*

a. Conduct a test to determine whether public colleges and universities and private colleges and universities pay different salaries when all ranks are combined.
b. For each rank, determine whether there is enough evidence to infer that the private college and university salaries differ from that of publicly funded colleges and universities.
c. If the answers to parts (a) and (b) differ, suggest a cause.
d. Conduct a test to determine whether your suggested cause is valid.

A16.12 XrA16-12 Millions of people suffer from migraine headaches. The costs in work days lost, medication, and treatment are measured in the billions of dollars. A study reported in the *Journal of the American Medical Association* (2005, 203: 2118–2125) described an experiment that examined whether acupuncture is an effective procedure in treating migraines. A random sample of 302 migraine patients was selected and divided into three groups. Group 1 was treated with acupuncture; group 2 was treated with sham acupuncture (patients believed that they were being treated with acupuncture but were not); and group 3 was not treated at all. The number of headache days per month was recorded for each patient before the treatments began. The number of headache days per month after treatment was also measured.

a. Conduct a test to determine whether there are differences in the number of headache days before treatment between the three groups of patients.
b. Test to determine whether differences exist after treatment. If so, what are the differences?
c. Why was the test in part (a) conducted?

A16.13 XrA16-13 The battle between customers and car dealerships is often intense. Customers want the lowest price, and dealers want to extract as much money as possible. One source of conflict is the trade-in car. Most dealers will offer a relatively low trade-in in anticipation of negotiating the final package. In an effort to determine how dealers operate, a consumer organization undertook an experiment. Seventy-two individuals were recruited. Each solicited an offer on his or her 5-year-old Toyota Camry. The exact same car was used throughout the experiment. The only variables were the age and gender of the "owner." The ages were categorized as (1) young, (2) middle, and (3) senior. The cash offers are stored in columns 1 and 2. Column 1 stores the data for female owners, and column 2 contains the offers made to male owners. The first 12 rows in both columns represent the offers made to young people, the next 12 rows represent the middle group, and the last 12 rows represent the elderly owners.

a. Can we infer that differences exist between the six groups?
b. If differences exist, determine whether the differences are due to gender, age, or some interaction.

A16.14 XrA16-14 In the presidential elections of 2000 and 2004, the vote in the state of Florida was crucial. It is important for the political parties to track party affiliation. Surveys in Broward and Miami-Dade counties were conducted in 1990, 1996, 2000, and 2004. The numbers of Democrats, Republicans, and other voters were recorded for both counties and for all four years. Test each of the following.

a. Party affiliation changed over the four surveys in Broward.
b. Party affiliation changed over the four surveys in Miami-Dade.
c. There were differences between Broward and Miami-Dade in 2004.

A16.15 XrA16-15 Auto manufacturers are required to test their vehicles for a variety of pollutants in the exhaust. The amount of pollutant varies even among identical vehicles so that several vehicles must be tested. The engineer in charge of testing has collected data (in grams per kilometer driven) on the amounts of two pollutants—carbon monoxide and nitrous oxide—for 50 identical vehicles. The engineer believes the company can save money by testing for only one of the pollutants because the two pollutants are closely linked; that is, if a car is emitting a large amount of carbon monoxide, it will also emit a large amount of nitrous oxide. Do the data support the engineer's belief?

A16.16 XrA16-16 In 2003, there were 129,142,000 workers in the United States (*Source:* U.S. Census Bureau). The general manager for a public transportation company wanted to learn more about how workers commute to work and how long it takes them. A random sample of workers was interviewed. Each reported how he or she typically get to work and how long it takes. Estimate with 95% confidence the total amount of time spent commuting. (Data for this exercise were adapted from the *Statistical Abstract of the United States, 2006*, Table 1083.)

GENERAL SOCIAL SURVEY EXERCISES

A16.17 GSS2008* Is there sufficient evidence to conclude that less than 50% of Americans support gun laws (GUNLAW)?

A16.18 GSS2008* Can we infer from the data that Democrats and Republicans (PARTYID: 0, 1 = Democrat, 5, 6 = Republican) differ in their position on whether the government should reduce income differences between rich and poor (EQWLTH)?

A16.19 GSS2008* How does income affect a person's response to the question, Should the government improve the living conditions of poor people (HELPPOOR)? Test the relationship between income (INCOME) and (HELPPOOR) to answer the question.

A16.20 GSS2008* Do the data allow us to infer that households with at least one union member (UNION: 1 = Respondent belongs, 2 = Spouse belongs, 3 = Both belong, 4 = Neither belong) differ from households with no union members with respect to their position on whether the government should improve the standard of living of poor people (HELPPOOR)?

A16.21 GSS2008* Is there sufficient evidence to conclude that people who have taken college-level science courses (COLSCINM: 1 = Yes, 2 = No) are more likely to answer the following question correctly (HOTCORE): Is the center of Earth very hot? 1 = Yes, 2 = No. Correct answer: Yes.

A16.22 GSS2006* Do larger companies pay better than smaller companies? Answer the question by testing to determine whether there is enough evidence to infer that there is a positive linear relationship between income (INCOME) and the number of people working in the company (NUMORG).

A16.23 GSS2004* Test to determine whether people who went bankrupt in the previous year (FINAN1: 1 = Yes, 2 = No) differ in their political affiliation (PARTYID: 0, 1 = Democrat; 2, 3, 4 = Independent; 5, 6 = Republican)?

A16.24 GSS2008* It seems reasonable to assume that the more one works, the greater the income. Test this assumption by analyzing the relationship between hours worked per week (HRS) and income (INCOME).

A16.25 GSS2004* Is there enough evidence to conclude that victims of a robbery [LAW1: Were you a victim of a robbery (mugging or stick-up) in the previous year? 1 = Yes, 2 = No] are less likely to favor requiring a police permit to buy a gun (GUNLAW: 1 = Favor, 2 = Oppose)?

A16.26 GSS2008* Do you get a more prestigious occupation if you acquire more education? Analyze the relationship between occupation prestige score (PRESTG80) and education (EDUC) to answer the question.

AMERICAN NATIONAL ELECTION SURVEY EXERCISES

A16.27 ANES2008* Newspaper readership is down all over North America. Newspaper publishers need to acquire more information to stop this worrying trend. Do more educated people spend more time reading newspapers? Conduct a test to determine whether there is evidence to infer that more education (EDUC) is related to more time reading newspapers (TIME3).

A16.28 ANES2008* In many cities, the network national news is broadcast at 6:30 or 7:00 P.M. In most cities, the national news is preceded by local news in the late afternoon or early evening. Do most viewers watch both news shows? To help answer this question, test to determine whether the number of days watching national news (DAYS1) is related to the number of days watching local news in the late afternoon or early evening (DAYS2).

CASE A16.1 Nutrition Education Programs*

DATA CA16-01

Nutrition education programs, which teach clients how to lose weight or reduce cholesterol levels through better eating patterns, have been growing in popularity. The nurse in charge of one such program at a local hospital wanted to know whether the programs actually work. A random sample was drawn of 33 clients who attended a nutrition education program for those with elevated cholesterol levels. The study recorded the weight, cholesterol levels, total dietary fat intake per average day, total dietary cholesterol intake per average day, and percent of daily calories from fat. These data were gathered both before and 3 months after the program. The researchers also determined the clients' genders, ages, and heights. The data are stored in the following way:

Column 1: Gender (1 = female, 2 = male)
Column 2: Age
Column 3: Height (in meters)
Columns 4 and 5: Weight, before and after (in kilograms)
Columns 6 and 7: Cholesterol level, before and after
Columns 8 and 9: Total dietary fat intake per average day, before and after (in grams)
Columns 10 and 11: Dietary cholesterol intake per average day, before and after (in milligrams)
Columns 12 and 13: Percent daily calories from fat, before and after

The nurse would like the following information:

a. In terms of each of weight, cholesterol level, fat intake, cholesterol intake, and calories from fat, is the program a success?
b. Does gender affect the amount of reduction in each of weight, cholesterol level, fat intake, cholesterol intake, and calories from fat?
c. Does age affect the amount of reduction in weight, cholesterol level, fat intake, cholesterol intake, and calories from fat cholesterol?

*The author would like to thank Karen Cavrag for writing this case.

This page intentionally left blank

© Comstock Images/Jupiterimages

17

MULTIPLE REGRESSION

17.1 *Model and Required Conditions*

17.2 *Estimating the Coefficients and Assessing the Model*

17.3 *Regression Diagnostics-II*

17.4 *Regression Diagnostics-III (Time Series)*

Appendix 17 *Review of Chapters 12 to 17*

General Social Survey

Variables That Affect Income

DATA
GSS2008*

In the Chapter 16 opening example, we used the General Social Survey to show that income and education are linearly related. This raises the question, What other variables affect one's income? To answer this question, we need to expand the simple linear regression technique used in the previous chapter to allow for more than one independent variable.

Here is a list of all the interval variables the General Social Survey created.

Age (AGE)

Years of education of respondent, spouse, father, and mother (EDUC, SPEDUC, PAEDUC, MAEDUC)

© mundoview/Shutterstock

Our answer appears on page 712.

Hours of work per week of respondent and of spouse (HRS and SPHRS)

Occupation prestige score of respondent, spouse, father, and mother (PRESTG80, SPPRES80, PAPRES80, MAPRES80)

Number of children (CHILDS)

Age when first child was born (AGEKDBRN)

Number of family members earning money (EARNRS)

Score on question "Should government reduce income differences between rich and poor?" (EQWLTH)

Score on question "Should government improve standard of living of poor people?" (HELPPOOR)

Score on question "Should government do more or less to solve country's problems?" (HELPNOT)

Score on question "Is it government's responsibility to help pay for doctor and hospital bills?" (HELPSICK)

Number of hours of television viewing per day (TVHOURS)

Years with current employer (CUREMPYR)

The goal is to create a regression analysis that includes all variables that you believe affect the amount of time spent watching television.

INTRODUCTION

In the previous chapter, we employed the simple linear regression model to analyze how one variable (the dependent variable y) is related to another interval variable (the independent variable x). The restriction of using only one independent variable was motivated by the need to simplify the introduction to regression analysis. Although there are a number of applications where we purposely develop a model with only one independent variable (see Section 4.6, for example), in general we prefer to include as many independent variables as are believed to affect the dependent variable. Arbitrarily limiting the number of independent variables also limits the usefulness of the model.

In this chapter, we allow for any number of independent variables. In so doing, we expect to develop models that fit the data better than would a simple linear regression model. We begin by describing the multiple regression model and listing the required conditions. We let the computer produce the required statistics and use them to assess the model's fit and diagnose violations of the required conditions. We use the model by interpreting the coefficients, predicting the particular value of the dependent variable, and estimating its expected value.

17.1 MODEL AND REQUIRED CONDITIONS

We now assume that k independent variables are potentially related to the dependent variable. Thus, the model is represented by the following equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

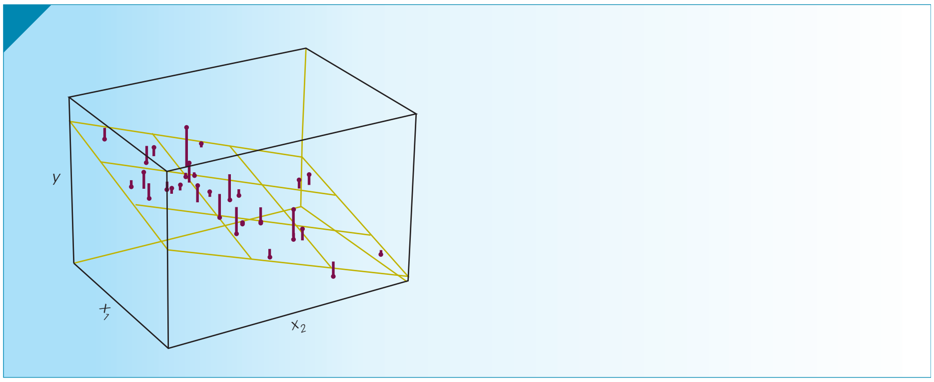
where y is the dependent variable, $x_1, x_2, \ldots, x_k$ are the independent variables, $\beta_0, \beta_1, \ldots, \beta_k$ are the coefficients, and ε is the error variable. The independent variables may actually be functions of other variables. For example, we might define some of the independent variables as follows:

$$x_2 = x_1^2$$
$$x_5 = x_3 x_4$$
$$x_7 = \log(x_6)$$

In Chapter 18, we will discuss how and under what circumstances such functions can be used in regression analysis.

The error variable is retained because, even though we have included additional independent variables, deviations between predicted values of y and actual values of y will still occur. Incidentally, when there is more than one independent variable in the regression model, we refer to the graphical depiction of the equation as a **response surface** rather than as a straight line. Figure 17.1 depicts a scatter diagram of a response surface with $k = 2$. (When $k = 2$, the regression equation creates a plane.) Of course, whenever k is greater than 2, we can only imagine the response surface; we cannot draw it.

FIGURE **17.1** Scatter Diagram and Response Surface with $k = 2$



An important part of the regression analysis comprises several statistical techniques that evaluate how well the model fits the data. These techniques require the following conditions, which we introduced in the previous chapter.

Required Conditions for Error Variable

1. The probability distribution of the error variable ε is normal.
2. The mean of the error variable is 0.
3. The standard deviation of ε is σ_ε, which is a constant.
4. The errors are independent.

In Section 16.6, we discussed how to recognize when the requirements are unsatisfied. Those same procedures can be used to detect violations of required conditions in the multiple regression model.

We now proceed as we did in Chapter 16. We discuss how the model's coefficients are estimated and how we assess the model's fit. However, there is one major difference between Chapters 16 and 17. In Chapter 16, we allowed for the possibility that some students will perform the calculations manually. The multiple regression model involves so many computations that it is virtually impossible to conduct the analysis without a computer. All analyses in this chapter will be performed by Excel and Minitab. Your job will be to interpret the output.

17.2 Estimating the Coefficients and Assessing the Model

The multiple regression equation is expressed similarly to the simple regression equation. The general form is

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \cdots + b_kx_k$$

where k is the number of independent variables.

The procedures introduced in Chapter 16 are extended to the multiple regression model. However, in Chapter 16, we first discussed how to interpret the coefficients and then discussed how to assess the model's fit. In practice, we reverse the process—that is, the first step is to determine how well the model fits. If the model's fit is poor, there is no point in a further analysis of the coefficients of that model. A much higher priority is assigned to the task of improving the model. We will discuss the art and science of model building in Chapter 18. In this chapter, we show how a regression analysis is performed. The steps we use are as follows:

1. Select variables that you believe are linearly related to the dependent variable.
2. Use a computer and software to generate the coefficients and the statistics used to assess the model.
3. Diagnose violations of required conditions. If there are problems, attempt to remedy them.
4. Assess the model's fit. Three statistics that perform this function are the standard error of estimate, the coefficient of determination, and the *F*-test of the analysis of variance. The first two were introduced in Chapter 16; the third will be introduced here.
5. If we are satisfied with the model's fit and that the required conditions are met, we can interpret the coefficients and test them as we did in Chapter 16. We use the model to predict a value of the dependent variable or estimate the expected value of the dependent variable.

We'll illustrate the procedure with the chapter-opening example.

Step 1: Select the Independent Variables

Here are the variables we believe may be linearly related to income.

Age (AGE): For most people income increases with age.

Years of education (EDUC): We've already shown that education is linearly related to income.

Hours of work per week (HRS): Obviously, more hours of work should equal more income.

Spouse's hours of work (SPHRS): It is possible that if one's spouse works more and earns more, the other spouse may choose to work less and thus earn less.

Occupation prestige score (PRESTG80): Occupations with higher prestige scores tend to pay more.

Number of children (CHILDS): Children are expensive, which may encourage their parents to work harder and thus earn more.

Number of family members earning money (EARNRS): As is the case with SPHRS, if more family members earn income, there may be less pressure on the respondent to work harder.

Years with current employer (CUREMPYR): This variable could be negatively or positively related to income.

You may be wondering why we don't simply include all the interval variables that are available to us. There are three reasons. First, the objective is to determine whether our hypothesized model is valid and whether the independent variables in the model are linearly related to the dependent variable. In other words, we should screen the independent variables and include only those that, in theory, affect the dependent variable.

Second, by including large numbers of independent variables, we increase the probability of Type I errors. For example, if we include 100 independent variables, none of which are related to the dependent variable, we're likely to conclude that 5 of them are linearly related to the dependent variable. This is a problem we discussed in Chapter 14.

Third, because of a problem called *multicollinearity* (described in Section 17.3), we may conclude that none of the independent variables are linearly related to the dependent variable when in fact one or more are.

Step 2: Use a Computer to Compute All Coefficients and Other Statistics

EXCEL

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.5809				
5	R Square	0.3374				
6	Adjusted R Square	0.3180				
7	Standard Error	33250				
8	Observations	282				
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression	1	153,716,984,625	19,214,623,078	17.38	7.02E-21
13	Residual	273	301,813,647,689	1,105,544,497		
14	Total	281	455,530,632,314			
15						
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
17	Intercept	−51785	19259	−2.69	0.0076	
18	AGE	461	237	1.95	0.0527	
19	EDUC	4101	848	4.84	0.0000	
20	HRS	620	173	3.59	0.0004	
21	SPHRS	−862	185	−4.67	4.71E-06	
22	PRESTG80	641	176	3.64	0.0003	
23	CHILDS	−331	1522	−0.22	0.8279	
24	EARNRS	687	2929	0.23	0.8147	
25	CUREMPYR	330	237	1.39	0.1649	

INSTRUCTIONS

1. Type or import the data so that the independent variables are in adjacent columns. Note that all rows with blanks (missing data) must be deleted.
2. Click **Data, Data Analysis,** and **Regression**.
3. Specify the **Input *Y* Range**, the **Input *X* Range**, and a value for α (.05).

MINITAB

Regression Analysis: Income versus AGE, EDUC, . . .

The regression equation is
Income = − 51785 + 461 Age + 4101 Educ + 620 Hrs − 862 Sphrs + 641 Prestg80 − 331 Childs + 687 Earnrs + 330 Curempyr
282 cases used, 1741 cases contain missing values

Predictor	Coef	SE Coef	T	P
Constant	−51785	19259	−2.69	0.008
Age	460.9	236.9	1.95	0.053
Educ	4100.9	847.7	4.84	0.000
Hrs	620.0	172.9	3.59	0.000
Sphrs	−862.2	184.6	−4.67	0.000
Prestg80	640.5	175.9	3.64	0.000
Childs	−331	1522	−0.22	0.828
Earnrs	687	2929	0.23	0.815
Curempyr	329.8	236.8	1.39	0.165

S = 33249.7 R-Sq = 33.7% R-Sq(adj) = 31.8%

INSTRUCTIONS

1. Click **Stat, Regression,** and **Regression . . . .**
2. Specify the dependent variable in the **Response** box and the independent variables in the **Predictors** box.

INTERPRET

The regression model is estimated by

$$\begin{aligned}\hat{y}(\text{INCOME}) = {} & -51{,}785 + 461\ \text{AGE} + 4101\ \text{EDUC} + 620\ \text{HRS} - 862\ \text{SPHRS} \\ & + 641\ \text{PRESTG80} - 331\ \text{CHILDS} + 687\ \text{EARNRS} \\ & +330\ \text{CUREMPYR}\end{aligned}$$

We assess the model in three ways: the standard error of estimate, the coefficient of determination (both introduced in Chapter 16) and the F-test of the analysis of variance (presented subsequently).

Standard Error of Estimate

Recall that σ_ε is the standard deviation of the error variable ε and that, because σ_ε is a population parameter, it is necessary to estimate its value by using s_ε. In multiple regression, the standard error of estimate is defined as follows.

Standard Error of Estimate

$$s_\varepsilon = \sqrt{\frac{\text{SSE}}{n - k - 1}}$$

where n is the sample size and k is the number of independent variables in the model.

As we noted in Chapter 16, each of our software packages reports the standard error of estimate in a different way.

EXCEL

	A	B
7	Standard Error	33250

MINITAB

S = 33249.7

INTERPRET

Recall that we judge the magnitude of the standard error of estimate relative to the values of the dependent variable, and particularly to the mean of y. In this example, $\bar{y} = 41{,}746$ (not shown in printouts). It appears that the standard error of estimate is quite large.

Coefficient of Determination

Recall from Chapter 16 that the coefficient of determination is defined as

$$R^2 = 1 - \frac{\text{SSE}}{\sum (y_i - \bar{y})^2}$$

EXCEL

	A	B
5	R Square	0.3374

MINITAB

R-Sq = 33.7%

INTERPRET

This means that 33.74% of the total variation in income is explained by the variation in the eight independent variables, whereas 66.26% remains unexplained.

Notice that Excel and Minitab print a second R^2 statistic, called the **coefficient of determination adjusted for degrees of freedom**, which has been adjusted to take into account the sample size and the number of independent variables. The rationale for this statistic is that, if the number of independent variables k is large relative to the sample size n, the unadjusted R^2 value may be unrealistically high. To understand this point, consider what would happen if the sample size is 2 in a simple linear regression model. The line would fit the data perfectly, resulting in $R^2 = 1$ when, in fact, there may be no linear relationship. To avoid creating a false impression, the adjusted R^2 is often calculated. Its formula follows.

Coefficient of Determination Adjusted for Degrees of Freedom

$$\text{Adjusted } R^2 = 1 - \frac{\text{SSE}/(n - k - 1)}{\sum (y_i - \bar{y})^2/(n - 1)} = 1 - \frac{\text{MSE}}{s_y^2}$$

If n is considerably larger than k, the unadjusted and adjusted R^2 values will be similar. But if SSE is quite different from 0 and k is large compared to n, the unadjusted and adjusted values of R^2 will differ substantially. If such differences exist, the analyst should be alerted to a potential problem in interpreting the coefficient of determination. In this example, the adjusted coefficient of determination is 31.80%, indicating that, no matter how we measure the coefficient of determination, the model's fit is not very good.

Testing the Validity of the Model

In the simple linear regression model, we tested the slope coefficient to determine whether sufficient evidence existed to allow us to conclude that there was a linear relationship between the independent variable and the dependent variable. However, because there is only one independent variable in that model, that same t-test was also tested to determine whether that model is valid. When there is more than one independent variable, we need another method to test the overall validity of the model. The technique is a version of the analysis of variance, which we introduced in Chapter 14.

To test the validity of the regression model, we specify the following hypotheses:

H_0: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$

H_1: At least one β_i is not equal to 0

If the null hypothesis is true, none of the independent variables $x_1, x_2, \ldots, x_k$ is linearly related to y, and therefore the model is invalid. If at least one β_i is not equal to 0, the model does have some validity.

When we discussed the coefficient of determination in Chapter 16, we noted that the total variation in the dependent variable [measured by $\sum (y_i - \bar{y})^2$] can be decomposed into two parts: the explained variation (measured by SSR) and the unexplained variation (measured by SSE); that is,

Total Variation in y = SSR + SSE

Furthermore, we established that, if SSR is large relative to SSE, the coefficient of determination will be high—signifying a good model. On the other hand, if SSE is large, most of the variation will be unexplained, which indicates that the model provides a poor fit and consequently has little validity.

The test statistic is the same one we encountered in Section 14.1, where we tested for the equivalence of two or more population means. To judge whether SSR is large enough relative to SSE to allow us to infer that at least one coefficient is not equal to 0, we compute the ratio of the two mean squares. (Recall that the mean square is the sum of squares divided by its degrees of freedom; recall, too, that the ratio of two mean squares is F-distributed as long as the underlying population is normal—a required condition for this application.) The calculation of the test statistic is summarized in an analysis of variance (ANOVA) table, whose general form appears in Table 17.1. The Excel and Minitab ANOVA tables are shown next.

TABLE **17.1** Analysis of Variance Table for Regression Analysis

SOURCE OF VARIATION	DEGREES OF FREEDOM	SUMS OF SQUARES	MEAN SQUARES	F-STATISTIC
Regression	k	SSR	MSR = SSR/k	F = MSR/MSE
Residual	$n - k - 1$	SSE	MSE = SSE/$(n - k - 1)$	
Total	$n - 1$	$\sum(y_i - \bar{y})^2$		

EXCEL

	A	B	C	D	E	F
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression	1	153,716,984,625	19,214,623,078	17.38	7.02E-21
13	Residual	273	301,813,647,689	1,105,544,497		
14	Total	281	455,530,632,314			

MINITAB

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	1.53717E+11	19,214,623,078	17.38	0.000
Residual Error	273	3.01814E+11	1,105,544,497		
Total	281	4.55531E+11			

A large value of F indicates that most of the variation in y is explained by the regression equation and that the model is valid. A small value of F indicates that most of the variation in y is unexplained. The rejection region allows us to determine whether F is large enough to justify rejecting the null hypothesis. For this test, the rejection region is

$$F > F_{\alpha,k,n-k-1}$$

In Example 17.1, the rejection region (assuming $\alpha = .05$) is

$$F > F_{\alpha,k,n-k-1} = F_{.05,8,273} \approx 1.98$$

As you can see from the printout, $F = 17.38$. The printout also includes the p-value of the test, which is 0. Obviously, there is a great deal of evidence to infer that the model is valid.

Although each assessment measurement offers a different perspective, all agree in their assessment of how well the model fits the data, because all are based on the sum of squares for error, SSE. The standard error of estimate is

$$s_\varepsilon = \sqrt{\frac{\text{SSE}}{n-k-1}}$$

and the coefficient of determination is

$$R^2 = 1 - \frac{\text{SSE}}{\sum (y_i - \bar{y})^2}$$

When the response surface hits every single point, SSE = 0. Hence, $s_\varepsilon = 0$ and $R^2 = 1$.

If the model provides a poor fit, we know that SSE will be large [its maximum value is $\sum (y_i - \bar{y})^2$], s_ε will be large, and [because SSE is close to $\sum (y_i - \bar{y})^2$] R^2 will be close to 0.

The F-statistic also depends on SSE. Specifically,

$$F = \frac{\text{MSR}}{\text{MSE}} = \frac{\left(\sum (y_i - \bar{y})^2 - \text{SSE}\right)/k}{\text{SSE}/(n-k-1)}$$

When SSE = 0,

$$F = \frac{\sum (y_i - \bar{y})^2/k}{0/(n-k-1)}$$

which is infinitely large. When SSE is large, SSE is close to $\sum (y_i - \bar{y})^2$ and F is quite small.

The relationship among s_ε, R^2, and F are summarized in Table 17.2.

TABLE **17.2** Relationship among SSE, s_ε, R^2, and F

SSE	s_ε	R^2	F	ASSESSMENT OF MODEL
0	0	1	∞	Perfect
Small	Small	Close to 1	Large	Good
Large	Large	Close to 0	Small	Poor
$\sum (y_i - \bar{y})^2$	$\sqrt{\frac{\sum (y_i - \bar{y})^2}{n-k-1}}$*	0	0	Useless

*When n is large and k is small, this quantity is approximately equal to the standard deviation of y.

If we're satisfied that the model fits the data as well as possible and that the required conditions are satisfied, we can interpret and test the individual coefficients and use the model to predict and estimate.

Interpreting the Coefficients

The coefficients $b_0, b_1, \ldots, b_k$ describe the relationship between each of the independent variables and the dependent variable in the sample. We need to use inferential methods (described below) to draw conclusions about the population. In our example, the sample consists of the 657 observations. The population is composed of all American adults.

Intercept The intercept is $b_0 = -51{,}785$. This is the average income when all the independent variables are zero. As we observed in Chapter 16, it is often misleading to try to interpret this value, particularly if 0 is outside the range of the values of the independent variables (as is the case here).

Age The relationship between income and age is described by $b_1 = 461$. From this number, we learn that for each additional year of age in this model, income increases on average by $461, assuming that the other independent variables in this model are held constant.

Education The coefficient $b_2 = 4{,}101$ specifies that in this sample for each additional year of education the income increases on average by $4,101, assuming the constancy of the other independent variables.

Hours of Work The relationship between hours of work per week is expressed by $b_3 = 620$. We interpret this number as the average increase in annual income for each additional hour of work per week, keeping the other independent variables fixed in this sample.

Spouse's Hours of Work The relationship between annual income and a spouse's hours of work per week is described in this sample by $b_4 = -862$, which we interpret to mean that for each additional hour a spouse works per week, income decreases on average by $862 when the other variables are constant.

Occupation Prestige Score In this sample, the relationship between annual income and occupation prestige score is described by $b_5 = 641$. For each additional unit increase in the occupation prestige score, annual income increases on average by $641, holding all other variables constant.

Number of Children The relationship between annual income and number of children is expressed by $b_6 = -331$, which tells us that for each additional child, annual income decreases on average by $331 in this sample.

Number of Family Members Earning Income In this dataset, the relationship between annual income and the number of family members who earn money is expressed by $b_7 = 687$, which tells us that for each additional family member earner, annual income increases on average by $687, assuming that the other independent variables are constant.

Number of Years with Current Job The coefficient of the last independent variable in this model is $b_8 = 330$. This number means that for each additional year of job tenure with the current company, annual income increases on average by $330, keeping the other independent variables constant in this sample.

Testing the Coefficients

In Chapter 16, we described how to test to determine whether there is sufficient evidence to infer that in the simple linear regression model x and y are linearly related. The null and alternative hypotheses were

$$H_0: \quad \beta_1 = 0$$

$$H_1: \quad \beta_1 \neq 0$$

The test statistic was

$$t = \frac{b_1 - \beta_1}{s_{b_1}}$$

which is Student t distributed with $\nu = n - 2$ degrees of freedom.

In the multiple regression model, we have more than one independent variable. For each such variable, we can test to determine whether there is enough evidence of a linear relationship between it and the dependent variable for the entire population when the other independent variables are included in the model.

Testing the Coefficients

$$H_0: \quad \beta_i = 0$$
$$H_1: \quad \beta_i \neq 0$$

(for $i = 1, 2, \ldots, k$); the test statistic is

$$t = \frac{b_i - \beta_i}{s_{b_i}}$$

which is Student t distributed with $\nu = n - k - 1$ degrees of freedom.

To illustrate, we test each coefficient in the multiple regression model in the chapter-opening example. The tests that follow are performed just as all other tests in this book have been performed. We set up the null and alternative hypotheses, identify the test statistic, and use the computer to calculate the value of the test statistic and its p-value. For each independent variable, we test ($i = 1, 2, 3, 4, 5, 6, 7, 8$).

$$H_0: \quad \beta_i = 0$$
$$H_1: \quad \beta_i \neq 0$$

Refer to page 696 and 697 and examine the computer output. The output includes the t-tests of β_i. The results of these tests pertain to the entire population of the United States in 2008. It is also important to add that these test results were determined when the other independent variables were included in the model. We add this statement because a simple linear regression will very likely result in different values of the test statistics and possibly the conclusion.

Test of β_1 (Coefficient of age)

Value of the test statistic: $t = 1.95$; p-value $= .0527$

Test of β_2 (Coefficient of education)

Value of the test statistic: $t = 4.84$; p-value $= 0$

Test of β_3 (Coefficient of number of hours of work per week)

Value of the test statistic: $t = 3.59$; p-value $= .0004$

Test of β_4 (Coefficient of spouse's number of hours of work per week)

Value of the test statistic: $t = -4.67$; p-value $= 0$

Test of β_5 (Coefficient of occupation prestige score)

Value of the test statistic: $t = 3.64$; p-value $= .0003$

Test of β_6 (Coefficient of number of children)

Value of the test statistic: $t = -.22$; p-value $= .8279$

Test of β_7 (Coefficient of number of earners in family)

Value of the test statistic: $t = .23$; p-value $= .8147$

Test of β_8 (Coefficient of years with current employer)

Value of the test statistic: $t = 1.39$; p-value $= .1649$

There is sufficient evidence at the 5% significance level to infer that each of the following variables is linearly related to income:

Education

Number of hours of work per week

Spouse's number of hours of work per week

Occupation prestige score

There is weak evidence to infer that income and age are linearly related.

In this model, there is not enough evidence to conclude that each of the following variables is linearly related to income:

Number of children

Number of earners in the family

Number of years with current employer

Note that this may mean that there is no evidence of a linear relationship between income and these three independent variables. However, it may also mean that there is a linear relationship between income and one or more of these variables, but because of a condition called *multicollinearity*, the t-tests revealed no linear relationship. We will discuss multicollinearity in Section 17.3.

A Cautionary Note About Interpreting the Results

Care should be taken when interpreting the results of this and other regression analyses. We might find that in one model there is enough evidence to conclude that a particular independent variable is linearly related to the dependent variable, but that no such evidence exists in another model. Consequently, whenever a particular t-test is *not* significant, we state that there is not enough evidence to infer that the independent and dependent variable are linearly related *in this model.* The implication is that another model may yield different conclusions.

Furthermore, if one or more of the required conditions are violated, the results may be invalid. In Section 16.6, we introduced the procedures that allow the statistics practitioner to examine the model's requirements. We will add to this discussion in Section 17.3. We also remind you that it is dangerous to extrapolate far outside the range of the observed values of the independent variables.

t-Tests and the Analysis of Variance

The t-tests of the individual coefficients allow us to determine whether $\beta_i \neq 0$ (for $i = 1, 2, \ldots, k$), which tells us whether a linear relationship exists between x_i and y. There is a t-test for each independent variable. Consequently, the computer automatically performs k t-tests. (It actually conducts $k + 1$ t-tests, including the one for the intercept β_0, which we usually ignore.) The F-test in the analysis of variance combines these t-tests into a single test. In other words, we test all the β_i at one time to determine whether at least one of them is not equal to 0. The question naturally arises, Why do we need the F-test if it is nothing more than the combination of the previously performed t-tests? Recall that we addressed this issue before. In Chapter 14, we pointed out that we can replace the analysis of variance by a series of t-tests of the difference between two means. However, by doing so, we increase the probability of making a Type I error. That means that even when there is no linear relationship between each of the independent variables and the dependent variable, multiple t-tests will likely show some are significant. As a result, you will conclude erroneously that, because at least one β_i is not equal to 0, the model is valid. The F-test, on the other hand, is performed only once. Because the probability that a Type I error will occur in a single trial is equal to α, the chance of erroneously concluding that the model is valid is substantially less with the F-test than with multiple t-tests.

There is another reason that the F-test is superior to multiple t-tests. Because of a commonly occurring problem called *multicollinearity*, the t-tests may indicate that some independent variables are not linearly related to the dependent variable, when in fact they are. The problem of multicollinearity does not affect the F-test, nor does it inhibit us from developing a model that fits the data well. Multicollinearity is discussed in Section 17.3.

The *F*-Test and the *t*-Test in the Simple Linear Regression Model

It is useful for you to know that we can use the F-test to test the validity of the simple linear regression model. However, this test is identical to the t-test of β_1. The t-test of β_1 in the simple linear regression model tells us whether that independent variable is linearly related to the dependent variable. However, because there is only one independent variable, the t-test of β_1 also tells us whether the model is valid, which is the purpose of the F-test.

The relationship between the t-test of β_i and the F-test can be explained mathematically. Statisticians can show that if we square a t-statistic with ν degrees of freedom, we produce an F-statistic with 1 and ν degrees of freedom. (We briefly discussed this relationship in Chapter 14.) To illustrate, consider Example 16.2 on page 641. We found the t-test of β_1 to be -13.44, with degrees of freedom equal to 98. The p-value was 5.75×10^{-24}. The output included the analysis of variance table where $F = 180.64$ and p-value was 5.75×10^{-24}. The t-statistic squared is $t^2 = (-13.44)^2 = 180.63$. (The difference is the result of rounding errors.) Notice that the degrees of freedom of the F-statistic are 1 and 98. Thus, we can use either test to test the validity of the simple linear regression model.

Using the Regression Equation

As was the case with simple linear regression, we can use the multiple regression equation in two ways: We can produce the prediction interval for a particular value of y, and we can produce the confidence interval estimate of the expected value of y. Like the other calculations associated with multiple regression, we call on the computer to do the work.

To illustrate, we'll predict the income of a 50-year-old, with 12 years of education, who works 40 hours per week, has a spouse who also works 40 hours per week (i.e., 2 earners in the family), has an occupation prestige score of 50, has 2 children, and has worked for the same company for 5 years.

As you discovered in the previous chapter, both Excel and Minitab output the prediction interval and interval estimate of the expected value of incomes for all people with the given variables.

EXCEL

	A	B	C
1	**Prediction Interval**		
2			
3			Margin
4			
5	Predicted value		45,168
6			
7	Prediction Interval		
8	Lower limit		−20,719
9	Upper limit		111,056
10			
11	Interval Estimate of Expected Value		
12	Lower limit		37,661
13	Upper limit		52,675

INSTRUCTIONS

See the instructions on page 669. In cells B284 to I284, we input the values 50 12 40 40 50 2 2 5, respectively. We specified 95% confidence.

MINITAB

Predicted Values for New Observations

New Obs	Fit	SE Fit	95% CI	95% PI
1	45,168	3,813	(37,661, 52,675)	(−20,719, 111,056)

Values of Predictors for New Observations

New Obs	Age	Educ	Hrs	Sphrs	Prestg80	Childs	Earnrs	Curempyr
1	50.0	12.0	40.0	40.0	50.0	2.00	2.00	5.00

INSTRUCTIONS

See the instructions on page 669. We input the values 50 12 40 40 50 2 2 5. We specified 95% confidence.

INTERPRET

The prediction interval is −20,719, 111,056. It is so wide as to be completely useless. To be useful in predicting values, the model must be considerably better. The confidence interval estimate of the expected income of a population is 37,661, 52,675.

EXERCISES

The following exercises require the use of a computer and statistical software. Exercises 17.1–17.4 can be solved manually. See Appendix A for the sample statistics. Use a 5% significance level.

17.1 Xr17-01 A developer who specializes in summer cottage properties is considering purchasing a large tract of land adjoining a lake. The current owner of the tract has already subdivided the land into separate building lots and has prepared the lots by removing some of the trees. The developer wants to forecast the value of each lot. From previous experience, she knows that the most important factors affecting the price of a lot are size, number of mature trees, and distance to the lake. From a nearby area, she gathers the relevant data for 60 recently sold lots.

a. Find the regression equation.
b. What is the standard error of estimate? Interpret its value.
c. What is the coefficient of determination? What does this statistic tell you?
d. What is the coefficient of determination, adjusted for degrees of freedom? Why does this value differ from the coefficient of determination? What does this tell you about the model?
e. Test the validity of the model. What does the *p*-value of the test statistic tell you?
f. Interpret each of the coefficients.
g. Test to determine whether each of the independent variables is linearly related to the price of the lot in this model.
h. Predict with 90% confidence the selling price of a 40,000-square-foot lot that has 50 mature trees and is 25 feet from the lake.
i. Estimate with 90% confidence the average selling price of 50,000-square-foot lots that have 10 mature trees and are 75 feet from the lake.

17.2 Xr17-02 Pat Statsdud, a student ranking near the bottom of the statistics class, decided that a certain amount of studying could actually improve final grades. However, too much studying would not be warranted because Pat's ambition (if that's what one could call it) was to ultimately graduate with the absolute minimum level of work. Pat was registered in a statistics course that had only 3 weeks to go before the final exam and for which the final grade was determined in the following way:

Total mark = 20% (Assignment)
+ 30% (Midterm test)
+ 50% (Final exam)

To determine how much work to do in the remaining 3 weeks, Pat needed to be able to predict the final exam mark on the basis of the assignment mark (worth 20 points) and the midterm mark (worth 30 points). Pat's marks on these were 12/20 and 14/30, respectively. Accordingly, Pat undertook the following analysis. The final exam mark, assignment mark, and midterm test mark for 30 students who took the statistics course last year were collected.

a. Determine the regression equation.
b. What is the standard error of estimate? Briefly describe how you interpret this statistic.
c. What is the coefficient of determination? What does this statistic tell you?
d. Test the validity of the model.
e. Interpret each of the coefficients.
f. Can Pat infer that the assignment mark is linearly related to the final grade in this model?
g. Can Pat infer that the midterm mark is linearly related to the final grade in this model?
h. Predict Pat's final exam mark with 95% confidence.
i. Predict Pat's final grade with 95% confidence.

17.3 Xr17-03 The president of a company that manufactures drywall wants to analyze the variables that affect demand for his product. Drywall is used to construct walls in houses and offices. Consequently, the president decides to develop a regression model in which the dependent variable is monthly sales of drywall (in hundreds of 4 ×8 sheets) and the independent variables are

Number of building permits issued in the county
Five-year mortgage rates (in percentage points)
Vacancy rate in apartments (in percentage points)
Vacancy rate in office buildings (in percentage points)
To estimate a multiple regression model, he took monthly observations from the past 2 years.

a. Analyze the data using multiple regression.
b. What is the standard error of estimate? Can you use this statistic to assess the model's fit? If so, how?
c. What is the coefficient of determination, and what does it tell you about the regression model?
d. Test the overall validity of the model.
e. Interpret each of the coefficients.
f. Test to determine whether each of the independent variables is linearly related to drywall demand in this model.
g. Predict next month's drywall sales with 95% confidence if the number of building permits is 50, the 5-year mortgage rate is 9.0%, and the vacancy rates are 3.6% in apartments and 14.3% in office buildings.

17.4 Xr17-04 The general manager of the Cleveland Indians baseball team is in the process of determining which minor-league players to draft. He is aware that his team needs home-run hitters and would like to find a way to predict the number of home runs a player will hit. Being an astute statistician, he gathers a random sample of players and records the number of home runs each player hit in his first two full years as a major-league player, the number of home runs he hit in his last full year in the minor leagues, his age, and the number of years of professional baseball.

a. Develop a regression model and use a software package to produce the statistics.
b. Interpret each of the coefficients.
c. How well does the model fit?
d. Test the model's validity.
e. Do each of the independent variables belong in the model?
f. Calculate the 95% interval of the number of home runs in the first two years of a player who is 25 years old, has played professional baseball for 7 years, and hit 22 home runs in his last year in the minor leagues.
g. Calculate the 95% interval of the expected number of home runs in the first two years of players who are 27 years old, have played professional baseball for 5 years, and hit 18 home runs in their last year in the minors.

APPLICATIONS in HUMAN RESOURCES MANAGEMENT


© Losevsky Pavel/Shutterstock

Severance Pay

In most firms, the entire issue of compensation falls into the domain of the human resources manager. The manager must ensure that the method used to determine compensation contributes to the firm's objectives. Moreover, the firm needs to ensure that discrimination or bias of any kind is not a factor. Another function of the personnel manager is to develop severance packages for employees whose services are no longer needed because of downsizing or merger. The size and nature of severance is rarely part of any working agreement and must be determined by a variety of factors. Regression analysis is often useful in this area.

17.5 Xr17-05 When one company buys another company, it is not unusual that some workers are terminated. The severance benefits offered to the laid-off workers are often the subject of dispute. Suppose that the Laurier Company recently bought the Western Company and subsequently terminated 20 of Western's employees. As part of the buyout agreement, it was promised that the severance packages offered to the former Western employees would be equivalent to those offered to Laurier employees who had been terminated in the past year. Thirty-six-year-old Bill Smith, a Western employee for the past 10 years, earning $32,000 per year, was one of those let go. His severance package included an offer of 5 weeks' severance pay. Bill complained that this offer was less than that offered to Laurier's employees when they were laid off, in contravention of the buyout agreement. A statistician was called in to settle the dispute. The statistician was told that severance is determined by three factors: age, length of service with the company, and pay. To determine how generous the severance package had been, a random sample of 50 Laurier ex-employees was taken. For each, the following variables were recorded:

Number of weeks of severance pay
Age of employee
Number of years with the company
Annual pay (in thousands of dollars)

a. Determine the regression equation.
b. Comment on how well the model fits the data.
c. Do all the independent variables belong in the equation? Explain.
d. Perform an analysis to determine whether Bill is correct in his assessment of the severance package.

17.6 Xr17-06 The admissions officer of a university is trying to develop a formal system to decide which students to admit to the university. She believes that determinants of success include the standard variables—high school grades and SAT scores. However, she also believes that students who have participated in extracurricular activities are more likely to succeed than those who have not. To investigate the issue, she randomly sampled 100 fourth-year students and recorded the following variables:

GPA for the first 3 years at the university (range: 0 to 12)
GPA from high school (range: 0 to 12)
SAT score (range: 400 to 1600)
Number of hours on average spent per week in organized extracurricular activities in the last year of high school

a. Develop a model that helps the admissions officer decide which students to admit and use the computer to generate the usual statistics.
b. What is the coefficient of determination? Interpret its value.
c. Test the overall validity of the model.
d. Test to determine whether each of the independent variables is linearly related to the dependent variable in this model.
e. Determine the 95% interval of the GPA for the first 3 years of university for a student whose high school GPA is 10, whose SAT score is 1200, and who worked an average of 2 hours per week on organized extracurricular activities in the last year of high school.
f. Find the 90% interval of the mean GPA for the first 3 years of university for all students whose high school GPA is 8, whose SAT score is 1100, and who worked an average of 10 hours per week on organized extracurricular activities in the last year of high school.

17.7 Xr17-07 The marketing manager for a chain of hardware stores needed more information about the effectiveness of the three types of advertising that the chain used. These are localized direct mailing (in which flyers describing sales and featured products are distributed to homes in the area surrounding a store), newspaper advertising, and local television advertisements. To determine which type is most effective, the manager collected 1 week's data from 100 randomly selected stores. For each store, the following variables were recorded:

Weekly gross sales
Weekly expenditures on direct mailing
Weekly expenditures on newspaper advertising
Weekly expenditures on television commercials

All variables were recorded in thousands of dollars.

a. Find the regression equation.
b. What are the coefficient of determination and the coefficient of determination adjusted for degrees of freedom? What do these statistics tell you about the regression equation?
c. What does the standard error of estimate tell you about the regression model?
d. Test the validity of the model.
e. Which independent variables are linearly related to weekly gross sales in this model? Explain.
f. Compute the 95% interval of the week's gross sales if a local store spent $800 on direct mailing, $1,200 on newspaper advertisements, and $2,000 on television commercials.
g. Calculate the 95% interval of the mean weekly gross sales for all stores that spend $800 on direct mailing, $1,200 on newspaper advertising, and $2,000 on television commercials.
h. Discuss the difference between the two intervals found in parts (f) and (g).

17.8 Xr17-08 For many cities around the world, garbage is an increasing problem. Many North American cities have virtually run out of space to dump the garbage. A consultant for a large American city decided to gather data about the problem. She took a random sample of houses and determined the following:

Y = the amount of garbage per average week (pounds)
X_1 = Size of the house (square feet)
X_2 = Number of children
X_3 = Number of adults who are usually home during the day

a. Conduct a regression analysis.
b. Is the model valid?
c. Interpret each of the coefficients.
d. Test to determine whether each of the independent variables is linearly related to the dependent variable.

17.9 Xr17-09 The administrator of a school board in a large county was analyzing the average mathematics test scores in the schools under her control. She noticed that there were dramatic differences in scores among the schools. In an attempt to improve the scores of all the schools, she attempted to determine the factors that account for the differences. Accordingly, she took a random sample of 40 schools across the county and, for each, determined the mean test score last year, the percentage of teachers in each school who have at least one university degree in mathematics, the mean age, and the mean annual income (in $1,000s) of the mathematics teachers.

a. Conduct a regression analysis to develop the equation.
b. Is the model valid?

c. Interpret and test the coefficients.
d. Predict with 95% confidence the test score at a school where 50% of the mathematics teachers have mathematics degrees, the mean age is 43, and the mean annual income is $48,300.

17.10 Xr17-10* Life insurance companies are keenly interested in predicting how long their customers will live because their premiums and profitability depend on such numbers. An actuary for one insurance company gathered data from 100 recently deceased male customers. He recorded the age at death of the customer plus the ages at death of his mother and father, the mean ages at death of his grandmothers, and the mean ages at death of his grandfathers.
a. Perform a multiple regression analysis on these data.
b. Is the model valid?
c. Interpret and test the coefficients.
d. Determine the 95% interval of the longevity of a man whose parents lived to the age of 70, whose grandmothers averaged 80 years, and whose grandfathers averaged 75 years.
e. Find the 95% interval of the mean longevity of men whose mothers lived to 75 years, whose fathers lived to 65 years, whose grandmothers averaged 85 years, and whose grandfathers averaged 75 years.

17.11 Xr17-11 University students often complain that universities reward professors for research but not for teaching, and they argue that professors react to this situation by devoting more time and energy to the publication of their findings and less time and energy to classroom activities. Professors counter that research and teaching go hand in hand: More research makes better teachers. A student organization at one university decided to investigate the issue. It randomly selected 50 economics professors who are employed by a multicampus university. The students recorded the salaries (in $1,000s) of the professors, their average teaching evaluations (on a 10-point scale), and the total number of journal articles published in their careers. Perform a complete analysis (produce the regression equation, assess it, and report your findings).

17.12 Xr17-12* One critical factor that determines the success of a catalog store chain is the availability of products that consumers want to buy. If a store is sold out, future sales to that customer are less likely. Accordingly, delivery trucks operating from a central warehouse regularly resupply stores. In an analysis of a chain's operations, the general manager wanted to determine the factors that are related to how long it takes to unload delivery trucks. A random sample of 50 deliveries to one store was observed. The times (in minutes) to unload the truck, the total number of boxes, and the total weight (in hundreds of pounds) of the boxes were recorded.

a. Determine the multiple regression equation.
b. How well does the model fit the data? Explain.
c. Interpret and test the coefficients.
d. Produce a 95% interval of the amount of time needed to unload a truck with 100 boxes weighing 5,000 pounds.
e. Produce a 95% interval of the average amount of time needed to unload trucks with 100 boxes weighing 5,000 pounds.

17.13 Xr17-13 Lotteries have become important sources of revenue for governments. Many people have criticized lotteries, however, referring to them as a tax on the poor and uneducated. In an examination of the issue, a random sample of 100 adults was asked how much they spend on lottery tickets and was interviewed about various socioeconomic variables. The purpose of this study is to test the following beliefs:

1. Relatively uneducated people spend more on lotteries than do relatively educated people.
2. Older people buy more lottery tickets than younger people.
3. People with more children spend more on lotteries than people with fewer children.
4. Relatively poor people spend a greater proportion of their income on lotteries than relatively rich people.

The following data were recorded:

Amount spent on lottery tickets as a percentage of total household income
Number of years of education
Age
Number of children
Personal income (in thousands of dollars)

a. Develop the multiple regression equation.
b. Is the model valid?
c. Test each of the beliefs. What conclusions can you draw?

17.14 Xr17-14* The MBA program at a large university is facing a pleasant problem—too many applicants. The current admissions policy requires students to have completed at least 3 years of work experience and an undergraduate degree with a B– average or better. Until 3 years ago, the school admitted any applicant who met these requirements. However, because the program recently converted from a 2-year program (four semesters) to a 1-year program (three semesters), the number of applicants has increased substantially. The dean, who teaches statistics courses, wants to raise the admissions standards by developing a method that more accurately predicts how well an applicant will perform in the MBA program. She believes that the primary determinants of success are the following:

Undergraduate grade point average (GPA)
Graduate Management Admissions Test (GMAT) score
Number of years of work experience

She randomly sampled students who completed the MBA and recorded their MBA program GPA, as well as the three variables listed here.

a. Develop a multiple regression model.
b. Test the model's validity.
c. Test to determine which of the independent variables is linearly related to MBA GPA.

APPLICATIONS in OPERATIONS MANAGEMENT

© Jeff Greenberg/ The Image Works

Location Analysis

Location analysis is one function of operations management. Deciding where to locate a plant, warehouse, or retail outlet is a critical decision for any organization. A large number of variables must be considered in this decision problem. For example, a production facility must be located close to suppliers of raw resources and supplies, skilled labor, and transportation to customers. Retail outlets must consider the type and number of potential customers. In the next example, we describe an application of regression analysis to find profitable locations for a motel chain.

17.15 Xr17-15 La Quinta Motor Inns is a moderately priced chain of motor inns located across the United States. Its market is the frequent business traveler. The chain recently launched a campaign to increase market share by building new inns. The management of the chain is aware of the difficulty in choosing locations for new motels. Moreover, making decisions without adequate information often results in poor decisions. Consequently, the chain's management acquired data on 100 randomly selected inns belonging to La Quinta. The objective was to predict which sites are likely to be profitable.

To measure profitability, La Quinta used *operating margin*, which is the ratio of the sum of profit, depreciation, and interest expenses divided by total revenue. (Although occupancy is often used as a measure of a motel's success, the company statistician concluded that occupancy was too unstable, especially during economic turbulence.) The higher the operating margin, the greater the success of the inn. La Quinta defines profitable inns as those with an operating margin in excess of 50%; unprofitable inns are those with margins of less than 30%. After a discussion with a number of experienced managers, La Quinta decided to select one or two independent variables from each of the following categories: competition, market awareness, demand generators, demographics, and physical location. To measure the degree of competition, they determined the total number of motel and hotel rooms within 3 miles of each La Quinta inn. Market awareness was measured by the number of miles to the closest competing motel. Two variables that represent sources of customers were chosen. The amount of office space and college and university enrollment in the surrounding community are demand generators. Both of these are measures of economic activity. A demographic variable that describes the community is the median household income. Finally, as a measure of the physical qualities of the location La Quinta chose the distance to the downtown core. These data are stored using the following format:

Column 1: y = operating margin, in percent
Column 2: x_1 = Total number of motel and hotel rooms within 3 miles of La Quinta inn

(Continued)

Column 3: x_2 = Number of miles to closest competition

Column 4: x_3 = Office space in thousands of square feet in surrounding community

Column 5: x_4 = College and university enrollment (in thousands) in nearby university or college

Column 6: x_5 = Median household income (in $thousands) in surrounding community

Column 7: x_6 = Distance (in miles) to the downtown core

Adapted from Sheryl E. Kimes and James A. Fitzsimmons, "Selecting Profitable Hotel Sites at La Quinta Motor Inns," *INTERFACES* 20 March–April 1990, pp. 12–20.

a. Develop a regression analysis.
b. Test to determine whether there is enough evidence to infer that the model is valid.
c. Test each of the slope coefficients.
d. Interpret the coefficients.
e. Predict with 95% confidence the operating margin of a site with the following characteristics.
There are 3,815 rooms within 3 miles of the site, the closest other hotel or motel is .9 miles away, the amount of office space is 476,000 square feet, there is one college and one university with a total enrollment of 24,500 students, the median income in the area is $35,000, and the distance to the downtown core is 11.2 miles.
f. Refer to part (e). Estimate with 95% confidence the mean operating margin of all La Quinta inns with those characteristics.

General Social Survey Exercises

17.16 GSS2008* How does the amount of education of one's parents (PAEDUC, MAEDUC) affect your education (EDUC)? Excel users note: You must delete rows with blanks.

a. Develop a regression model.
b. Test the validity of the model.
c. Test the two slope coefficients.
d. Interpret the coefficients.

17.17 GSS2008* What determines people's opinion on the following question? Should the government reduce income differences between rich and poor (EQWLTH)? (1 = government should reduce differences, 2–7 = No government action.)

a. Develop a regression analysis using demographic variables education (EDUC), age, (AGE), number of children (CHILDS), and occupation prestige score (PRESTG80).
b. Test the model's validity.
c. Test each of the slope coefficients.
d. Interpret the coefficient of determination.

17.18 GSS2008* The Nielsen ratings estimate the numbers of televisions tuned to various channels. However, television executives need more information. The General Social Survey may be the source of this information. Respondents were asked to report the number of hours per average day of television viewing (TVHOURS). Conduct a regression analysis using the following independent variables

Education (EDUC)
Age (AGE)
Hours of work (HRS)
Number of children (CHILDS)
Number of family members earning money (EARNRS)
Occupation prestige score (PRESTG80)

a. Test the model's validity.
b. Test each slope coefficient.
c. Determine the coefficient of determination and describe what it tells you.

17.19 GSS2008* What determines people's opinion on the following question? Should the government improve the standard of living of poor people (HELPPOOR)? (1 = Government act; 2–5 = People should help themselves).

a. Develop a regression analysis using demographic variables education (EDUC), age, (AGE), number of children (CHILDS), and occupation prestige score (PRESTG80.)
b. Test the model's validity.
c. Test each of the slope coefficients.
d. Interpret the coefficient of determination.

17.20 GSS2006* Xr17-20 Use the General Social Survey of 2006 to undertake a regression analysis of income (INCOME) using the following independent variables. (Because the GSS2008 file is so large we deleted the blanks and stored the variables in Xr17-20.)

Age (AGE)
Education (EDUC)
Hours of work (HRS)
Number of children (CHILDS)
Age when first child was born (AGEKDBRN)
Years with current job (YEARSJOB)
Number of days per month working extra hours (MOREDAYS)
Number of people working for company (NUMORG)

a. Test the model's validity.
b. Test each of the slope coefficients.

AMERICAN NATIONAL ELECTION SURVEY EXERCISES

17.21 ANES2008* With voter turnout during presidential elections around 50%, a vital task for politicians is to try to predict who will actually vote. Develop a regression model to predict intention to vote (DEFINITE) using the following demographic independent variables:

Age (AGE)
Education (EDUC)
Income (INCOME)

a. Determine the regression equation.
b. Test the model's validity.
c. Test to determine whether there is sufficient evidence to infer a linear relationship between the dependent variable and each independent variable.

17.22 ANES2008* Does watching news on television or reading newspapers provide indicators of who will vote? Conduct a regression analysis with intention to vote (DEFINITE) as the dependent variable and the following independent variables:

Number of days in previous week watching national news on television (DAYS1)
Number of days in previous week watching local television news in afternoon or early evening (DAYS2)
Number of days in previous week watching local television news in late evening (DAYS3)
Number of days in previous week reading a daily newspaper (DAYS4)
Number of days in previous week reading a daily newspaper on the Internet (DAYS5)
Number of days in previous week listening to news on radio (DAYS6)

a. Compute the regression equation.
b. Is there enough evidence to conclude that the model is valid?
c. Test each slope coefficient.

17.3 REGRESSION DIAGNOSTICS—II

In Section 16.7, we discussed how to determine whether the required conditions are unsatisfied. The same procedures can be used to diagnose problems in the multiple regression model. Here is a brief summary of the diagnostic procedure we described in Chapter 16.

Calculate the residuals and check the following:

1. *Is the error variable nonnormal?* Draw the histogram of the residuals.

2. *Is the error variance constant?* Plot the residuals versus the predicted values of y.

3. *Are the errors independent (time-series data)?* Plot the residuals versus the time periods.

4. *Are there observations that are inaccurate or do not belong to the target population?* Double-check the accuracy of outliers and influential observations.

If the error is nonnormal and/or the variance is not a constant, several remedies can be attempted. These are beyond the level of this book.

Outliers and influential observations are checked by examining the data in question to ensure accuracy.

Nonindependence of a time series can sometimes be detected by graphing the residuals and the time periods and looking for evidence of autocorrelation. In Section 17.4, we introduce the Durbin–Watson test, which tests for one form of autocorrelation. We will offer a corrective measure for nonindependence.

There is another problem that is applicable to multiple regression models only. *Multicollinearity* is a condition wherein the independent variables are highly correlated. Multicollincarity distorts the *t*-tests of the coefficients, making it difficult to determine whether any of the independent variables are linearly related to the dependent variable. It also makes interpreting the coefficients problematic. We will discuss this condition and its remedy next.

Multicollinearity

Multicollinearity (also called *collinearity* and *intercorrelation*) is a condition that exists when the independent variables are correlated with one another. The adverse effect of multicollinearity is that the estimated regression coefficients of the independent variables that are correlated tend to have large sampling errors. There are two consequences of multicollinearity. First, because the variability of the coefficients is large, the sample coefficient may be far from the actual population parameter, including the possibility that the statistic and parameter may have opposite signs. Second, when the coefficients are tested, the *t*-statistics will be small, which leads to the inference that there is no linear relationship between the affected independent variables and the dependent variable. In some cases, this inference will be wrong. Fortunately, multicollinearity does not affect the *F*-test of the analysis of variance.

Consider the chapter-opening example where we found that age and years with current employer were not statistically significant at the 5% significance level. However, if we test the coefficient of correlation between income and age and between income and years with current employer, both will be statistically significant. The Excel printout is shown below. How do we explain the apparent contradiction between the multiple regression *t*-tests of the coefficients of age and of years with current employer and the results of the *t*-test of the correlation coefficients? The answer is multicollinearity.

	A	B
1	**Correlation (Pearson)**	
2		
3	*INCOME and AGE*	
4	Pearson Coefficient of Correlation	0.1883
5	t Stat	3.2083
6	df	280
7	P(T<=t) one tail	0.0007
8	t Critical one tail	1.6503
9	P(T<=t) two tail	0.0015
10	t Critical two tail	1.9685

	A	B
1	**Correlation (Pearson)**	
2		
3	*INCOME and CUREMPYR*	
4	Pearson Coefficient of Correlation	0.1972
5	t Stat	3.3652
6	df	280
7	P(T<=t) one tail	0.0004
8	t Critical one tail	1.6503
9	P(T<=t) two tail	0.0009
10	t Critical two tail	1.9685

There is a relatively high degree of correlation between age and years at current job. This should not be surprising because it is not likely that young people will have been at the same job for many years. As a result, multicollinearity affected the results of the multiple regression t-tests so that it appears that both age and years at current job are not significantly significant when, in fact, both variables are linearly related to income.

Another problem caused by multicollinearity is the interpretation of the coefficients. We interpret the coefficients as measuring the change in the dependent variable when the corresponding independent variable increases by one unit while all the other independent variables are held constant. This interpretation may be impossible when the independent variables are highly correlated because when the independent variable increases by one unit, some or all of the other independent variables will change.

This raises two important questions for the statistics practitioner. First, how do we recognize the problem of multicollinearity when it occurs? Second, how do we avoid or correct it?

Multicollinearity exists in virtually all multiple regression models. In fact, finding two completely uncorrelated variables is rare. The problem becomes serious, however, only when two or more independent variables are highly correlated. Unfortunately, we do not have a critical value that indicates when the correlation between two independent variables is large enough to cause problems. To complicate the issue, multicollinearity also occurs when a combination of several independent variables is correlated with another independent variable or with a combination of other independent variables. Consequently, even with access to all the correlation coefficients, determining when the multicollinearity problem has reached the serious stage may be extremely difficult. A good indicator of the problem is a large F-statistic but small t-statistics.

Minimizing the effect of multicollinearity is often easier than correcting it. The statistics practitioner must try to include independent variables that are independent of each other. Another alternative is to use a stepwise regression package. *Forward stepwise regression* brings independent variables into the equation one at a time. Only if an independent variable improves the model's fit is it included. If two variables are strongly correlated, the inclusion of one of them in the model makes the second one unnecessary. *Backward stepwise regression* starts with all the independent variables included in the equation and removes variables if they are not strongly related to the dependent variable. Because the stepwise technique excludes redundant variables, it minimizes multicollinearity. Stepwise regression is presented in Chapter 18.

EXERCISES

The following exercises require a computer and software.

17.23 Compute the residuals and the predicted values for the regression analysis in Exercise 17.1.

a. Is the normality requirement violated? Explain.
b. Is the variance of the error variable constant? Explain.

17.24 Calculate the coefficients of correlation for each pair of independent variables in Exercise 17.1. What do these statistics tell you about the independent variables and the *t*-tests of the coefficients?

17.25 Refer to Exercise 17.2.
a. Determine the residuals and predicted values.
b. Does it appear that the normality requirement is violated? Explain.
c. Is the variance of the error variable constant? Explain.
d. Determine the coefficient of correlation between the assignment mark and the midterm mark. What does this statistic tell you about the *t*-tests of the coefficients?

17.26 Compute the residuals and predicted values for the regression analysis in Exercise 17.3.
a. Does it appear that the error variable is not normally distributed?
b. Is the variance of the error variable constant?
c. Is multicollinearity a problem?

17.27 Refer to Exercise 17.4. Find the coefficients of correlation of the independent variables.
a. What do these correlations tell you about the independent variables?
b. What do they say about the *t*-tests of the coefficients?

17.28 Calculate the residuals and predicted values for the regression analysis in Exercise 17.5.
a. Does the error variable appear to be normally distributed?
b. Is the variance of the error variable constant?
c. Is multicollinearity a problem?

17.29 Are the required conditions satisfied in Exercise 17.6?

17.30 Refer to Exercise 17.7.
a. Conduct an analysis of the residuals to determine whether any of the required conditions are violated.
b. Does it appear that multicollinearity is a problem?
c. Identify any observations that should be checked for accuracy.

17.31 Are the required conditions satisfied for the regression analysis in Exercise 17.8?

17.32 Determine whether the required conditions are satisfied in Exercise 17.9

17.33 Refer to Exercise 17.10. Calculate the residuals and predicted values.
a. Is the normality requirement satisfied?
b. Is the variance of the error variable constant?
c. Is multicollinearity a problem?

17.34 Determine whether there are violations of the required conditions in the regression model used in Exercise 17.11.

17.35 Determine whether the required conditions are satisfied in Exercise 17.12.

17.36 Refer to Exercise 17.13.
a. Are the required conditions satisfied?
b. Is multicollinearity a problem? If so, explain the consequences.

17.37 Refer to Exercise 17.14. Are the required conditions satisfied?

17.38 Refer to Exercise 17.15. Check the required conditions.

17.4 Regression Diagnostics—III (Time Series)

In Chapter 16, we pointed out that, in general, we check to see whether the errors are independent when the data constitute a *times series*—data gathered sequentially over a series of time periods. In Section 16.6, we described the graphical procedure for determining whether the required condition that the errors are independent is violated. We plot the residuals versus the time periods and look for patterns. In this section, we augment that procedure with the **Durbin–Watson test**.

Durbin–Watson Test

The Durbin–Watson test allows the statistics practitioner to determine whether there is evidence of **first-order autocorrelation**—a condition in which a relationship exists between consecutive residuals e_i and e_{i-1}, where i is the time period. The Durbin–Watson statistic is defined as

$$d = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

The range of the values of d is

$$0 \le d \le 4$$

where small values of d ($d < 2$) indicate a positive first-order autocorrelation and large values of d ($d > 2$) imply a negative first-order autocorrelation. Positive first-order autocorrelation is a common occurrence in business and economic time series. It occurs when consecutive residuals tend to be similar. In that case, $(e_i - e_{i-1})^2$ will be small, producing a small value for d. Negative first-order autocorrelation occurs when consecutive residuals differ widely. For example, if positive and negative residuals generally alternate, $(e_i - e_{i-1})^2$ will be large; as a result, d will be greater than 2. Figures 17.2 and 17.3 depict positive first-order autocorrelation, whereas Figure 17.4 illustrates negative autocorrelation. Notice that in Figure 17.2 the first residual is a small number; the second residual, also a small number, is somewhat larger; and that trend continues. In Figure 17.3, the first residual is large and, in general, succeeding residuals decrease. In both figures, consecutive residuals are similar. In Figure 17.4, the first residual is a positive number and is followed by a negative residual. The remaining residuals follow this pattern (with some exceptions). Consecutive residuals are quite different.

FIGURE **17.2** **Positive First-Order Autocorrelation**

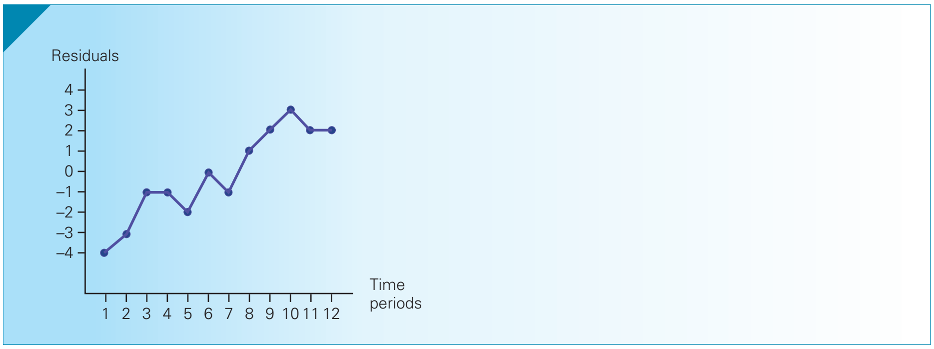


FIGURE **17.3** **Positive First-Order Autocorrelation**

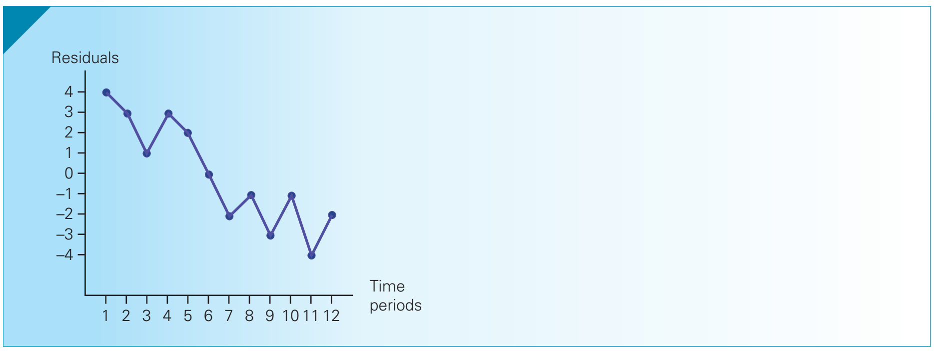


FIGURE 17.4 Negative First-Order Autocorrelation

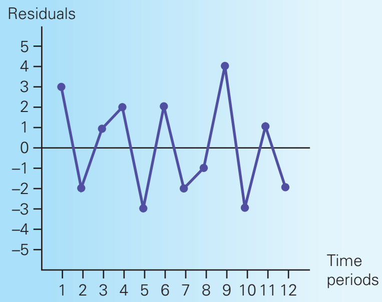


Table 8 in Appendix B is designed to test for positive first-order autocorrelation by providing values of d_L and d_U for a variety of values of n and k and for $\alpha = .01$ and .05.

The decision is made in the following way. If $d < d_L$, we conclude that there is enough evidence to show that positive first-order autocorrelation exists. If $d > d_U$, we conclude that there is not enough evidence to show that positive first-order autocorrelation exists. And if $d_L \le d \le d_U$, the test is inconclusive. The recommended course of action when the test is inconclusive is to continue testing with more data until a conclusive decision can be made.

For example, to test for positive first-order autocorrelation with $n = 20$, $k = 3$, and $\alpha = .05$, we test the following hypotheses:

H_0: There is no first-order autocorrelation.

H_1: There is positive first-order autocorrelation.

The decision is made as follows:

If $d < d_L = 1.00$, reject the null hypothesis in favor of the alternative hypothesis.

If $d > d_U = 1.68$, do not reject the null hypothesis.

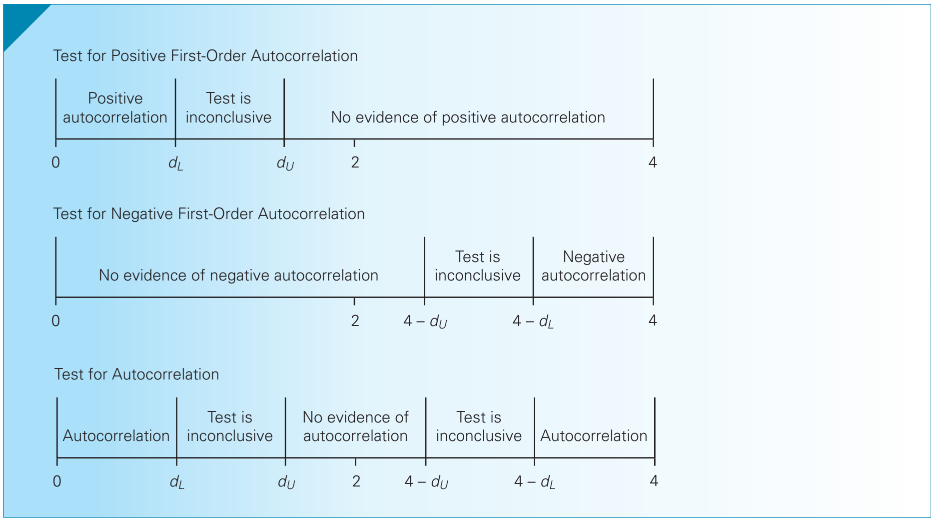
If $1.00 \le d \le 1.68$, the test is inconclusive.

To test for negative first-order autocorrelation, we change the critical values. If $d > 4 - d_L$, we conclude that negative first-order autocorrelation exists. If $d < 4 - d_U$, we conclude that there is not enough evidence to show that negative first-order autocorrelation exists. If $4 - d_U \le d \le 4 - d_L$, the test is inconclusive.

We can also test simply for first-order autocorrelation by combining the two one-tail tests. If $d < d_L$ or $d > 4 - d_L$, we conclude that autocorrelation exists. If $d_U \le d \le 4 - d_U$, we conclude that there is no evidence of autocorrelation. If $d_L \le d \le d_U$ or $4 - d_U \le d \le 4 - d_L$, the test is inconclusive. The significance level will be 2α (where α is the one-tail significance level). Figure 17.5 describes the range of values of d and the conclusion for each interval.

For time-series data, we add the Durbin–Watson test to our list of regression diagnostics. In other words, we determine whether the error variable is normally distributed with constant variance (as we did in Section 17.3), we identify outliers and (if our software allows it) influential observations that should be verified, and we conduct the Durbin–Watson test.

FIGURE 17.5 Durbin-Watson Test



EXAMPLE 17.1

DATA
Xm17-01

Christmas Week Ski Lift Sales

Christmas week is a critical period for most ski resorts. Because many students and adults are free from other obligations, they are able to spend several days indulging in their favorite pastime, skiing. A large proportion of gross revenue is earned during this period. A ski resort in Vermont wanted to determine the effect that weather had on its sales of lift tickets. The manager of the resort collected data on the number of lift tickets sold during Christmas week (y), the total snowfall in inches (x_1), and the average temperature in degrees Fahrenheit (x_2) for the past 20 years. Develop the multiple regression model and diagnose any violations of the required conditions.

SOLUTION

The model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

EXCEL

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.3465				
5	R Square	0.1200				
6	Adjusted R Square	0.0165				
7	Standard Error	1712				
8	Observations	20				
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression	2	6,793,798	3,396,899	1.16	0.3373
13	Residual	17	49,807,214	2,929,836		
14	Total	19	56,601,012			
15						
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
17	Intercept	8308	904	9.19	5.24E-08	
18	Snowfall	74.59	51.57	1.45	0.1663	
19	Temperature	–8.75	19.70	–0.44	0.6625	

MINITAB

Regression Analysis: Tickets versus Snowfall, Temperature

The regression equation is
Tickets = 8308 + 74.6 Snowfall − 8.8 Temperature

Predictor	Coef	SE Coef	T	P
Constant	8308.0	903.7	9.19	0.000
Snowfall	74.59	51.57	1.45	0.166
Temperature	−8.75	19.70	−0.44	0.662

S = 1712   R-Sq = 12.0%   R-Sq(adj) = 1.7%

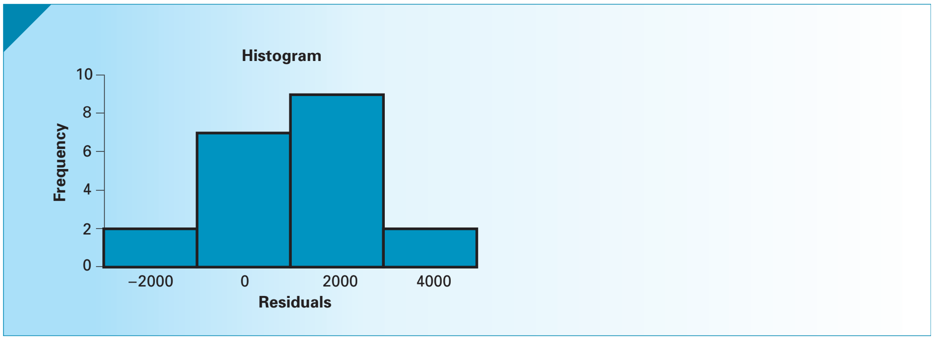
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	6,793,798	3,396,899	1.16	0.337
Residual Error	17	49,807,214	2,929,836		
Total	19	56,601,012			

INTERPRET

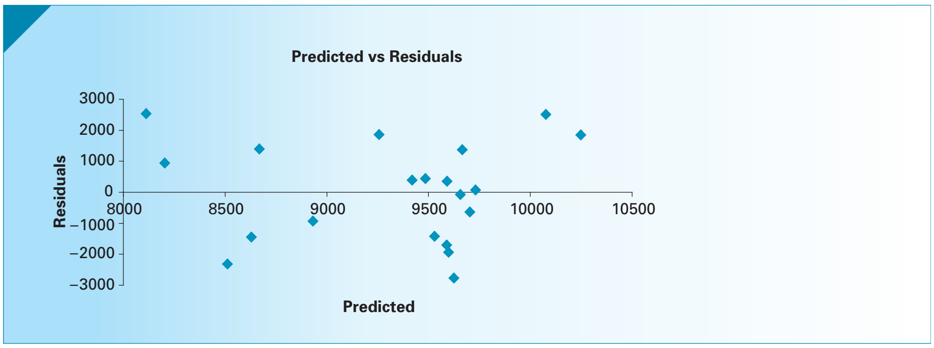
As you can see, the coefficient of determination is small ($R^2 = 12\%$) and the p-value of the F-test is .3373, both of which indicate that the model is poor. We used Excel to draw the histogram (Figure 17.6) of the residuals and plot the predicted values of y versus the residuals in Figure 17.7. Because the observations constitute a time series, we also used Excel to plot the time periods (years) versus the residuals (Figure 17.8).

FIGURE **17.6** Histogram of Residuals in Example 17.1



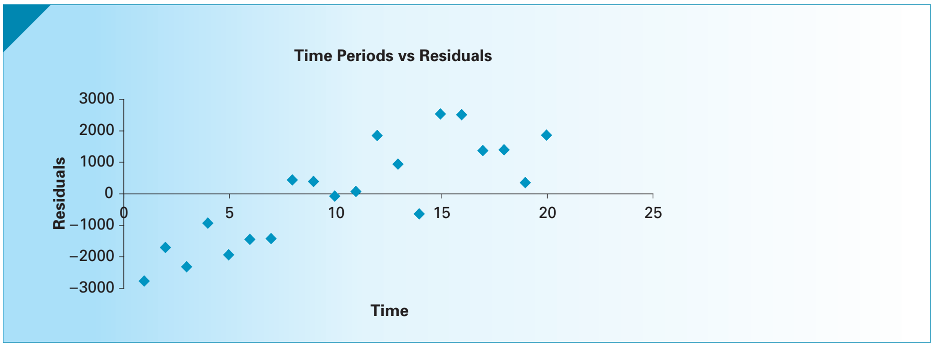
The histogram reveals that the error may be normally distributed.

FIGURE **17.7** Plot of Predicted Values versus Residuals in Example 17.1



There does not appear to be any evidence of heteroscedasticity.

FIGURE 17.8 Plot of Time Periods versus Residuals in Example 17.1



This graph reveals a serious problem. There is a strong relationship between consecutive values of the residuals, which indicates that the requirement that the errors are independent has been violated. To confirm this diagnosis, we instructed Excel and Minitab to calculate the Durbin–Watson statistic.

EXCEL

	A	B	C
1	Durbin–Watson Statistic		
2			
3	d = 0.5931		

INSTRUCTIONS

Proceed through the usual steps to conduct a regression analysis and print the residuals (see page 672). Highlight the entire list of residuals and click **Add-Ins, Data Analysis Plus**, and **Durbin–Watson Statistic**.

MINITAB

```
Durbin–Watson statistic = 0.593140
```

INSTRUCTIONS

Follow the instructions on page 673. Before clicking **OK**, click **Options . . .** and **Durbin–Watson statistic.**

The critical values are determined by noting that $n = 20$ and $k = 2$ (there are two independent variables in the model). If we wish to test for positive first-order autocorrelation with $\alpha = .05$, we find in Table 8(a) in Appendix B

$$d_L = 1.10 \quad \text{and} \quad d_U = 1.54$$

The null and alternative hypotheses are

H_0: There is no first-order autocorrelation.

H_1: There is positive first-order autocorrelation.

The rejection region is $d < d_L = 1.10$. Because $d = .59$, we reject the null hypothesis and conclude that there is enough evidence to infer that positive first-order autocorrelation exists.

Autocorrelation usually indicates that the model needs to include an independent variable that has a time-ordered effect on the dependent variable. The simplest such independent variable represents the time periods. To illustrate, we included a third independent variable that records the number of years since the year the data were gathered. Thus, $x_3 = 1, 2, \ldots, 20$. The new model is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

EXCEL

	A	B	C	D	E	F
1	SUMMARY OUTPUT					
2						
3	*Regression Statistics*					
4	Multiple R	0.8608				
5	R Square	0.7410				
6	Adjusted R Square	0.6924				
7	Standard Error	957				
8	Observations	20				
9						
10	ANOVA					
11		*df*	*SS*	*MS*	*F*	*Significance F*
12	Regression	3	41,940,217	13,980,072	15.26	0.0001
13	Residual	16	14,660,795	916,300		
14	Total	19	56,601,012			
15						
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	
17	Intercept	5966	631.3	9.45	6.00E-08	
18	Snowfall	70.18	28.85	2.43	0.0271	
19	Temperature	−9.23	11.02	−0.84	0.4145	
20	Time	230.0	37.13	6.19	1.29E-05	

MINITAB

Regression Analysis: Tickets versus Snowfall, Temperature, Time

The regression equation is
Tickets = 5966 + 70.2 Snowfall − 9.2 Temperature + 230 Time

Predictor	Coef	SE Coef	T	P
Constant	5965.6	631.3	9.45	0.000
Snowfall	70.18	28.85	2.43	0.027
Temperature	−9.23	11.02	−0.84	0.414
Time	229.97	37.13	6.19	0.000

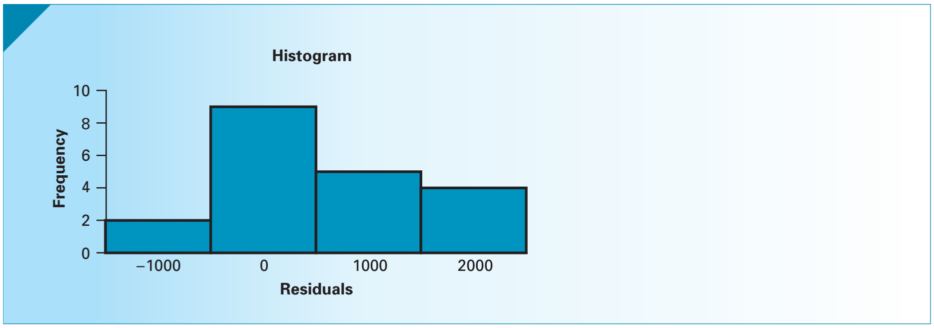
S = 957.2 R-Sq = 74.1% R-Sq(adj) = 69.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	41,940,217	13,980,072	15.26	0.000
Residual Error	16	14,660,795	916,300		
Total	19	56,601,012			

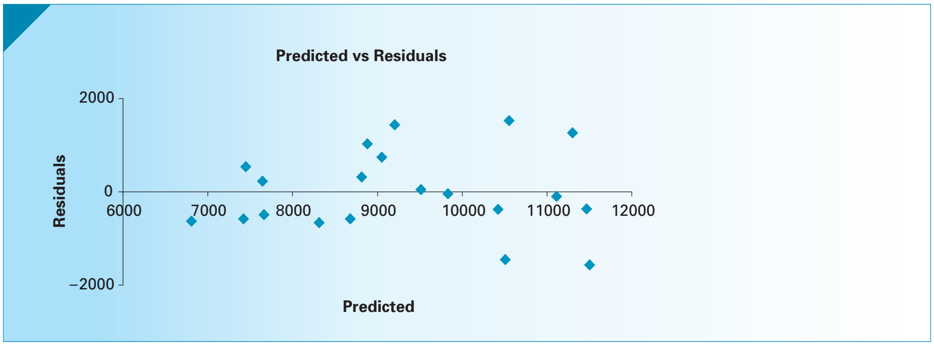
As we did before, we calculate the residuals and conduct regression diagnostics using Excel. The results are shown in Figures 17.9–17.11.

FIGURE **17.9** Histogram of Residuals in Example 17.1 (Time variable included)



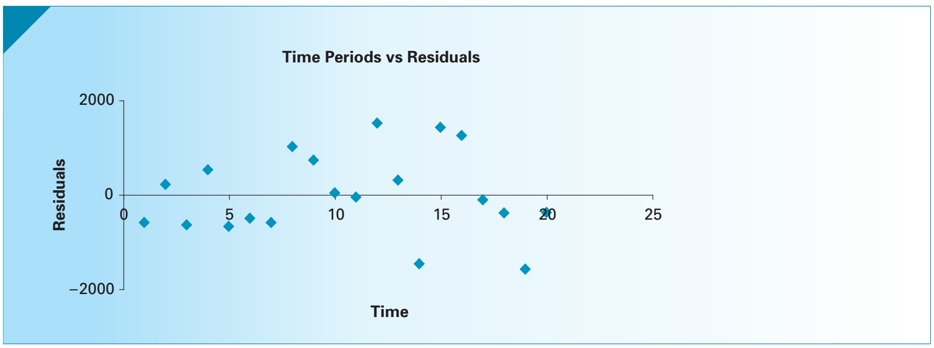
The histogram reveals that the error may be normally distributed.

FIGURE **17.10** Plot of Predicted Values versus Residuals in Example 17.1 (Time variable included)



The error variable variance appears to be constant.

FIGURE **17.11** Plot of Time Periods versus Residuals in Example 17.1 (Time variable included)



There is no sign of autocorrelation. To confirm our diagnosis, we conducted the Durbin–Watson test.

EXCEL

	A	B	C
1	Durbin–Watson Statistic		
2			
3	d = 1.885		

MINITAB

```
Durbin–Watson statistic = 1.88499
```

From Table 8(a) in Appendix B, we find the critical values of the Durbin–Watson test. With $k = 3$ and $n = 20$, we find

$$d_L = 1.00 \quad \text{and} \quad d_U = 1.68$$

Because $d > 1.68$, we conclude that there is not enough evidence to infer the presence of positive first-order autocorrelation.

Notice that the model is improved dramatically. The *F*-test tells us that the model is valid. The *t*-tests tell us that both the amount of snowfall and time are significantly linearly related to the number of lift tickets. This information could prove useful in advertising for the resort. For example, the resort could emphasize any recent snowfall in its advertising. If no new snow has fallen, the resort might emphasize its snow-making facilities.

Developing an Understanding of Statistical Concepts

Notice that the addition of the time variable explained a large proportion of the variation in the number of lift tickets sold; that is, the resort experienced a relatively steady increase in sales over the past 20 years. Once this variable was included in the model, the amount of snowfall became significant because it was able to explain some of the remaining variation in lift ticket sales. Without the time variable, the amount of snowfall and the temperature were unable to explain a significant proportion of the variation in ticket sales. The graph of the residuals versus the time periods and the Durbin–Watson test enabled us to identify the problem and correct it. In overcoming the autocorrelation problem, we improved the model so that we identified the amount of snowfall as an important variable in determining ticket sales. This result is quite common. Correcting a violation of a required condition will frequently improve the model.

Exercises

17.39 Perform the Durbin–Watson test at the 5% significance level to determine whether positive first-order autocorrelation exists when $d = 1.10$, $n = 25$, and $k = 3$.

17.40 Determine whether negative first-order autocorrelation exists when $d = 2.85$, $n = 50$, and $k = 5$. (Use a 1% significance level.)

17.41 Given the following information, perform the Durbin–Watson test to determine whether first-order autocorrelation exists.

$n = 25 \quad k = 5 \quad \alpha = .10 \quad d = .90$

17.42 Test the following hypotheses with $\alpha = .05$.

H_0: There is no first-order autocorrelation.
H_1: There is positive first-order autocorrelation.
$n = 50 \quad k = 2 \quad d = 1.38$

17.43 Test the following hypotheses with $\alpha = .02$.

H_0: There is no first-order autocorrelation.
H_1: There is first-order autocorrelation.
$n = 90 \quad k = 5 \quad d = 1.60$

17.44 Test the following hypotheses with $\alpha = .05$.

H_0: There is no first-order autocorrelation.
H_1: There is negative first-order autocorrelation.
$n = 33 \quad k = 4 \quad d = 2.25$

The following exercises require a computer and software.

17.45 Xr17-45 Observations of variables y, x_1, and x_2 were taken over 100 consecutive time periods.

a. Conduct a regression analysis of these data.
b. Plot the residuals versus the time periods. Describe the graph.
c. Perform the Durbin–Watson test. Is there evidence of autocorrelation? Use $\alpha = .10$.
d. If autocorrelation was detected in part (c), propose an alternative regression model to remedy the problem. Use the computer to generate the statistics associated with this model.
e. Redo parts (b) and (c). Compare the two models.

17.46 Xr17-46 Weekly sales of a company's product (y) and those of its main competitor (x) were recorded for one year.

a. Conduct a regression analysis of these data.
b. Plot the residuals versus the time periods. Does there appear to be autocorrelation?
c. Perform the Durbin–Watson test. Is there evidence of autocorrelation? Use $\alpha = .10$.
d. If autocorrelation was detected in part (c), propose an alternative regression model to remedy the problem. Use the computer to generate the statistics associated with this model.
e. Redo parts (b) and (c). Compare the two models.

17.47 Refer to Exercise 17.3. Is there evidence of positive first-order autocorrelation?

17.48 Refer to Exercise 16.99. Determine whether there is evidence of first-order autocorrelation.

17.49 Xr17-49 The manager of a tire store in Minneapolis has been concerned with the high cost of inventory. The current policy is to stock all the snow tires that are predicted to sell over the entire winter at the beginning of the season (end of October). The manager can reduce inventory costs by having suppliers deliver snow tires regularly from October to February. However, he needs to be able to predict weekly sales to avoid stockouts that will ultimately lose sales. To help develop a forecasting model, he records the number of snow tires sold weekly during the last winter and the amount of snowfall (in inches) in each week.

a. Develop a regression model and use a software package to produce the statistics.
b. Perform a complete diagnostic analysis to determine whether the required conditions are satisfied.
c. If one or more conditions are unsatisfied, attempt to remedy the problem.
d. Use whatever procedures you wish to assess how well the new model fits the data.
e. Interpret and test each of the coefficients.

CHAPTER SUMMARY

The multiple regression model extends the model introduced in Chapter 16. The statistical concepts and techniques are similar to those presented in simple linear regression. We assess the model in three ways: standard error of estimate, the coefficient of determination (and the coefficient of determination adjusted for degrees of freedom), and the F-test of the analysis of variance. We can use the t-tests of the coefficients to determine whether each of the independent variables is linearly related to the dependent variable. As we did in Chapter 16, we showed how to diagnose violations of the required conditions and to identify other problems. We introduced multicollinearity and demonstrated its effect and its remedy. Finally, we presented the Durbin–Watson test to detect first-order autocorrelation.

IMPORTANT TERMS

Response surface 694
Coefficient of determination adjusted for degrees of freedom 698
Multicollinearity 714
Durbin–Watson test 716
First-order autocorrelation 716

SYMBOLS

Symbol	Pronounced	Represents
β_i	Beta sub *i* or beta *i*	Coefficient of *i*th independent variable
b_i	*b* sub *i* or *b i*	Sample coefficient

FORMULAS

Standard error of estimate

$$s_\varepsilon = \sqrt{\frac{\text{SSE}}{n-k-1}}$$

Test statistic for β_i

$$t = \frac{b_i - \beta_i}{s_{b_i}}$$

Coefficient of determination

$$R^2 = \frac{s_{xy}^2}{s_x^2 s_y^2} = 1 - \frac{\text{SSE}}{\sum (y_i - \bar{y})^2}$$

Adjusted coefficient of determination

$$\text{Adjusted } R^2 = 1 - \frac{\text{SSE}/(n-k-1)}{\sum (y_i - \bar{y})^2/(n-1)}$$

Mean square for error

$$\text{MSE} = \text{SSE}/k$$

Mean square for regression

$$\text{MSR} = \text{SSR}/(n-k-1)$$

F-statistic

$$F = \text{MSR}/\text{MSE}$$

Durbin–Watson statistic

$$d = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

COMPUTER OUTPUT AND INSTRUCTIONS

Technique	Excel	Minitab
Regression	696	697
Prediction interval	706	706
Durbin–Watson statistic	721	721

CHAPTER EXERCISES

The following exercises require the use of a computer and statistical software. Use a 5% significance level.

17.50 Xr17-50 The agronomist referred to in Exercise 16.101 believed that the amount of rainfall as well as the amount of fertilizer used would affect the crop yield. She redid the experiment in the following way. Thirty greenhouses were rented. In each, the amount of fertilizer and the amount of water were varied. At the end of the growing season, the amount of corn was recorded.

a. Determine the sample regression line, and interpret the coefficients.
b. Do these data allow us to infer that there is a linear relationship between the amount of fertilizer and the crop yield?
c. Do these data allow us to infer that there is a linear relationship between the amount of water and the crop yield?
d. What can you say about the fit of the multiple regression model?

e. Is it reasonable to believe that the error variable is normally distributed with constant variance?
f. Predict the crop yield when 100 kilograms of fertilizer and 1,000 liters of water are applied. Use a confidence level of 95%.

17.51 Xr16-12* Exercise 16.12 addressed the problem of determining the relationship between the price of apartment buildings and number of square feet. Hoping to improve the predictive capability of the model the real estate agent also recorded the number of apartments, the age, and the number of floors.
a. Calculate the regression equation.
b. Is the model valid?
c. Compare your answer with that of Exercise 16.12.

17.52 Xr16-16* In Exercise 16.16, a statistics practitioner examined the relationship between office rents and the city's office vacancy rate. The model appears to be quite poor. It was decided to add another variable that measures the state of the economy. The city's unemployment rate was chosen for this purpose.
a. Determine the regression equation.
b. Determine the coefficient of determination and describe what this value means.
c. Test the model's validity in explaining office rent.
d. Determine which of the two independent variables is linearly related to rents.
e. Determine whether the error is normally distributed with a constant variance.
f. Determine whether there is evidence of autocorrelation.
g. Predict with 95% confidence the office rent in a city whose vacancy rate is 10% and whose unemployment rate is 7%.

CASE 17.1 An Analysis of Mutual Fund Managers, Part 1*

© Vicki Beaver

DATA C17-01a

There are thousands of mutual funds available (see page 181 for a brief introduction to mutual funds). There is no shortage of sources of information about them. Newspapers regularly report the value of each unit, mutual fund companies and brokers advertise extensively, and there are books on the subject. Many of the advertisements imply that individuals should invest in the advertiser's mutual fund because it has performed well in the past. Unfortunately, there is little evidence to infer that past performance is a predictor of the future. However, it may be possible to acquire useful information by examining the managers of mutual funds. Several researchers have studied the issue. One project gathered data concerning the performance of 2,029 funds.

The performance of each fund was measured by its risk-adjusted excess return, which is the difference between the return on investment of the fund and a return that is considered a standard. The standard is based on a variety of variables, including the risk-free rate.

Four variables describe the fund manager: age, tenure (how many years the manager has been in charge), whether the manager had an MBA ($1 = $ yes, $0 = $ no), and a measure of the quality of the manager's education [the average Scholastic Achievement Test (SAT) score of students at the university where the manager received his or her undergraduate degree].

Conduct an analysis of the data. Discuss how the average SAT score of the manager's alma mater, whether he or she has an MBA, and his or her age and tenure are related to the performance of the fund.

*This case is based on "Are Some Mutual Fund Managers Better Than Others? Cross-Sectional Patterns in Behavior and Performance," Judith Chevalier and Glenn Ellison, Working Paper 5852, National Bureau of Economic Research.

CASE 17.2 An Analysis of Mutual Fund Managers, Part 2


© Rtimages/Shutterstock

DATA
C17-02a
C17-02b

In addition to analyzing the relationship between the managers' characteristic and the performance of the fund, researchers wanted to determine whether the same characteristics are related to the behavior of the fund. In particular, they wanted to know whether the risk of the fund and its management expense ratio (MER) were related to the manager's age, tenure, university SAT score, and whether he or she had an MBA.

In Section 4.6, we introduced the market model wherein we measure the systematic risk of stocks by the stock's beta. The beta of a portfolio is the average of the betas of the stocks that make up the portfolio. File C17-02a stores the same managers' characteristics as those in file C17-01. However, the first column contains the betas of the mutual funds.

To analyze the management expense ratios, it was decided to include a measure of the size of the fund. The logarithm of the funds' assets (in $millions) was recorded with the MER. These data are stored in file C17-02b.

Analyze both sets of data and write a brief report of your findings.

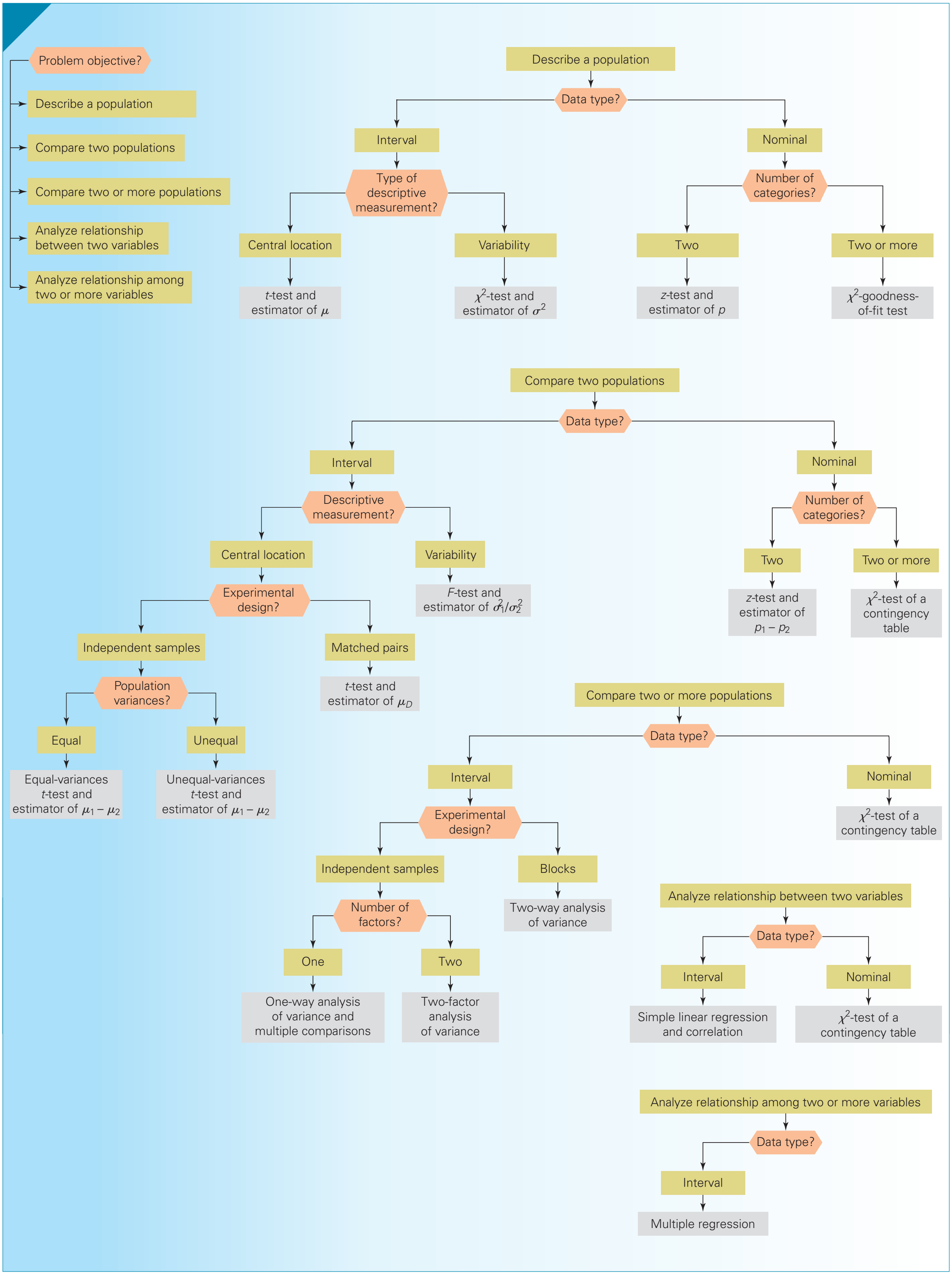
APPENDIX 17 / REVIEW OF CHAPTERS 12 TO 17

Table A17.1 presents a list of inferential methods presented thus far, and Figure A17.1 depicts a flowchart designed to help students identify the correct statistical technique.

TABLE A17.1 Summary of Statistical Techniques in Chapters 12 to 17

t-test of μ
Estimator of μ (including estimator of $N\mu$)
χ^2 test of σ^2
Estimator of σ^2
z-test of p
Estimator of p (including estimator of Np)
Equal-variances t-test of $\mu_1 - \mu_2$
Equal-variances estimator of $\mu_1 - \mu_2$
Unequal-variances t-test of $\mu_1 - \mu_2$
Unequal-variances estimator of $\mu_1 - \mu_2$
t-test of μ_D
Estimator of μ_D
F-test of σ_1^2/σ_2^2
Estimator of σ_1^2/σ_2^2
z-test of $p_1 - p_2$ (Case 1)
z-test of $p_1 - p_2$ (Case 2)
Estimator of $p_1 - p_2$
One-way analysis of variance (including multiple comparisons)
Two-way (randomized blocks) analysis of variance
Two-factor analysis of variance
χ^2-goodness-of-fit test
χ^2-test of a contingency table
Simple linear regression and correlation (including t-tests of β_1 and ρ, and prediction and confidence intervals)
Multiple regression (including t-tests of β_i, F-test, and prediction and confidence intervals)

FIGURE A17.1 Flowchart of Techniques in Chapters 12 to 17



EXERCISES

A17.1 XrA17-01 Garlic has long been considered a remedy to ward off the common cold. A British researcher organized an experiment to see if this generally held belief is true. A random sample of 146 volunteers was recruited. Half the sample took one capsule of an allicin-containing garlic supplement each day. The others took a placebo. The results for each volunteer after the winter months were recorded in the following way.

Column
1. Identification number
2. 1 = allicin-containing capsule; 2 = placebo
3. Suffered a cold (1 = no, 2 = yes)
4. If individual caught a cold, the number of days until recovery (999 was recorded if no cold)

a. Can the researcher conclude that garlic does help prevent colds?
b. Does garlic reduce the number of days until recovery if a cold was caught?

A17.2 XrA17-02 Because shelf space is a limited resource for a retail store, product selection, shelf-space allocation, and shelf-space placement decisions must be made according to a careful analysis of profitability and inventory turnover. The manager of a chain of variety stores wishes to see whether shelf location affects the sales of a canned soup. She believes that placing the product at eye level will result in greater sales than will placing the product on a lower shelf. She observed the number of sales of the product in 40 different stores. Sales were observed over 2 weeks, with product placement at eye level one week and on a lower shelf the other week. Can we conclude that placement of the product at eye level significantly increases sales?

A17.3 XrA17-03 In an effort to explain the results of Exercise A15.9, a researcher recorded the distances for the random sample of British and American golf courses. Can we infer that British courses are shorter than American courses?

A17.4 XrA17-04 It is generally assumed that alcohol consumption tends to make drinkers more impulsive. However, a recent study in the journal *Alcohol and Alcoholism* may contradict this assumption. The study took a random sample of 76 male undergraduate students and divided them into three groups. One group remained sober; the second group was given flavored drinks with not enough alcohol to intoxicate; and the students in third group were intoxicated. Each student was offered a chance of receiving $15 at the end of the session or double that amount later. The results were recorded using the following format:

Column 1: Group number
Column 2: Code 1 = chose $15, 2 = chose $30 later

Do the data allow us to infer that there is a relationship between the choices students make and their level of intoxication?

A17.5 XrA17-05 Refer to Exercise 13.35. The executive did a further analysis by taking another random sample. This time she tracked the number of customers who have had an accident in the last 5 years. For each she recorded the total amount of repairs and the credit score. Do these data allow the executive to conclude that the higher the credit score the lower the cost of repairs will be?

A17.6 XrA17-06 The U.S. National Endowment for the Arts conducts surveys of American adults to determine, among other things, their participation in various arts activities. A recent survey asked a random sample of American adults whether they participate in photography. The responses are 1 = yes and 2 = no. There were 205.8 million American adults. Estimate with 95% confidence the number of American adults are participate in photography. (Adapted from the *Statistical Abstract of the United States, 2006*, Table 1228.)

A17.7 XrA17-07 Mouth-to-mouth resuscitation has long been considered better than chest compression for people who have suffered a heart attack. To determine if this indeed is the better way, Japanese researchers analyzed 4,068 adult patients who had cardiac arrest witnessed by bystanders. Of those, 439 received only chest compressions from bystanders and 712 received conventional CPR compressions and breaths. The results for each group was recorded where 1 = did not survive with good neurological function and 2 = did survive with good neurological function. What conclusions can be drawn from these data?

A17.8 XrA17-08 Refer to Exercise A15.6. The financial analyst undertook another project wherein respondents were also asked the age of the head of the household. The choices are

1. Younger than 25
2. 25 to 34
3. 35 to 44
4. 45 to 54
5. 55 to 64
6. 65 and older

The responses to questions about ownership of mutual funds is No = 1 and Yes = 2. Do these data allow us to infer that the age of the head of the household is related to whether he or she owns mutual funds? (*Source:* Adapted from the *Statistical Abstract of the United States, 2006*, Table 1200.)

A17.9 XrA17-09 Over one decade (1995–2005), the number of hip and knee replacement surgeries increased by 87%. Because the costs of hip and knee replacements are so expensive, private health-insurance and government-operated health-care plans have become more concerned. To get more information, random samples of people who had hip replacements in 1995 and in 2005 were drawn. From the files, the ages of the patients were recorded. Is there enough evidence to infer that the ages of people who require hip replacements are getting smaller? (*Source:* Canadian Joint Replacement Registry.)

A17.10 XrA17-10 Refer to Exercise A17.9. Weight is a major factor that determines whether a person will need a hip or knee replacement and at what age. To learn more about the topic, a medical researcher randomly sampled individuals who had hip replacement (code = 1) and knee replacement (code = 2) and one of the following categories:

1. Underweight 2. Normal range
3. Overweight but not obese 4. Obese

Do the data allow the researcher to conclude that weight and the joint needing replacement are related?

A17.11 XrA17-11 Television shows with large amounts of sex or violence tend to attract more viewers. Advertisers want large audiences, but they also want viewers to remember the brand names of their products. A study was undertaken to determine the effect that shows with sex and violence have on their viewers. A random sample of 328 adults was divided into three groups. Group 1 watched violent programs, group 2 watched sexually explicit shows, and group 3 watched neutral shows. The researchers spliced nine 30-second commercials for a wide range of products. After the show, the subjects were quizzed to see if they could recall the brand name of the products. They were also asked to name the brands 24 hours later. The number of correct answers was recorded. Conduct a test to determine whether differences exist between the three groups of viewers and which type of program does best in brand recall. Results were published in the *Journal of Applied Psychology* (*National Post*, August 16, 2004).

A17.12 XrA17-12 In an effort to explain to customers why their electricity bills have been so high lately, and how customers could save money by reducing the thermostat settings on both space heaters and water heaters, a public utility commission has collected total kilowatt consumption figures for last year's winter months, as well as thermostat settings on space and water heaters, for 100 homes.

a. Determine the regression equation.
b. Determine the coefficient of determination and describe what it tells you.
c. Test the validity of the model.
d. Find the 95% interval of the electricity consumption of a house whose space heater thermostat is set at 70 and whose water heater thermostat is set at 130.
e. Calculate the 95% interval of the average electricity consumption for houses whose space heater thermostat is set at 70 and whose water heater thermostat is set at 130.

A17.13 XrA17-13 An economist wanted to learn more about total compensation packages. She conducted a survey of 858 workers and asked all to report their hourly wages or salaries, their total benefits, and whether the companies they worked for produced goods or services. Determine whether differences exist between goods-producing and services-producing firms in terms of hourly wages and total benefits. (Adapted from the *Statistical Abstract of the United States, 2006*, Table 637.)

A17.14 XrA17-14 Professional athletes in North America are paid very well for their ability to play games that amateurs play for fun. To determine the factors that influence a team to pay a hockey player's salary, an MBA student randomly selected 50 hockey players who played in the 1992–1993 and 1993–1994 seasons. He recorded their salaries at the end of the 1993–1994 season as well as a number of performance measures in the previous two seasons. The following data were recorded.

Columns 1 and 2: Games played in 1992–1993 and 1993–1994
Columns 3 and 4: Goals scored in 1992–1993 and 1993–1994
Columns 5 and 6: Assists recorded in 1992–1993 and 1993–1994
Columns 7 and 8: Plus/minus score in 1992–1993 and 1993–1994
Columns 9 and 10: Penalty minutes served in 1992–1993 and 1993–1994
Column 11: Salary in U.S. dollars

(Plus/minus is the number of goals scored by the player's team minus the number of goals scored by the opposing team while the player is on the ice.) Develop a model that analyzes the relationship between salary and the performance measures. Describe your findings. (The author wishes to thank Gordon Barnett for writing this exercise.)

A17.15 XrA17-15 The risks associated with smoking are well known. Virtually all physicals recommend that their patients quit. This raises the question, What are the risks for people who quit smoking compared to continuing smokers and those who have never smoked? In a study described in the *Journal of Internal Medicine* [Feb. 2004, 255(2): 266–272], researchers took samples of each of the following groups.

Group 1: Never smokers
Group 2: Continuing smokers
Group 3: Smokers who quit

At the beginning of the 10-year research project, there were 238 people who had never smoked and 155 smokers. Over the year, 39 smokers quit. The weight gain, increase in systolic (SBP) blood pressure, and increase in diastolic (DBP) blood pressure were measured and recorded. Determine whether differences exist between the three groups in terms of weight gain, increases in systolic blood pressure, and increases in diastolic blood pressure and which groups differ.

A17.16 XrA17-16 A survey was conducted among Canadian farmers, who were each asked to report the number of acres in his or her farm. There were a total of 229,373 farms in Canada in 2006 (*Source:* Statistics Canada). Estimate with 95% confidence the total amount of area (in acres) that was farmed in Canada in 2006.

General Social Survey Exercises

A17.17 GSS2008* Do the data allow us to infer that households with at least one union member (UNION: 1 = Respondent belongs, 2 = Spouse belongs, 3 = Both belong, 4 = Neither belong) differ from households with no union members with respect to their position on whether the government should do more or less to solve the country's problems (HELPNOT: 1 = Government should do more; 2, 3, 4, 5 = Government does too much)?

A17.18 GSS2008* Estimate with 95% confidence the proportion of Americans who are divorced (DIVORCE: 1 = Yes, 2 = No).

A17.19 GSS2008* Is there sufficient evidence to conclude that people who have taken college-level science courses (COLSCINM: 1 = Yes, 2 = No) are more likely to answer the following question correctly (ODDS1): A doctor tells a couple that there is one chance in four that their child will have an inherited disease. Does this mean that if the first child has the illness, the next three will not? 1 = Yes, 2 = No. Correct answer: No.

A17.20 GSS2008* Estimate with 95% confidence the mean job tenure (CUREMPYR).

A17.21 GSS2008* Is there sufficient evidence to infer that the three groups of conservatives (POLVIEWS: 5 = slightly conservative, 6 = conservative, 7 = extremely conservative) differ in support for capital punishment (CAPPUN: 1 = Favor, 2 = Oppose)?

A17.22 GSS2008* Do older people watch more television? To answer the question, analyze the relationship between age (AGE) and the amount of time spent watching television (TVHOURS).

A17.23 GSS2008* If a person has a higher income, is he or she more likely to believe that the government should do more to solve the country's problems? Conduct a test of the relationship between income (INCOME) and HELPNOT to answer the question.

A17.24 GSS2006* Is there enough evidence to conclude that people with vision problems (DISABLD2: Do you have a vision problem that prevents you from reading a newspaper even when wearing glasses or contacts: 1 = Yes, 2 = No) are more likely to believe that it is the government's responsibility to help pay for doctor and hospital bills (1 = Government should help; 2, 3 ,4, 5 = People should help themselves)?

A17.25 GSS2008* Do the children of men with prestigious occupations have prestigious occupations themselves? Conduct a test to determine whether there is a positive linear relationship between PRESTG80 and PAPRES80.

A17.26 GSS2008* Is there a relationship between the number of hours a husband works and the number of hours his wife works? Answer the question by conducting a test of the two variables (HRS and SPHRS).

American National Election Survey Exercises

A17.27 ANES2008* Are older people more likely to vote? One way to answer this question is to conduct a test to determine whether there is enough evidence to conclude that age (AGE) and intention to vote (DEFINITE) are positively related.

A17.28 ANES2008* Is the amount of time a person watches television news per day affected by his or her education? Test to determine whether TIME2 and EDUC are linearly related.

CASE A17.1 Testing a More Effective Device to Keep Arteries Open

DATA CA17-01

A stent is a metal mesh cylinder that holds a coronary artery open after a blockage has been removed. However, in many patients the stents, which are made from bare metal, become blocked as well. One cause of the reoccurrence of blockages is the body's rejection of the foreign object. In a study published in the *New England Journal of Medicine* (January 2004), a new polymer-based stent was tested. After insertion, the new stents slowly release a drug (paclitaxel) to prevent the rejection problem. A sample was recruited of 1,314 patients who were receiving a stent in a single, previously untreated coronary artery blockage. A total of 652 were randomly assigned to receive a bare-metal stent, and 662 to receive an identical-looking polymer drug-releasing stent. The results were recorded in the following way:

Column 1: Patient identification number
Column 2: Stent type (1 = bare metal, 2 = polymer based)
Column 3: Reference-vessel diameter (the diameter of the artery that is blocked, in millimeters)
Column 4: Lesion length (the length of the blockage, in millimeters)

Reference-vessel diameters and lesion lengths were measured before the stents were inserted.

The following data were recorded 12 months after the stents were inserted.

Column 5: Blockage reoccurrence after 9 months (2 = yes, 1 = no)
Column 6: Blockage that needed to be reopened (2 = yes, 1 = no)
Column 7: Death from cardiac causes (2 = yes, 1 = no)
Column 8: Stroke caused by stent (2 = yes, 1 = no)

a. Using the variables stored in columns 3 through 8, determine whether there is enough evidence to infer that the polymer-based stent is superior to the bare-metal stent.

b. As a laboratory researcher in the pharmaceutical company write a report that describes this experiment and the results.

CASE A17.2 Automobile Crashes and the Ages of Drivers*

DATA
CA17-02

Setting premiums for insurance is a complex task. If the premium is too high, the insurance company will lose customers; if it is too low, the company will lose money. Statistics plays a critical role in almost all aspects of the insurance business. As part of a statistical analysis, an insurance company in Florida studied the relationship between the severity of car crashes and the ages of the drivers. A random sample of crashes in 2002 in the state of Florida was drawn. For each crash, the age category of the driver was recorded as well as whether the driver was injured or killed. The data were stored as follows:

Column 1: Crash number

Column 2: Age category

1. 5 to 34
2. 35 to 44
3. 45 to 54
4. 55 to 64
5. 65 and over

Column 3: Medical status of driver

1 = Uninjured
2 = Injured (but not killed)
3 = Killed

a. Is there enough evidence to conclude that age and medical status of the driver in car crashes are related?

b. Estimate with 95% confidence the proportion of all Florida drivers in crashes in 2002 who were uninjured.

*Adapted from Florida Department of Highway Safety and Vehicles as reported in the *Miami Herald* January 1, 2004, p. 2B.

APPENDIX A

DATA FILE SAMPLE STATISTICS

Chapter 10

10.30 $\bar{x} = 252.38$
10.31 $\bar{x} = 1{,}810.16$
10.32 $\bar{x} = 12.10$
10.33 $\bar{x} = 10.21$
10.34 $\bar{x} = .510$
10.35 $\bar{x} = 26.81$
10.36 $\bar{x} = 19.28$
10.37 $\bar{x} = 15.00$
10.38 $\bar{x} = 585{,}063$
10.39 $\bar{x} = 14.98$
10.40 $\bar{x} = 27.19$

Chapter 11

11.35 $\bar{x} = 5{,}065$
11.36 $\bar{x} = 29{,}120$
11.37 $\bar{x} = 569$
11.38 $\bar{x} = 19.13$
11.39 $\bar{x} = -1.20$
11.40 $\bar{x} = 55.8$
11.41 $\bar{x} = 5.04$
11.42 $\bar{x} = 19.39$
11.43 $\bar{x} = 105.7$
11.44 $\bar{x} = 4.84$
11.45 $\bar{x} = 5.64$
11.46 $\bar{x} = 29.92$
11.47 $\bar{x} = 231.56$

Chapter 12

12.31 $\bar{x} = 7.15, s = 1.65, n = 200$
12.32 $\bar{x} = 4.66, s = 2.37, n = 240$
12.33 $\bar{x} = 17.00, s = 4.31, n = 162$
12.34 $\bar{x} = 15{,}137, s = 5{,}263, n = 306$
12.35 $\bar{x} = 59.04, s = 20.62, n = 122$
12.36 $\bar{x} = 2.67, s = 2.50, n = 188$
12.37 $\bar{x} = 34.49, s = 7.82, n = 900$
12.38 $\bar{x} = 422.36, s = 122.77, n = 176$
12.39 $\bar{x} = 13.94, s = 2.16, n = 212$
12.40 $\bar{x} = 15.27, s = 5.72, n = 116$
12.41 $\bar{x} = 3.79, s = 4.25, n = 564$
12.42 $\bar{x} = 89.27, s = 17.30, n = 85$
12.43 $\bar{x} = 15.02, s = 8.31, n = 83$
12.44 $\bar{x} = 96{,}100, s = 34{,}468, n = 473$
12.45 $\bar{x} = 1.507, s = .640, n = 473$
12.63 $s^2 = 270.58, n = 25$
12.64 $s^2 = 22.56, n = 245$
12.65 $s^2 = 4.72, n = 90$
12.66 $s^2 = 174.47, n = 100$
12.67 $s^2 = 19.68, n = 25$
12.91 $n(1) = 466, n(2) = 55$
12.93 $n(1) = 140, n(2) = 59, n(3) = 39, n(4) = 106, n(5) = 47$
12.94 $n(1) = 153, n(2) = 24$
12.95 $n(1) = 92, n(2) = 28$
12.96 $n(1) = 603, n(2) = 905$
12.97 $n(1) = 92, n(2) = 334$
12.98 $n(1) = 57, n(2) = 35, n(3) = 4, n(4) = 4$
12.100 $n(1) = 245, n(2) = 745, n(3) = 238, n(4) = 1319, n(5) = 2453$
12.101 $n(1) = 786, n(2) = 254$
12.102 $n(1) = 518, n(2) = 132$
12.124 $n(1) = 81, n(2) = 47, n(3) = 167, n(4) = 146, n(5) = 34$
12.125 $n(1) = 63, n(2) = 125, n(3) = 45, n(4) = 87$
12.126 $n(1) = 418, n(2) = 536, n(3) = 882$
12.127 $n(1) = 290, n(2) = 35$
12.128 $n(1) = 72, n(2) = 77, n(3) = 37, n(4) = 50, n(5) = 176$
12.129 $n(1) = 289, n(2) = 51$

Chapter 13

13.17 Tastee: $\bar{x}_1 = 36.93, s_1 = 4.23, n_1 = 15$
Competitor: $\bar{x}_2 = 31.36, s_2 = 3.35, n_2 = 25$
13.18 Oat bran: $\bar{x}_1 = 10.01, s_1 = 4.43, n_1 = 120$
Other: $\bar{x}_2 = 9.12, s_2 = 4.45, n_2 = 120$
13.19 18-to-34: $\bar{x}_1 = 58.99, s_1 = 30.77, n_1 = 250$
35-to-50: $\bar{x}_2 = 52.96, s_2 = 43.32, n_2 = 250$
13.20 2 yrs ago: $\bar{x}_1 = 59.81, s_1 = 7.02, n_1 = 125$
This year: $\bar{x}_2 = 57.40, s_2 = 6.99, n_2 = 159$
13.21 Male: $\bar{x}_1 = 10.23, s_1 = 2.87, n_1 = 100$
Female: $\bar{x}_2 = 9.66, s_2 = 2.90, n_2 = 100$
13.22 A: $\bar{x}_1 = 115.50, s_1 = 21.69, n_1 = 30$
B: $\bar{x}_2 = 110.20, s_2 = 21.93, n_2 = 30$
13.23 Men: $\bar{x}_1 = 5.56, s_1 = 5.36, n_1 = 306$
Women: $\bar{x}_2 = 5.49, s_2 = 5.58, n_2 = 290$
13.24 A: $\bar{x}_1 = 70.42, s_1 = 20.54, n_1 = 24$
B: $\bar{x}_2 = 56.44, s_2 = 9.03, n_2 = 16$
13.25 Successful: $\bar{x}_1 = 5.02, s_1 = 1.39, n_1 = 200$
Unsuccessful: $\bar{x}_2 = 7.80, s_2 = 3.09, n_2 = 200$
13.26 Phone: $\bar{x}_1 = .646, s_1 = .045, n_1 = 125$
Not: $\bar{x}_2 = .601, s_2 = .053, n_2 = 145$
13.27 Chitchat: $\bar{x}_1 = .654, s_1 = .048, n_1 = 95$
Political: $\bar{x}_2 = .662, s_2 = .045, n_2 = 90$
13.28 Planner: $\bar{x}_1 = 6.18, s_1 = 1.59, n_1 = 64$
Broker: $\bar{x}_2 = 5.94, s_2 = 1.61, n_2 = 81$
13.29 Textbook: $\bar{x}_1 = 63.71, s_1 = 5.90, n_1 = 173$
No book: $\bar{x}_2 = 66.80, s_2 = 6.85, n_2 = 202$
13.30 Wendy's: $\bar{x}_1 = 149.85, s_1 = 21.82, n_1 = 213$
McDonald's: $\bar{x}_2 = 154.43, s_2 = 23.64, n_2 = 202$
13.31 Men: $\bar{x}_1 = 488.4, s_1 = 19.6, n_1 = 124$
Women: $\bar{x}_2 = 498.1, s_2 = 21.9, n_2 = 187$
13.32 Applied: $\bar{x}_1 = 130.93, s_1 = 31.99, n_1 = 100$
Contacted: $\bar{x}_2 = 126.14, s_2 = 26.00, n_2 = 100$
13.33 New: $\bar{x}_1 = 73.60, s_1 = 15.60, n_1 = 20$
Existing: $\bar{x}_2 = 69.20, s_2 = 15.06, n_2 = 20$
13.34 Fixed: $\bar{x}_1 = 60{,}245, s_1 = 10{,}506, n_1 = 90$
Commission: $\bar{x}_2 = 63{,}563, s_2 = 10{,}755, n_2 = 90$
13.35 Accident: $\bar{x}_1 = 633.97, s_1 = 49.45, n_1 = 93$
No accident: $\bar{x}_2 = 661.86, s_2 = 52.69, n_2 = 338$
13.36 Cork: $\bar{x}_1 = 14.20, s_1 = 2.84, n_1 = 130$
Metal: $\bar{x}_2 = 11.27, s_2 = 4.42, n_2 = 130$
13.37 Before: $\bar{x}_1 = 496.9, s_1 = 73.8, n_1 = 355$
After: $\bar{x}_2 = 511.3, s_2 = 69.1, n_2 = 288$
13.57 $D = X$[This year] $- X$[5 years ago]: $\bar{x}_D = 12.4, s_D = 99.1, n_D = 150$
13.58 $D = X$[Waiter] $- X$[Waitress]: $\bar{x}_D = -1.16, s_D = 2.22, n_D = 50$
13.59 $D = X$[This year] $- X$[Last year]: $\bar{x}_D = 19.75, n_D = 30.63, n_D = 40$
13.60 $D = X$[Uninsulated] $- X$[Insulated]: $\bar{x}_D = 57.40, s_D = 13.14, n_D = 15$
13.61 $D = X$[Men] $- X$[Women]: $\bar{x}_D = -42.94, s_D = 317.16, n_D = 45$
13.62 $D = X$[Last year] $- X$[Previous year]: $\bar{x}_D = -183.35, s_D = 1{,}568.94, n_D = 170$
13.63 $D = X$[This year] $- X$[Last year]: $\bar{x}_D = .0422, s_D = .1634, n_D = 38$
13.64 $D = X$[Company 1] $- X$[Company 2]: $\bar{x}_D = 520.85, s_D = 1{,}854.92, n_D = 55$

13.65 $D = X[\text{New}] - X[\text{Existing}]$: $\bar{x}_D = 4.55$, $s_D = 7.22$, $n_D = 20$

13.67 $D = X[\text{Finance}] - X[\text{Marketing}]$: $\bar{x}_D = 4{,}587$, $s_D = 22{,}851$, $n_D = 25$

13.69 **a.** $D = X[\text{After}] - X[\text{Before}]$: $\bar{x}_D = -.10$, $s_D = 1.95$, $n_D = 42$

b. $D = X[\text{After}] - X[\text{Before}]$: $\bar{x}_D = 1.24$, $s_D = 2.83$, $n_D = 98$

13.81 Week 1: $s_1^2 = 19.38$, $n_1 = 100$
Week 2: $s_2^2 = 12.70$, $n_2 = 100$

13.82 A: $s_1^2 = 41{,}309$, $n_1 = 100$
B: $s_2^2 = 19{,}850$, $n_2 = 100$

13.83 Portfolio 1: $s_1^2 = .0261$, $n_1 = 52$
Portfolio 2: $s_2^2 = .0875$, $n_2 = 52$

13.84 Teller 1: $s_1^2 = 3.35$, $n_1 = 100$
Teller 2: $s_2^2 = 10.95$, $n_2 = 100$

13.101 Lexus: $n(1) = 33$, $n(2) = 317$
Acura: $n(1) = 33$, $n(2) = 261$

13.102 Smokers: $n_1(1) = 28$; $n_1(2) = 10$
Nonsmokers: $n_2(1) = 150$; $n_2(2) = 12$

13.103 This year: $n_1(1) = 306$; $n_1(2) = 171$
10 years ago: $n_2(1) = 304$; $n_2(2) = 158$

13.104 Canada: $n_1(1) = 230$; $n_1(2) = 215$
U.S.: $n_2(1) = 165$; $n_2(2) = 275$

13.105 A: $n_1(1) = 189$; $n_1(2) = 11$
B: $n_2(1) = 178$; $n_2(2) = 22$

13.106 High school: $n_1(1) = 27$; $n_1(2) = 167$
Postsecondary: $n_2(1) = 17$; $n_2(2) = 63$

13.107 2008: $n_1(1) = 63$; $n_1(2) = 41$
2011: $n_2(1) = 81$; $n_2(2) = 44$

13.108 Canada: Nov $n_1(1) = 244$; $n_1(2) = 62$; $n_1(3) = 62$; $n_1(4) = 19$
Canada: Dec: $n_2(1) = 162$; $n_2(2) = 53$; $n_2(3) = 53$; $n_2(4) = 41$
U.S.: Nov: $n_3(1) = 232$; $n_3(2) = 95$; $n_3(3) = 90$; $n_3(4) = 52$
U.S.: Dec: $n_4(1) = 185$; $n_4(2) = 92$; $n_4(3) = 84$; $n_4(4) = 40$
Britain: Nov: $n_5(1) = 160$; $n_5(2) = 85$; $n_5(3) = 72$; $n_5(4) = 24$
Britain: Dec: $n_6(1) = 129$; $n_6(2) = 84$; $n_6(3) = 60$; $n_6(4) = 27$

13.109 Canada: 2008 $n_1(1) = 192$; $n_1(2) = 373$
Canada: 2009: $n_2(1) = 154$; $n_2(2) = 438$
U.S.: 2008: $n_3(1) = 157$; $n_3(2) = 446$
U.S.: 2009: $n_4(1) = 106$; $n_4(2) = 480$
Britain: 2008: $n_5(1) = 117$; $n_5(2) = 332$
Britain: 2009: $n_6(1) = 72$; $n_6(2) = 405$

13.110 Health conscious: $n_1(1) = 199$; $n_1(2) = 32$
Not health conscious: $n_2(1) = 563$; $n_2(2) = 56$

13.111 Segment 1: $n(1) = 68$, $n(2) = 95$
Segment 2: $n(1) = 20$, $n(2) = 34$
Segment 3: $n(1) = 10$, $n(2) = 13$
Segment 4: $n(1) = 29$, $n(2) = 79$

13.112 Source 1: $n_1(1) = 344$, $n_1(2) = 38$
Source 2: $n_2(1) = 275$, $n_2(2) = 41$

Chapter 14

14.9

Sample	$\bar{x}_i$	s_i^2	n_i
1	68.83	52.28	20
2	65.08	37.38	26
3	62.01	63.46	16
4	64.64	56.88	19

14.10

Sample	$\bar{x}_i$	s_i^2	n_i
1	90.17	991.5	30
2	95.77	900.9	30
3	106.8	928.7	30
4	111.2	1,023	30

14.11

Sample	$\bar{x}_i$	s_i^2	n_i
1	196.8	914.1	41
2	207.8	861.1	73
3	223.4	1,195	86
4	232.7	1,080	79

14.12

Sample	$\bar{x}_i$	s_i^2	n_i
1	164.6	1,164	25
2	185.6	1,719	25
3	154.8	1,113	25
4	182.6	1,657	25
5	178.9	841.8	25

14.13

Sample	$\bar{x}_i$	s_i^2	n_i
1	22.21	121.6	39
2	18.46	90.39	114
3	15.49	85.25	81
4	9.31	65.40	67

14.14

Sample	$\bar{x}_i$	s_i^2	n_i
1	551.5	2,742	20
2	576.8	2,641	20
3	559.5	3,129	20

14.15

Sample	$\bar{x}_i$	s_i^2	n_i
1	5.81	6.22	100
2	5.30	4.05	100
3	5.33	3.90	100

14.16

Sample	$\bar{x}_i$	s_i^2	n_i
1	74.10	250.0	30
2	75.67	184.2	30
3	78.50	233.4	30
4	81.30	242.9	30

14.17

	Size		
Sample	$\bar{x}_i$	s_i^2	n_i
1	24.97	48.23	50
2	21.65	54.54	50
3	17.84	33.85	50

	Nicotine		
Sample	$\bar{x}_i$	s_i^2	n_i
1	15.52	3.72	50
2	13.39	3.59	50
3	10.08	3.83	50

14.18 a.

Sample	$\bar{x}_i$	s_i^2	n_i
1	31.30	28.34	63
2	34.42	23.20	81
3	37.38	31.16	40
4	39.93	72.03	111

b.

Sample	$\bar{x}_i$	s_i^2	n_i
1	37.22	39.82	63
2	38.91	40.85	81
3	41.48	61.38	40
4	41.75	46.59	111

c.

Sample	$\bar{x}_i$	s_i^2	n_i
1	11.75	3.93	63
2	12.41	3.39	81
3	11.73	4.26	40
4	11.89	4.30	111

14.19

Sample	$\bar{x}_i$	s_i^2	n_i
1	153.6	654.3	20
2	151.5	924.0	20
3	133.3	626.8	20

14.20

Sample	$\bar{x}_i$	s_i^2	n_i
1	18.54	178.0	61
2	19.34	171.4	83
3	20.29	297.5	91

14.39

Sample	$\bar{x}_i$	s_i^2	n_i
1	61.60	80.49	10
2	57.30	70.46	10
3	61.80	22.18	10
4	51.80	75.29	10

14.41

Sample	$\bar{x}_i$	s_i^2	n_i
1	53.17	194.6	30
2	49.37	152.6	30
3	44.33	129.9	30

14.59 $k = 3$, $b = 12$, SST = 204.2, SSB = 1,150.2, SSE = 495.1

14.60 $k = 3$, $b = 20$, SST = 7,131, SSB = 177,465, SSE = 1,098

14.61 $k = 3$, $b = 20$, SST = 10.26, SSB = 3,020.30, SSE = 226.71

14.62 $k = 4$, $b = 30$, SST = 4,206, SSB = 126,843, SSE = 5,764

14.63 $k = 7$, $b = 200$, SST = 28,674, SSB = 209,835, SSE = 479,125

14.64 $k = 5$, $b = 36$, SST = 1,406.4, SSB = 7,309.7, SSE = 4,593.9

14.65 $k = 4$, $b = 21$, SST = 563.82, SSB = 1,327.33, SSE = 748.70

Chapter 15

15.7 $n(1) = 28$, $n(2) = 17$, $n(3) = 19$, $n(4) = 17$, $n(5) = 19$

15.8 $n(1) = 41$, $n(2) = 107$, $n(3) = 66$, $n(4) = 19$

15.9 $n(1) = 114$, $n(2) = 92$, $n(3) = 84$, $n(4) = 101$, $n(5) = 107$, $n(6) = 102$

15.10 $n(1) = 11$, $n(2) = 32$, $n(3) = 62$, $n(4) = 29$, $n(5) = 16$

15.11 $n(1) = 8$, $n(2) = 4$, $n(3) = 3$, $n(4) = 8$, $n(5) = 2$

15.12 $n(1) = 159$, $n(2) = 28$, $n(3) = 47$, $n(4) = 16$

15.13 $n(1) = 36$, $n(2) = 58$, $n(3) = 74$, $n(4) = 29$

15.14 $n(1) = 408$, $n(2) = 571$, $n(3) = 221$

15.15 $n(1) = 19$, $n(2) = 23$, $n(3) = 14$, $n(4) = 194$

15.16 $n(1) = 63$, $n(2) = 125$, $n(3) = 45$, $n(4) = 87$

15.31

	Newspaper			
Occupation	**G&M**	**Post**	**Star**	**Sun**
Blue collar	27	18	38	37
White collar	29	43	21	15
Professional	33	51	22	20

15.32

	Actual	
Predicted	**Positive**	**Negative**
Positive	65	64
Negative	39	48

15.33

	Last			
Second-last	**1**	**2**	**3**	**4**
1	39	36	51	23
2	36	32	46	20
3	54	46	65	29
4	24	20	28	10

15.34

Education	**Continuing**	**Quitter**
1	34	23
2	251	212
3	159	248
4	16	57

15.35

	Heartburn Condition			
Source	**1**	**2**	**3**	**4**
ABC	60	23	13	25
CBS	65	19	14	28
NBC	73	26	9	24
Newspaper	67	11	10	7
Radio	57	16	9	14
None	47	21	10	10

15.36

	Degree			
University	**B.A.**	**B.Eng.**	**B.B.A.**	**Other**
1	44	11	34	11
2	52	14	27	7
3	31	27	18	24
4	40	12	42	6

15.37

	Financial Ties	
Results	**Yes**	**No**
Favorable	29	1
Neutral	10	7
Critical	9	14

15.38

	Degree			
Approach	**1**	**2**	**3**	**4**
1	51	8	5	11
2	24	14	12	8
3	26	9	19	8

Chapter 16

16.6 Lengths: $\bar{x} = 38.00$, $s_x^2 = 193.90$
Test: $\bar{y} = 13.80$, $s_y^2 = 47.96$;
$n = 60$, $s_{xy} = 51.86$

16.7 Floors: $\bar{x} = 13.68$, $s_x^2 = 59.32$
Price: $\bar{y} = 210.42$, $s_y^2 = 496.41$;
$n = 50$, $s_{xy} = 86.93$

16.8 Age: $\bar{x} = 45.49$, $s_x^2 = 107.51$
Time: $\bar{y} = 11.55$, $s_y^2 = 42.54$;
$n = 229$, $s_{xy} = 9.67$

16.9 Age: $\bar{x} = 37.28$, $s_x^2 = 55.11$
Employment: $\bar{y} = 26.28$, $s_y^2 = 4.00$;
$n = 80$, $s_{xy} = -6.44$

16.10 Cigarettes: $\bar{x} = 37.64$, $s_x^2 = 108.3$
Days: $\bar{y} = 14.43$, $s_y^2 = 19.80$;
$n = 231$, $s_{xy} = 20.55$

16.11 Distance: $\bar{x} = 4.88$, $s_x^2 = 4.27$
Percent: $\bar{y} = 49.22$, $s_y^2 = 243.94$;
$n = 85$, $s_{xy} = 22.83$

16.12 Size: $\bar{x} = 53.93$, $s_x^2 = 688.18$
Price: $\bar{y} = 6{,}465$, $s_y^2 = 11{,}918{,}489$;
$n = 40$, $s_{xy} = 30{,}945$

16.13 Hours: $\bar{x} = 1{,}199$, $s_x^2 = 59{,}153$
Price: $\bar{y} = 27.73$, $s_y^2 = 3.62$;
$n = 60$, $s_{xy} = -81.78$

16.14 Occupants: $\bar{x} = 4.75$, $s_x^2 = 4.84$
Electricity: $\bar{y} = 762.6$, $s_y^2 = 56{,}725$;
$n = 200$, $s_{xy} = 310.0$

16.15 Income: $\bar{x} = 59.42$, $s_x^2 = 115.24$
Food: $\bar{y} = 270.3$, $s_y^2 = 1{,}797.25$;
$n = 150$, $s_{xy} = 225.66$

16.16 Vacancy: $\bar{x} = 11.33$, $s_x^2 = 35.47$
Rent: $\bar{y} = 17.20$, $s_y^2 = 11.24$;
$n = 30$, $s_{xy} = -10.78$

16.17 Height: $\bar{x} = 68.95$, $s_x^2 = 9.97$
Income: $\bar{y} = 59.59$, $s_y^2 = 71.95$;
$n = 250$, $s_{xy} = 6.02$

16.18 Test: $\bar{x} = 79.47$, $s_x^2 = 16.07$
Nondefective: $\bar{y} = 93.89$, $s_y^2 = 1.28$;
$n = 45$, $s_{xy} = .83$

16.99 Ads: $\bar{x} = 4.12$, $s_x^2 = 3.47$
Customers: $\bar{y} = 384.81$, $s_y^2 = 18{,}552$;
$n = 26$, $s_{xy} = 74.02$

16.100 Age: $\bar{x} = 113.35$, $s_x^2 = 378.77$
Repairs: $\bar{y} = 395.21$, $s_y^2 = 4{,}094.79$;
$n = 20$, $s_{xy} = 936.82$

16.101 Fertilizer: $\bar{x} = 300$, $s_x^2 = 20{,}690$
Yield: $\bar{y} = 318.60$, $s_y^2 = 5{,}230$;
$n = 30$, $s_{xy} = 2{,}538$

16.102 Tar: $\bar{x} = 12.22$, $s_x^2 = 32.10$
Nicotine: $\bar{y} = .88$, $s_y^2 = .13$;
$n = 25$, $s_{xy} = 1.96$

16.103 Television: $\bar{x} = 30.43$, $s_x^2 = 99.11$,
Debt: $\bar{y} = 126{,}604$,
$s_y^2 = 2{,}152{,}602{,}614$; $n = 430$,
$s_{xy} = 255{,}877$

16.104 Test: $\bar{x} = 71.92$, $s_x^2 = 90.97$
Nondefective: $\bar{y} = 94.44$,
$s_y^2 = 11.84$; $n = 50$, $s_{xy} = 13.08$

Chapter 17

17.1 $R^2 = .2425$, $R^2(adjusted) = .2019$,
$s_\varepsilon = 40.24$, $F = 5.97$, p-value $= .0013$

	Coefficients	Standard Error	t-Statistic	p-Value
Intercept	51.39	23.52	2.19	.0331
Lot size	.700	.559	1.25	.2156
Trees	.679	.229	2.96	.0045
Distance	−.378	.195	−1.94	.0577

17.2 $R^2 = .7629$, $R^2(adjusted) = .7453$,
$s_\varepsilon = 3.75$, $F = 43.43$, p-value $= 0$

	Coefficients	Standard Error	t-Statistic	p-Value
Intercept	13.01	3.53	3.69	.0010
Assignment	.194	.200	.97	.3417
Midterm	1.11	.122	9.12	0

17.3 $R^2 = .8935$, $R^2(adjusted) = .8711$,
$s_\varepsilon = 40.13$, $F = 39.86$, p-value $= 0$

	Coefficients	Standard Error	t-Statistic	p-Value
Intercept	−111.83	134.34	−.83	.4155
Permits	4.76	.395	12.06	0
Mortgage	16.99	15.16	1.12	.2764
Apartment vacancy	−10.53	6.39	−1.65	.1161
Office vacancy	1.31	2.79	.47	.6446

17.4 $R^2 = .3511$, $R^2(adjusted) = .3352$,
$s_\varepsilon = 6.99$, $F = 22.01$, p-value $= 0$

	Coefficients	Standard Error	t-Statistic	p-Value
Intercept	−1.97	9.55	−.21	.8369
Minor HR	.666	.087	7.64	0
Age	.136	.524	.26	.7961
Years pro	1.18	.671	1.75	.0819

APPENDIX B

TABLES

TABLE 1 Binomial Probabilities

Tabulated values are $P(X \le k) = \sum_{x=0}^{k} p(x_i)$. (Values are rounded to four decimal places.)

$n = 5$

	p														
k	**0.01**	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
0	0.9510	0.7738	0.5905	0.3277	0.2373	0.1681	0.0778	0.0313	0.0102	0.0024	0.0010	0.0003	0.0000	0.0000	0.0000
1	0.9990	0.9774	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0156	0.0067	0.0005	0.0000	0.0000
2	1.0000	0.9988	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.1035	0.0579	0.0086	0.0012	0.0000
3	1.0000	1.0000	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.3672	0.2627	0.0815	0.0226	0.0010
4	1.0000	1.0000	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.7627	0.6723	0.4095	0.2262	0.0490

$n = 6$

	p														
k	**0.01**	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
0	0.9415	0.7351	0.5314	0.2621	0.1780	0.1176	0.0467	0.0156	0.0041	0.0007	0.0002	0.0001	0.0000	0.0000	0.0000
1	0.9985	0.9672	0.8857	0.6554	0.5339	0.4202	0.2333	0.1094	0.0410	0.0109	0.0046	0.0016	0.0001	0.0000	0.0000
2	1.0000	0.9978	0.9842	0.9011	0.8306	0.7443	0.5443	0.3438	0.1792	0.0705	0.0376	0.0170	0.0013	0.0001	0.0000
3	1.0000	0.9999	0.9987	0.9830	0.9624	0.9295	0.8208	0.6563	0.4557	0.2557	0.1694	0.0989	0.0159	0.0022	0.0000
4	1.0000	1.0000	0.9999	0.9984	0.9954	0.9891	0.9590	0.8906	0.7667	0.5798	0.4661	0.3446	0.1143	0.0328	0.0015
5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9959	0.9844	0.9533	0.8824	0.8220	0.7379	0.4686	0.2649	0.0585

$n = 7$

	p														
k	**0.01**	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
0	0.9321	0.6983	0.4783	0.2097	0.1335	0.0824	0.0280	0.0078	0.0016	0.0002	0.0001	0.0000	0.0000	0.0000	0.0000
1	0.9980	0.9556	0.8503	0.5767	0.4449	0.3294	0.1586	0.0625	0.0188	0.0038	0.0013	0.0004	0.0000	0.0000	0.0000
2	1.0000	0.9962	0.9743	0.8520	0.7564	0.6471	0.4199	0.2266	0.0963	0.0288	0.0129	0.0047	0.0002	0.0000	0.0000
3	1.0000	0.9998	0.9973	0.9667	0.9294	0.8740	0.7102	0.5000	0.2898	0.1260	0.0706	0.0333	0.0027	0.0002	0.0000
4	1.0000	1.0000	0.9998	0.9953	0.9871	0.9712	0.9037	0.7734	0.5801	0.3529	0.2436	0.1480	0.0257	0.0038	0.0000
5	1.0000	1.0000	1.0000	0.9996	0.9987	0.9962	0.9812	0.9375	0.8414	0.6706	0.5551	0.4233	0.1497	0.0444	0.0020
6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9984	0.9922	0.9720	0.9176	0.8665	0.7903	0.5217	0.3017	0.0679

TABLE 1 (*Continued*)

n = 8

	p														
k	**0.01**	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
0	0.9227	0.6634	0.4305	0.1678	0.1001	0.0576	0.0168	0.0039	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9973	0.9428	0.8131	0.5033	0.3671	0.2553	0.1064	0.0352	0.0085	0.0013	0.0004	0.0001	0.0000	0.0000	0.0000
2	0.9999	0.9942	0.9619	0.7969	0.6785	0.5518	0.3154	0.1445	0.0498	0.0113	0.0042	0.0012	0.0000	0.0000	0.0000
3	1.0000	0.9996	0.9950	0.9437	0.8862	0.8059	0.5941	0.3633	0.1737	0.0580	0.0273	0.0104	0.0004	0.0000	0.0000
4	1.0000	1.0000	0.9996	0.9896	0.9727	0.9420	0.8263	0.6367	0.4059	0.1941	0.1138	0.0563	0.0050	0.0004	0.0000
5	1.0000	1.0000	1.0000	0.9988	0.9958	0.9887	0.9502	0.8555	0.6846	0.4482	0.3215	0.2031	0.0381	0.0058	0.0001
6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9915	0.9648	0.8936	0.7447	0.6329	0.4967	0.1869	0.0572	0.0027
7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9961	0.9832	0.9424	0.8999	0.8322	0.5695	0.3366	0.0773

n = 9

	p														
k	**0.01**	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
0	0.9135	0.6302	0.3874	0.1342	0.0751	0.0404	0.0101	0.0020	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9966	0.9288	0.7748	0.4362	0.3003	0.1960	0.0705	0.0195	0.0038	0.0004	0.0001	0.0000	0.0000	0.0000	0.0000
2	0.9999	0.9916	0.9470	0.7382	0.6007	0.4628	0.2318	0.0898	0.0250	0.0043	0.0013	0.0003	0.0000	0.0000	0.0000
3	1.0000	0.9994	0.9917	0.9144	0.8343	0.7297	0.4826	0.2539	0.0994	0.0253	0.0100	0.0031	0.0001	0.0000	0.0000
4	1.0000	1.0000	0.9991	0.9804	0.9511	0.9012	0.7334	0.5000	0.2666	0.0988	0.0489	0.0196	0.0009	0.0000	0.0000
5	1.0000	1.0000	0.9999	0.9969	0.9900	0.9747	0.9006	0.7461	0.5174	0.2703	0.1657	0.0856	0.0083	0.0006	0.0000
6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9750	0.9102	0.7682	0.5372	0.3993	0.2618	0.0530	0.0084	0.0001
7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9962	0.9805	0.9295	0.8040	0.6997	0.5638	0.2252	0.0712	0.0034
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9980	0.9899	0.9596	0.9249	0.8658	0.6126	0.3698	0.0865

TABLE 1 (*Continued*)

n = 10

k	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
3	1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000
4	1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000
5	1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000
6	1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000
7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001
8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956

n = 15

k	0.01	0.05	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90	0.95	0.99
0	0.8601	0.4633	0.2059	0.0352	0.0134	0.0047	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9904	0.8290	0.5490	0.1671	0.0802	0.0353	0.0052	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.9996	0.9638	0.8159	0.3980	0.2361	0.1268	0.0271	0.0037	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	1.0000	0.9945	0.9444	0.6482	0.4613	0.2969	0.0905	0.0176	0.0019	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
4	1.0000	0.9994	0.9873	0.8358	0.6865	0.5155	0.2173	0.0592	0.0093	0.0007	0.0001	0.0000	0.0000	0.0000	0.0000
5	1.0000	0.9999	0.9978	0.9389	0.8516	0.7216	0.4032	0.1509	0.0338	0.0037	0.0008	0.0001	0.0000	0.0000	0.0000
6	1.0000	1.0000	0.9997	0.9819	0.9434	0.8689	0.6098	0.3036	0.0950	0.0152	0.0042	0.0008	0.0000	0.0000	0.0000
7	1.0000	1.0000	1.0000	0.9958	0.9827	0.9500	0.7869	0.5000	0.2131	0.0500	0.0173	0.0042	0.0000	0.0000	0.0000
8	1.0000	1.0000	1.0000	0.9992	0.9958	0.9848	0.9050	0.6964	0.3902	0.1311	0.0566	0.0181	0.0003	0.0000	0.0000
9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9662	0.8491	0.5968	0.2784	0.1484	0.0611	0.0022	0.0001	0.0000
10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9907	0.9408	0.7827	0.4845	0.3135	0.1642	0.0127	0.0006	0.0000
11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824	0.9095	0.7031	0.5387	0.3518	0.0556	0.0055	0.0000
12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963	0.9729	0.8732	0.7639	0.6020	0.1841	0.0362	0.0004
13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9948	0.9647	0.9198	0.8329	0.4510	0.1710	0.0096
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9953	0.9866	0.9648	0.7941	0.5367	0.1399

TABLE 1 (*Continued*)

n = 20

	p														
k	**0.01**	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
0	0.8179	0.3585	0.1216	0.0115	0.0032	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9831	0.7358	0.3917	0.0692	0.0243	0.0076	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.9990	0.9245	0.6769	0.2061	0.0913	0.0355	0.0036	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	1.0000	0.9841	0.8670	0.4114	0.2252	0.1071	0.0160	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	1.0000	0.9974	0.9568	0.6296	0.4148	0.2375	0.0510	0.0059	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	1.0000	0.9997	0.9887	0.8042	0.6172	0.4164	0.1256	0.0207	0.0016	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	1.0000	1.0000	0.9976	0.9133	0.7858	0.6080	0.2500	0.0577	0.0065	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
7	1.0000	1.0000	0.9996	0.9679	0.8982	0.7723	0.4159	0.1316	0.0210	0.0013	0.0002	0.0000	0.0000	0.0000	0.0000
8	1.0000	1.0000	0.9999	0.9900	0.9591	0.8867	0.5956	0.2517	0.0565	0.0051	0.0009	0.0001	0.0000	0.0000	0.0000
9	1.0000	1.0000	1.0000	0.9974	0.9861	0.9520	0.7553	0.4119	0.1275	0.0171	0.0039	0.0006	0.0000	0.0000	0.0000
10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.8725	0.5881	0.2447	0.0480	0.0139	0.0026	0.0000	0.0000	0.0000
11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9435	0.7483	0.4044	0.1133	0.0409	0.0100	0.0001	0.0000	0.0000
12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9790	0.8684	0.5841	0.2277	0.1018	0.0321	0.0004	0.0000	0.0000
13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423	0.7500	0.3920	0.2142	0.0867	0.0024	0.0000	0.0000
14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793	0.8744	0.5836	0.3828	0.1958	0.0113	0.0003	0.0000
15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941	0.9490	0.7625	0.5852	0.3704	0.0432	0.0026	0.0000
16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9840	0.8929	0.7748	0.5886	0.1330	0.0159	0.0000
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9964	0.9645	0.9087	0.7939	0.3231	0.0755	0.0010
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9924	0.9757	0.9308	0.6083	0.2642	0.0169
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9992	0.9968	0.9885	0.8784	0.6415	0.1821

TABLE 1 (*Continued*)

$n = 25$

	p														
k	**0.01**	**0.05**	**0.10**	**0.20**	**0.25**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.75**	**0.80**	**0.90**	**0.95**	**0.99**
0	0.7778	0.2774	0.0718	0.0038	0.0008	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9742	0.6424	0.2712	0.0274	0.0070	0.0016	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.9980	0.8729	0.5371	0.0982	0.0321	0.0090	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
3	0.9999	0.9659	0.7636	0.2340	0.0962	0.0332	0.0024	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
4	1.0000	0.9928	0.9020	0.4207	0.2137	0.0905	0.0095	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
5	1.0000	0.9988	0.9666	0.6167	0.3783	0.1935	0.0294	0.0020	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
6	1.0000	0.9998	0.9905	0.7800	0.5611	0.3407	0.0736	0.0073	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
7	1.0000	1.0000	0.9977	0.8909	0.7265	0.5118	0.1536	0.0216	0.0012	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
8	1.0000	1.0000	0.9995	0.9532	0.8506	0.6769	0.2735	0.0539	0.0043	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
9	1.0000	1.0000	0.9999	0.9827	0.9287	0.8106	0.4246	0.1148	0.0132	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000
10	1.0000	1.0000	1.0000	0.9944	0.9703	0.9022	0.5858	0.2122	0.0344	0.0018	0.0002	0.0000	0.0000	0.0000	0.0000
11	1.0000	1.0000	1.0000	0.9985	0.9893	0.9558	0.7323	0.3450	0.0778	0.0060	0.0009	0.0001	0.0000	0.0000	0.0000
12	1.0000	1.0000	1.0000	0.9996	0.9966	0.9825	0.8462	0.5000	0.1538	0.0175	0.0034	0.0004	0.0000	0.0000	0.0000
13	1.0000	1.0000	1.0000	0.9999	0.9991	0.9940	0.9222	0.6550	0.2677	0.0442	0.0107	0.0015	0.0000	0.0000	0.0000
14	1.0000	1.0000	1.0000	1.0000	0.9998	0.9982	0.9656	0.7878	0.4142	0.0978	0.0297	0.0056	0.0000	0.0000	0.0000
15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9868	0.8852	0.5754	0.1894	0.0713	0.0173	0.0001	0.0000	0.0000
16	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9957	0.9461	0.7265	0.3231	0.1494	0.0468	0.0005	0.0000	0.0000
17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9988	0.9784	0.8464	0.4882	0.2735	0.1091	0.0023	0.0000	0.0000
18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9927	0.9264	0.6593	0.4389	0.2200	0.0095	0.0002	0.0000
19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9980	0.9706	0.8065	0.6217	0.3833	0.0334	0.0012	0.0000
20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9905	0.9095	0.7863	0.5793	0.0980	0.0072	0.0000
21	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9976	0.9668	0.9038	0.7660	0.2364	0.0341	0.0001
22	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996	0.9910	0.9679	0.9018	0.4629	0.1271	0.0020
23	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9984	0.9930	0.9726	0.7288	0.3576	0.0258
24	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9992	0.9962	0.9282	0.7226	0.2222

TABLE 2 Poisson Probabilities

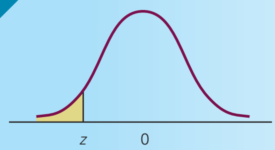
Tabulated values are $P(X \leq k) = \sum_{x=0}^{k} p(xi)$. (Values are rounded to four decimal places.)

	μ															
k	**0.10**	**0.20**	**0.30**	**0.40**	**0.50**	**1.0**	**1.5**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**	**4.5**	**5.0**	**5.5**	**6.0**
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067	0.0041	0.0025
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404	0.0266	0.0174
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1736	0.1247	0.0884	0.0620
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650	0.2017	0.1512
4		1.0000	1.0000	0.9999	0.9998	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405	0.3575	0.2851
5				1.0000	1.0000	0.9994	0.9955	0.9834	0.9580	0.9161	0.8576	0.7851	0.7029	0.6160	0.5289	0.4457
6						0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622	0.6860	0.6063
7						1.0000	0.9998	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666	0.8095	0.7440
8							1.0000	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319	0.8944	0.8472
9								1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682	0.9462	0.9161
10									0.9999	0.9997	0.9990	0.9972	0.9933	0.9863	0.9747	0.9574
11									1.0000	0.9999	0.9997	0.9991	0.9976	0.9945	0.9890	0.9799
12										1.0000	0.9999	0.9997	0.9992	0.9980	0.9955	0.9912
13											1.0000	0.9999	0.9997	0.9993	0.9983	0.9964
14												1.0000	0.9999	0.9998	0.9994	0.9986
15													1.0000	0.9999	0.9998	0.9995
16														1.0000	0.9999	0.9998
17															1.0000	0.9999
18																1.0000
19																
20																

TABLE 2 (*Continued*)

	μ												
k	6.50	7.00	7.50	8.00	8.50	9.00	9.50	10	11	12	13	14	15
0	0.0015	0.0009	0.0006	0.0003	0.0002	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.0113	0.0073	0.0047	0.0030	0.0019	0.0012	0.0008	0.0005	0.0002	0.0001	0.0000	0.0000	0.0000
2	0.0430	0.0296	0.0203	0.0138	0.0093	0.0062	0.0042	0.0028	0.0012	0.0005	0.0002	0.0001	0.0000
3	0.1118	0.0818	0.0591	0.0424	0.0301	0.0212	0.0149	0.0103	0.0049	0.0023	0.0011	0.0005	0.0002
4	0.2237	0.1730	0.1321	0.0996	0.0744	0.0550	0.0403	0.0293	0.0151	0.0076	0.0037	0.0018	0.0009
5	0.3690	0.3007	0.2414	0.1912	0.1496	0.1157	0.0885	0.0671	0.0375	0.0203	0.0107	0.0055	0.0028
6	0.5265	0.4497	0.3782	0.3134	0.2562	0.2068	0.1649	0.1301	0.0786	0.0458	0.0259	0.0142	0.0076
7	0.6728	0.5987	0.5246	0.4530	0.3856	0.3239	0.2687	0.2202	0.1432	0.0895	0.0540	0.0316	0.0180
8	0.7916	0.7291	0.6620	0.5925	0.5231	0.4557	0.3918	0.3328	0.2320	0.1550	0.0998	0.0621	0.0374
9	0.8774	0.8305	0.7764	0.7166	0.6530	0.5874	0.5218	0.4579	0.3405	0.2424	0.1658	0.1094	0.0699
10	0.9332	0.9015	0.8622	0.8159	0.7634	0.7060	0.6453	0.5830	0.4599	0.3472	0.2517	0.1757	0.1185
11	0.9661	0.9467	0.9208	0.8881	0.8487	0.8030	0.7520	0.6968	0.5793	0.4616	0.3532	0.2600	0.1848
12	0.9840	0.9730	0.9573	0.9362	0.9091	0.8758	0.8364	0.7916	0.6887	0.5760	0.4631	0.3585	0.2676
13	0.9929	0.9872	0.9784	0.9658	0.9486	0.9261	0.8981	0.8645	0.7813	0.6815	0.5730	0.4644	0.3632
14	0.9970	0.9943	0.9897	0.9827	0.9726	0.9585	0.9400	0.9165	0.8540	0.7720	0.6751	0.5704	0.4657
15	0.9988	0.9976	0.9954	0.9918	0.9862	0.9780	0.9665	0.9513	0.9074	0.8444	0.7636	0.6694	0.5681
16	0.9996	0.9990	0.9980	0.9963	0.9934	0.9889	0.9823	0.9730	0.9441	0.8987	0.8355	0.7559	0.6641
17	0.9998	0.9996	0.9992	0.9984	0.9970	0.9947	0.9911	0.9857	0.9678	0.9370	0.8905	0.8272	0.7489
18	0.9999	0.9999	0.9997	0.9993	0.9987	0.9976	0.9957	0.9928	0.9823	0.9626	0.9302	0.8826	0.8195
19	1.0000	1.0000	0.9999	0.9997	0.9995	0.9989	0.9980	0.9965	0.9907	0.9787	0.9573	0.9235	0.8752
20			1.0000	0.9999	0.9998	0.9996	0.9991	0.9984	0.9953	0.9884	0.9750	0.9521	0.9170
21				1.0000	0.9999	0.9998	0.9996	0.9993	0.9977	0.9939	0.9859	0.9712	0.9469
22					1.0000	0.9999	0.9999	0.9997	0.9990	0.9970	0.9924	0.9833	0.9673
23						1.0000	0.9999	0.9999	0.9995	0.9985	0.9960	0.9907	0.9805
24							1.0000	1.0000	0.9998	0.9993	0.9980	0.9950	0.9888
25									0.9999	0.9997	0.9990	0.9974	0.9938
26									1.0000	0.9999	0.9995	0.9987	0.9967
27										0.9999	0.9998	0.9994	0.9983
28										1.0000	0.9999	0.9997	0.9991
29											1.0000	0.9999	0.9996
30												0.9999	0.9998
31												1.0000	0.9999
32													1.0000

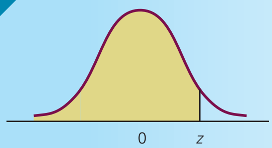
TABLE 3 Cumulative Standardized Normal Probabilities



$P(-\infty < Z < z)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

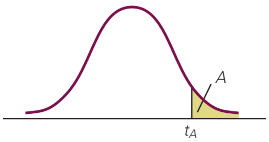
TABLE 3 (*Continued*)



$P(-\infty < Z < z)$.

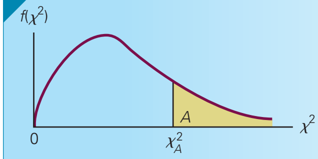
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

TABLE 4
Critical Values of the Student t Distribution



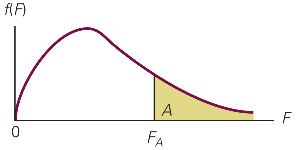
Degrees of Freedom	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
35	1.306	1.690	2.030	2.438	2.724
40	1.303	1.684	2.021	2.423	2.704
45	1.301	1.679	2.014	2.412	2.690
50	1.299	1.676	2.009	2.403	2.678
55	1.297	1.673	2.004	2.396	2.668
60	1.296	1.671	2.000	2.390	2.660
65	1.295	1.669	1.997	2.385	2.654
70	1.294	1.667	1.994	2.381	2.648
75	1.293	1.665	1.992	2.377	2.643
80	1.292	1.664	1.990	2.374	2.639
85	1.292	1.663	1.988	2.371	2.635
90	1.291	1.662	1.987	2.368	2.632
95	1.291	1.661	1.985	2.366	2.629
100	1.290	1.660	1.984	2.364	2.626
110	1.289	1.659	1.982	2.361	2.621
120	1.289	1.658	1.980	2.358	2.617
130	1.288	1.657	1.978	2.355	2.614
140	1.288	1.656	1.977	2.353	2.611
150	1.287	1.655	1.976	2.351	2.609
160	1.287	1.654	1.975	2.350	2.607
170	1.287	1.654	1.974	2.348	2.605
180	1.286	1.653	1.973	2.347	2.603
190	1.286	1.653	1.973	2.346	2.602
200	1.286	1.653	1.972	2.345	2.601
∞	1.282	1.645	1.960	2.326	2.576

TABLE 5 Critical Values of the χ^2 Distribution



Degrees of Freedom	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000039	0.000157	0.000982	0.00393	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	4.61	5.99	7.38	9.21	10.6
3	0.072	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	34.4	37.7	40.6	44.3	46.9
26	11.2	12.2	13.8	15.4	17.3	35.6	38.9	41.9	45.6	48.3
27	11.8	12.9	14.6	16.2	18.1	36.7	40.1	43.2	47.0	49.6
28	12.5	13.6	15.3	16.9	18.9	37.9	41.3	44.5	48.3	51.0
29	13.1	14.3	16.0	17.7	19.8	39.1	42.6	45.7	49.6	52.3
30	13.8	15.0	16.8	18.5	20.6	40.3	43.8	47.0	50.9	53.7
40	20.7	22.2	24.4	26.5	29.1	51.8	55.8	59.3	63.7	66.8
50	28.0	29.7	32.4	34.8	37.7	63.2	67.5	71.4	76.2	79.5
60	35.5	37.5	40.5	43.2	46.5	74.4	79.1	83.3	88.4	92.0
70	43.3	45.4	48.8	51.7	55.3	85.5	90.5	95.0	100	104
80	51.2	53.5	57.2	60.4	64.3	96.6	102	107	112	116
90	59.2	61.8	65.6	69.1	73.3	108	113	118	124	128
100	67.3	70.1	74.2	77.9	82.4	118	124	130	136	140

TABLE 6(a) Critical Values of the F-Distribution: $A = .05$

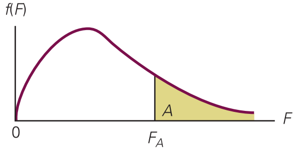


ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	NUMERATOR DEGREES OF FREEDOM																			
DENOMINATOR DEGREES OF FREEDOM																				
1	161	199	216	225	230	234	237	239	241	242	243	244	245	245	246	246	247	247	248	248
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71	8.70	8.69	8.68	8.67	8.67	8.66
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87	5.86	5.84	5.83	5.82	5.81	5.80
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64	4.62	4.60	4.59	4.58	4.57	4.56
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96	3.94	3.92	3.91	3.90	3.88	3.87
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53	3.51	3.49	3.48	3.47	3.46	3.44
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24	3.22	3.20	3.19	3.17	3.16	3.15
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03	3.01	2.99	2.97	2.96	2.95	2.94
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86	2.85	2.83	2.81	2.80	2.79	2.77
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.69	2.67	2.66	2.65
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64	2.62	2.60	2.58	2.57	2.56	2.54
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55	2.53	2.51	2.50	2.48	2.47	2.46
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48	2.46	2.44	2.43	2.41	2.40	2.39
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42	2.40	2.38	2.37	2.35	2.34	2.33
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37	2.35	2.33	2.32	2.30	2.29	2.28
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33	2.31	2.29	2.27	2.26	2.24	2.23
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29	2.27	2.25	2.23	2.22	2.20	2.19
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26	2.23	2.21	2.20	2.18	2.17	2.16
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.22	2.20	2.18	2.17	2.15	2.14	2.12
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17	2.15	2.13	2.11	2.10	2.08	2.07
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13	2.11	2.09	2.07	2.05	2.04	2.03
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.12	2.09	2.07	2.05	2.03	2.02	2.00	1.99
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	2.09	2.06	2.04	2.02	2.00	1.99	1.97	1.96
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04	2.01	1.99	1.98	1.96	1.95	1.93
35	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.07	2.04	2.01	1.99	1.96	1.94	1.92	1.91	1.89	1.88
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95	1.92	1.90	1.89	1.87	1.85	1.84
45	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	2.01	1.97	1.94	1.92	1.89	1.87	1.86	1.84	1.82	1.81
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.92	1.89	1.87	1.85	1.83	1.81	1.80	1.78
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86	1.84	1.82	1.80	1.78	1.76	1.75
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.86	1.84	1.81	1.79	1.77	1.75	1.74	1.72
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.84	1.82	1.79	1.77	1.75	1.73	1.72	1.70
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.90	1.86	1.83	1.80	1.78	1.76	1.74	1.72	1.70	1.69
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.82	1.79	1.77	1.75	1.73	1.71	1.69	1.68
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.87	1.83	1.80	1.78	1.75	1.73	1.71	1.69	1.67	1.66
140	3.91	3.06	2.67	2.44	2.28	2.16	2.08	2.01	1.95	1.90	1.86	1.82	1.79	1.76	1.74	1.72	1.70	1.68	1.66	1.65
160	3.90	3.05	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.81	1.78	1.75	1.73	1.71	1.69	1.67	1.65	1.64
180	3.89	3.05	2.65	2.42	2.26	2.15	2.06	1.99	1.93	1.88	1.84	1.81	1.77	1.75	1.72	1.70	1.68	1.66	1.64	1.63
200	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.77	1.74	1.72	1.69	1.67	1.66	1.64	1.62
∞	3.84	3.00	2.61	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.72	1.69	1.67	1.64	1.62	1.60	1.59	1.57

ν_2 \ ν_1	22	24	26	28	30	35	40	45	50	60	70	80	90	100	120	140	160	180	200	∞
	NUMERATOR DEGREES OF FREEDOM																			
1	249	249	249	250	250	251	251	251	252	252	252	253	253	253	253	253	254	254	254	254
2	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5
3	8.65	8.64	8.63	8.62	8.62	8.60	8.59	8.59	8.58	8.57	8.57	8.56	8.56	8.55	8.55	8.55	8.54	8.54	8.54	8.53
4	5.79	5.77	5.76	5.75	5.75	5.73	5.72	5.71	5.70	5.69	5.68	5.67	5.67	5.66	5.66	5.65	5.65	5.65	5.65	5.63
5	4.54	4.53	4.52	4.50	4.50	4.48	4.46	4.45	4.44	4.43	4.42	4.41	4.41	4.41	4.40	4.39	4.39	4.39	4.39	4.37
6	3.86	3.84	3.83	3.82	3.81	3.79	3.77	3.76	3.75	3.74	3.73	3.72	3.72	3.71	3.70	3.70	3.70	3.69	3.69	3.67
7	3.43	3.41	3.40	3.39	3.38	3.36	3.34	3.33	3.32	3.30	3.29	3.29	3.28	3.27	3.27	3.26	3.26	3.25	3.25	3.23
8	3.13	3.12	3.10	3.09	3.08	3.06	3.04	3.03	3.02	3.01	2.99	2.99	2.98	2.97	2.97	2.96	2.96	2.95	2.95	2.93
9	2.92	2.90	2.89	2.87	2.86	2.84	2.83	2.81	2.80	2.79	2.78	2.77	2.76	2.76	2.75	2.74	2.74	2.73	2.73	2.71
10	2.75	2.74	2.72	2.71	2.70	2.68	2.66	2.65	2.64	2.62	2.61	2.60	2.59	2.59	2.58	2.57	2.57	2.57	2.56	2.54
11	2.63	2.61	2.59	2.58	2.57	2.55	2.53	2.52	2.51	2.49	2.48	2.47	2.46	2.46	2.45	2.44	2.44	2.43	2.43	2.41
12	2.52	2.51	2.49	2.48	2.47	2.44	2.43	2.41	2.40	2.38	2.37	2.36	2.36	2.35	2.34	2.33	2.33	2.33	2.32	2.30
13	2.44	2.42	2.41	2.39	2.38	2.36	2.34	2.33	2.31	2.30	2.28	2.27	2.27	2.26	2.25	2.25	2.24	2.24	2.23	2.21
14	2.37	2.35	2.33	2.32	2.31	2.28	2.27	2.25	2.24	2.22	2.21	2.20	2.19	2.19	2.18	2.17	2.17	2.16	2.16	2.13
15	2.31	2.29	2.27	2.26	2.25	2.22	2.20	2.19	2.18	2.16	2.15	2.14	2.13	2.12	2.11	2.11	2.10	2.10	2.10	2.07
16	2.25	2.24	2.22	2.21	2.19	2.17	2.15	2.14	2.12	2.11	2.09	2.08	2.07	2.07	2.06	2.05	2.05	2.04	2.04	2.01
17	2.21	2.19	2.17	2.16	2.15	2.12	2.10	2.09	2.08	2.06	2.05	2.03	2.03	2.02	2.01	2.00	2.00	1.99	1.99	1.96
18	2.17	2.15	2.13	2.12	2.11	2.08	2.06	2.05	2.04	2.02	2.00	1.99	1.98	1.98	1.97	1.96	1.96	1.95	1.95	1.92
19	2.13	2.11	2.10	2.08	2.07	2.05	2.03	2.01	2.00	1.98	1.97	1.96	1.95	1.94	1.93	1.92	1.92	1.91	1.91	1.88
20	2.10	2.08	2.07	2.05	2.04	2.01	1.99	1.98	1.97	1.95	1.93	1.92	1.91	1.91	1.90	1.89	1.88	1.88	1.88	1.84
22	2.05	2.03	2.01	2.00	1.98	1.96	1.94	1.92	1.91	1.89	1.88	1.86	1.86	1.85	1.84	1.83	1.82	1.82	1.82	1.78
24	2.00	1.98	1.97	1.95	1.94	1.91	1.89	1.88	1.86	1.84	1.83	1.82	1.81	1.80	1.79	1.78	1.78	1.77	1.77	1.73
26	1.97	1.95	1.93	1.91	1.90	1.87	1.85	1.84	1.82	1.80	1.79	1.78	1.77	1.76	1.75	1.74	1.73	1.73	1.73	1.69
28	1.93	1.91	1.90	1.88	1.87	1.84	1.82	1.80	1.79	1.77	1.75	1.74	1.73	1.73	1.71	1.71	1.70	1.69	1.69	1.65
30	1.91	1.89	1.87	1.85	1.84	1.81	1.79	1.77	1.76	1.74	1.72	1.71	1.70	1.70	1.68	1.68	1.67	1.66	1.66	1.62
35	1.85	1.83	1.82	1.80	1.79	1.76	1.74	1.72	1.70	1.68	1.66	1.65	1.64	1.63	1.62	1.61	1.61	1.60	1.60	1.56
40	1.81	1.79	1.77	1.76	1.74	1.72	1.69	1.67	1.66	1.64	1.62	1.61	1.60	1.59	1.58	1.57	1.56	1.55	1.55	1.51
45	1.78	1.76	1.74	1.73	1.71	1.68	1.66	1.64	1.63	1.60	1.59	1.57	1.56	1.55	1.54	1.53	1.52	1.52	1.51	1.47
50	1.76	1.74	1.72	1.70	1.69	1.66	1.63	1.61	1.60	1.58	1.56	1.54	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.44
60	1.72	1.70	1.68	1.66	1.65	1.62	1.59	1.57	1.56	1.53	1.52	1.50	1.49	1.48	1.47	1.46	1.45	1.44	1.44	1.39
70	1.70	1.67	1.65	1.64	1.62	1.59	1.57	1.55	1.53	1.50	1.49	1.47	1.46	1.45	1.44	1.42	1.42	1.41	1.40	1.35
80	1.68	1.65	1.63	1.62	1.60	1.57	1.54	1.52	1.51	1.48	1.46	1.45	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.33
90	1.66	1.64	1.62	1.60	1.59	1.55	1.53	1.51	1.49	1.46	1.44	1.43	1.42	1.41	1.39	1.38	1.37	1.36	1.36	1.30
100	1.65	1.63	1.61	1.59	1.57	1.54	1.52	1.49	1.48	1.45	1.43	1.41	1.40	1.39	1.38	1.36	1.35	1.35	1.34	1.28
120	1.63	1.61	1.59	1.57	1.55	1.52	1.50	1.47	1.46	1.43	1.41	1.39	1.38	1.37	1.35	1.34	1.33	1.32	1.32	1.26
140	1.62	1.60	1.57	1.56	1.54	1.51	1.48	1.46	1.44	1.41	1.39	1.38	1.36	1.35	1.33	1.32	1.31	1.30	1.30	1.23
160	1.61	1.59	1.57	1.55	1.53	1.50	1.47	1.45	1.43	1.40	1.38	1.36	1.35	1.34	1.32	1.31	1.30	1.29	1.28	1.22
180	1.60	1.58	1.56	1.54	1.52	1.49	1.46	1.44	1.42	1.39	1.37	1.35	1.34	1.33	1.31	1.30	1.29	1.28	1.27	1.20
200	1.60	1.57	1.55	1.53	1.52	1.48	1.46	1.43	1.41	1.39	1.36	1.35	1.33	1.32	1.30	1.29	1.28	1.27	1.26	1.19
∞	1.54	1.52	1.50	1.48	1.46	1.42	1.40	1.37	1.35	1.32	1.29	1.28	1.26	1.25	1.22	1.21	1.19	1.18	1.17	1.00

DENOMINATOR DEGREES OF FREEDOM

TABLE 6(b) Values of the *F*-Distribution: $A = .025$

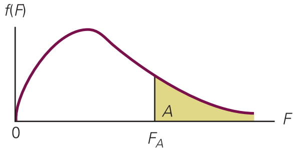


NUMERATOR DEGREES OF FREEDOM (ν_1); DENOMINATOR DEGREES OF FREEDOM (ν_2)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	648	799	864	900	922	937	948	957	963	969	973	977	980	983	985	987	989	990	992	993
2	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.4
3	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.4	14.3	14.3	14.3	14.3	14.2	14.2	14.2	14.2	14.2
4	12.2	10.6	10.0	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.79	8.75	8.71	8.68	8.66	8.63	8.61	8.59	8.58	8.56
5	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	6.52	6.49	6.46	6.43	6.40	6.38	6.36	6.34	6.33
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41	5.37	5.33	5.30	5.27	5.24	5.22	5.20	5.18	5.17
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.71	4.67	4.63	4.60	4.57	4.54	4.52	4.50	4.48	4.47
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.24	4.20	4.16	4.13	4.10	4.08	4.05	4.03	4.02	4.00
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.91	3.87	3.83	3.80	3.77	3.74	3.72	3.70	3.68	3.67
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.66	3.62	3.58	3.55	3.52	3.50	3.47	3.45	3.44	3.42
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.47	3.43	3.39	3.36	3.33	3.30	3.28	3.26	3.24	3.23
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.32	3.28	3.24	3.21	3.18	3.15	3.13	3.11	3.09	3.07
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.20	3.15	3.12	3.08	3.05	3.03	3.00	2.98	2.96	2.95
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.09	3.05	3.01	2.98	2.95	2.92	2.90	2.88	2.86	2.84
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	3.01	2.96	2.92	2.89	2.86	2.84	2.81	2.79	2.77	2.76
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.93	2.89	2.85	2.82	2.79	2.76	2.74	2.72	2.70	2.68
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.87	2.82	2.79	2.75	2.72	2.70	2.67	2.65	2.63	2.62
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.81	2.77	2.73	2.70	2.67	2.64	2.62	2.60	2.58	2.56
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.76	2.72	2.68	2.65	2.62	2.59	2.57	2.55	2.53	2.51
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.72	2.68	2.64	2.60	2.57	2.55	2.52	2.50	2.48	2.46
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.65	2.60	2.56	2.53	2.50	2.47	2.45	2.43	2.41	2.39
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.59	2.54	2.50	2.47	2.44	2.41	2.39	2.36	2.35	2.33
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59	2.54	2.49	2.45	2.42	2.39	2.36	2.34	2.31	2.29	2.28
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55	2.49	2.45	2.41	2.37	2.34	2.32	2.29	2.27	2.25	2.23
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.46	2.41	2.37	2.34	2.31	2.28	2.26	2.23	2.21	2.20
35	5.48	4.11	3.52	3.18	2.96	2.80	2.68	2.58	2.50	2.44	2.39	2.34	2.30	2.27	2.23	2.21	2.18	2.16	2.14	2.12
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.33	2.29	2.25	2.21	2.18	2.15	2.13	2.11	2.09	2.07
45	5.38	4.01	3.42	3.09	2.86	2.70	2.58	2.49	2.41	2.35	2.29	2.25	2.21	2.17	2.14	2.11	2.09	2.07	2.04	2.03
50	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32	2.26	2.22	2.18	2.14	2.11	2.08	2.06	2.03	2.01	1.99
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.22	2.17	2.13	2.09	2.06	2.03	2.01	1.98	1.96	1.94
70	5.25	3.89	3.31	2.97	2.75	2.59	2.47	2.38	2.30	2.24	2.18	2.14	2.10	2.06	2.03	2.00	1.97	1.95	1.93	1.91
80	5.22	3.86	3.28	2.95	2.73	2.57	2.45	2.35	2.28	2.21	2.16	2.11	2.07	2.03	2.00	1.97	1.95	1.92	1.90	1.88
90	5.20	3.84	3.26	2.93	2.71	2.55	2.43	2.34	2.26	2.19	2.14	2.09	2.05	2.02	1.98	1.95	1.93	1.91	1.88	1.86
100	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.24	2.18	2.12	2.08	2.04	2.00	1.97	1.94	1.91	1.89	1.87	1.85
120	5.15	3.80	3.23	2.89	2.67	2.52	2.39	2.30	2.22	2.16	2.10	2.05	2.01	1.98	1.94	1.92	1.89	1.87	1.84	1.82
140	5.13	3.79	3.21	2.88	2.66	2.50	2.38	2.28	2.21	2.14	2.09	2.04	2.00	1.96	1.93	1.90	1.87	1.85	1.83	1.81
160	5.12	3.78	3.20	2.87	2.65	2.49	2.37	2.27	2.19	2.13	2.07	2.03	1.99	1.95	1.92	1.89	1.86	1.84	1.82	1.80
180	5.11	3.77	3.19	2.86	2.64	2.48	2.36	2.26	2.19	2.12	2.07	2.02	1.98	1.94	1.91	1.88	1.85	1.83	1.81	1.79
200	5.10	3.76	3.18	2.85	2.63	2.47	2.35	2.26	2.18	2.11	2.06	2.01	1.97	1.93	1.90	1.87	1.84	1.82	1.80	1.78
∞	5.03	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.99	1.95	1.90	1.87	1.83	1.80	1.78	1.75	1.73	1.71

ν_2 \ ν_1 (Denominator / Numerator degrees of freedom)	22	24	26	28	30	35	40	45	50	60	70	80	90	100	120	140	160	180	200	∞
1	995	997	999	1000	1001	1004	1006	1007	1008	1010	1011	1012	1013	1013	1014	1015	1015	1015	1016	1018
2	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5
3	14.1	14.1	14.1	14.1	14.1	14.1	14.0	14.0	14.0	14.0	14.0	14.0	14.0	14.0	13.9	13.9	13.9	13.9	13.9	13.9
4	8.53	8.51	8.49	8.48	8.46	8.43	8.41	8.39	8.38	8.36	8.35	8.33	8.33	8.32	8.31	8.30	8.30	8.29	8.29	8.26
5	6.30	6.28	6.26	6.24	6.23	6.20	6.18	6.16	6.14	6.12	6.11	6.10	6.09	6.08	6.07	6.06	6.06	6.05	6.05	6.02
6	5.14	5.12	5.10	5.08	5.07	5.04	5.01	4.99	4.98	4.96	4.94	4.93	4.92	4.92	4.90	4.90	4.89	4.89	4.88	4.85
7	4.44	4.41	4.39	4.38	4.36	4.33	4.31	4.29	4.28	4.25	4.24	4.23	4.22	4.21	4.20	4.19	4.18	4.18	4.18	4.14
8	3.97	3.95	3.93	3.91	3.89	3.86	3.84	3.82	3.81	3.78	3.77	3.76	3.75	3.74	3.73	3.72	3.71	3.71	3.70	3.67
9	3.64	3.61	3.59	3.58	3.56	3.53	3.51	3.49	3.47	3.45	3.43	3.42	3.41	3.40	3.39	3.38	3.38	3.37	3.37	3.33
10	3.39	3.37	3.34	3.33	3.31	3.28	3.26	3.24	3.22	3.20	3.18	3.17	3.16	3.15	3.14	3.13	3.13	3.12	3.12	3.08
11	3.20	3.17	3.15	3.13	3.12	3.09	3.06	3.04	3.03	3.00	2.99	2.97	2.96	2.96	2.94	2.94	2.93	2.92	2.92	2.88
12	3.04	3.02	3.00	2.98	2.96	2.93	2.91	2.89	2.87	2.85	2.83	2.82	2.81	2.80	2.79	2.78	2.77	2.77	2.76	2.73
13	2.92	2.89	2.87	2.85	2.84	2.80	2.78	2.76	2.74	2.72	2.70	2.69	2.68	2.67	2.66	2.65	2.64	2.64	2.63	2.60
14	2.81	2.79	2.77	2.75	2.73	2.70	2.67	2.65	2.64	2.61	2.60	2.58	2.57	2.56	2.55	2.54	2.54	2.53	2.53	2.49
15	2.73	2.70	2.68	2.66	2.64	2.61	2.59	2.56	2.55	2.52	2.51	2.49	2.48	2.47	2.46	2.45	2.44	2.44	2.44	2.40
16	2.65	2.63	2.60	2.58	2.57	2.53	2.51	2.49	2.47	2.45	2.43	2.42	2.40	2.40	2.38	2.37	2.37	2.36	2.36	2.32
17	2.59	2.56	2.54	2.52	2.50	2.47	2.44	2.42	2.41	2.38	2.36	2.35	2.34	2.33	2.32	2.31	2.30	2.29	2.29	2.25
18	2.53	2.50	2.48	2.46	2.44	2.41	2.38	2.36	2.35	2.32	2.30	2.29	2.28	2.27	2.26	2.25	2.24	2.23	2.23	2.19
19	2.48	2.45	2.43	2.41	2.39	2.36	2.33	2.31	2.30	2.27	2.25	2.24	2.23	2.22	2.20	2.19	2.19	2.18	2.18	2.13
20	2.43	2.41	2.39	2.37	2.35	2.31	2.29	2.27	2.25	2.22	2.20	2.19	2.18	2.17	2.16	2.15	2.14	2.13	2.13	2.09
22	2.36	2.33	2.31	2.29	2.27	2.24	2.21	2.19	2.17	2.14	2.13	2.11	2.10	2.09	2.08	2.07	2.06	2.05	2.05	2.00
24	2.30	2.27	2.25	2.23	2.21	2.17	2.15	2.12	2.11	2.08	2.06	2.05	2.03	2.02	2.01	2.00	1.99	1.99	1.98	1.94
26	2.24	2.22	2.19	2.17	2.16	2.12	2.09	2.07	2.05	2.03	2.01	1.99	1.98	1.97	1.95	1.94	1.94	1.93	1.92	1.88
28	2.20	2.17	2.15	2.13	2.11	2.08	2.05	2.03	2.01	1.98	1.96	1.94	1.93	1.92	1.91	1.90	1.89	1.88	1.88	1.83
30	2.16	2.14	2.11	2.09	2.07	2.04	2.01	1.99	1.97	1.94	1.92	1.90	1.89	1.88	1.87	1.86	1.85	1.84	1.84	1.79
35	2.09	2.06	2.04	2.02	2.00	1.96	1.93	1.91	1.89	1.86	1.84	1.82	1.81	1.80	1.79	1.77	1.77	1.76	1.75	1.70
40	2.03	2.01	1.98	1.96	1.94	1.90	1.88	1.85	1.83	1.80	1.78	1.76	1.75	1.74	1.72	1.71	1.70	1.70	1.69	1.64
45	1.99	1.96	1.94	1.92	1.90	1.86	1.83	1.81	1.79	1.76	1.74	1.72	1.70	1.69	1.68	1.66	1.66	1.65	1.64	1.59
50	1.96	1.93	1.91	1.89	1.87	1.83	1.80	1.77	1.75	1.72	1.70	1.68	1.67	1.66	1.64	1.63	1.62	1.61	1.60	1.55
60	1.91	1.88	1.86	1.83	1.82	1.78	1.74	1.72	1.70	1.67	1.64	1.63	1.61	1.60	1.58	1.57	1.56	1.55	1.54	1.48
70	1.88	1.85	1.82	1.80	1.78	1.74	1.71	1.68	1.66	1.63	1.60	1.59	1.57	1.56	1.54	1.53	1.52	1.51	1.50	1.44
80	1.85	1.82	1.79	1.77	1.75	1.71	1.68	1.65	1.63	1.60	1.57	1.55	1.54	1.53	1.51	1.49	1.48	1.47	1.47	1.40
90	1.83	1.80	1.77	1.75	1.73	1.69	1.66	1.63	1.61	1.58	1.55	1.53	1.52	1.50	1.48	1.47	1.46	1.45	1.44	1.37
100	1.81	1.78	1.76	1.74	1.71	1.67	1.64	1.61	1.59	1.56	1.53	1.51	1.50	1.48	1.46	1.45	1.44	1.43	1.42	1.35
120	1.79	1.76	1.73	1.71	1.69	1.65	1.61	1.59	1.56	1.53	1.50	1.48	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.31
140	1.77	1.74	1.72	1.69	1.67	1.63	1.60	1.57	1.55	1.51	1.48	1.46	1.45	1.43	1.41	1.39	1.38	1.37	1.36	1.28
160	1.76	1.73	1.70	1.68	1.66	1.62	1.58	1.55	1.53	1.50	1.47	1.45	1.43	1.42	1.39	1.38	1.36	1.35	1.35	1.26
180	1.75	1.72	1.69	1.67	1.65	1.61	1.57	1.54	1.52	1.48	1.46	1.43	1.42	1.40	1.38	1.36	1.35	1.34	1.33	1.25
200	1.74	1.71	1.68	1.66	1.64	1.60	1.56	1.53	1.51	1.47	1.45	1.42	1.41	1.39	1.37	1.35	1.34	1.33	1.32	1.23
∞	1.67	1.64	1.61	1.59	1.57	1.52	1.49	1.46	1.43	1.39	1.36	1.33	1.31	1.30	1.27	1.25	1.23	1.22	1.21	1.00

TABLE 6(c) Values of the F-Distribution: $A = .01$

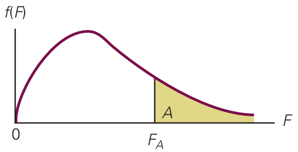


NUMERATOR DEGREES OF FREEDOM (ν_1); DENOMINATOR DEGREES OF FREEDOM (ν_2)

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4052	4999	5403	5625	5764	5859	5928	5981	6022	6056	6083	6106	6126	6143	6157	6170	6181	6192	6201	6209
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1	27.0	26.9	26.9	26.8	26.8	26.8	26.7	26.7
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.5	14.4	14.3	14.2	14.2	14.2	14.1	14.1	14.0	14.0
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89	9.82	9.77	9.72	9.68	9.64	9.61	9.58	9.55
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.60	7.56	7.52	7.48	7.45	7.42	7.40
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41	6.36	6.31	6.28	6.24	6.21	6.18	6.16
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61	5.56	5.52	5.48	5.44	5.41	5.38	5.36
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	5.05	5.01	4.96	4.92	4.89	4.86	4.83	4.81
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60	4.56	4.52	4.49	4.46	4.43	4.41
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.34	4.29	4.25	4.21	4.18	4.15	4.12	4.10
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05	4.01	3.97	3.94	3.91	3.88	3.86
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.91	3.86	3.82	3.78	3.75	3.72	3.69	3.66
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75	3.70	3.66	3.62	3.59	3.56	3.53	3.51
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56	3.52	3.49	3.45	3.42	3.40	3.37
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.50	3.45	3.41	3.37	3.34	3.31	3.28	3.26
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.40	3.35	3.31	3.27	3.24	3.21	3.19	3.16
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.32	3.27	3.23	3.19	3.16	3.13	3.10	3.08
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24	3.19	3.15	3.12	3.08	3.05	3.03	3.00
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13	3.09	3.05	3.02	2.99	2.96	2.94
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.07	3.02	2.98	2.94	2.91	2.88	2.85	2.83
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.98	2.93	2.89	2.85	2.82	2.79	2.76	2.74
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96	2.90	2.86	2.81	2.78	2.75	2.72	2.69	2.66
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90	2.84	2.79	2.75	2.72	2.68	2.65	2.63	2.60
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74	2.70	2.66	2.63	2.60	2.57	2.55
35	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.80	2.74	2.69	2.64	2.60	2.56	2.53	2.50	2.47	2.44
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56	2.52	2.48	2.45	2.42	2.39	2.37
45	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74	2.67	2.61	2.55	2.51	2.46	2.43	2.39	2.36	2.34	2.31
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	2.63	2.56	2.51	2.46	2.42	2.38	2.35	2.32	2.29	2.27
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44	2.39	2.35	2.31	2.28	2.25	2.22	2.20
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	2.45	2.40	2.35	2.31	2.27	2.23	2.20	2.18	2.15
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.36	2.31	2.27	2.23	2.20	2.17	2.14	2.12
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52	2.45	2.39	2.33	2.29	2.24	2.21	2.17	2.14	2.11	2.09
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.31	2.27	2.22	2.19	2.15	2.12	2.09	2.07
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34	2.28	2.23	2.19	2.15	2.12	2.09	2.06	2.03
140	6.82	4.76	3.92	3.46	3.15	2.93	2.77	2.64	2.54	2.45	2.38	2.31	2.26	2.21	2.17	2.13	2.10	2.07	2.04	2.01
160	6.80	4.74	3.91	3.44	3.13	2.92	2.75	2.62	2.52	2.43	2.36	2.30	2.24	2.20	2.15	2.11	2.08	2.05	2.02	1.99
180	6.78	4.73	3.89	3.43	3.12	2.90	2.74	2.61	2.51	2.42	2.35	2.28	2.23	2.18	2.14	2.10	2.07	2.04	2.01	1.98
200	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27	2.22	2.17	2.13	2.09	2.06	2.03	2.00	1.97
∞	6.64	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.19	2.13	2.08	2.04	2.00	1.97	1.94	1.91	1.88

ν_2 \ ν_1	22	24	26	28	30	35	40	45	50	60	70	80	90	100	120	140	160	180	200	∞
	NUMERATOR DEGREES OF FREEDOM																			
DENOMINATOR DEGREES OF FREEDOM																				
1	6223	6235	6245	6253	6261	6276	6287	6296	6303	6313	6321	6326	6331	6334	6339	6343	6346	6348	6350	6366
2	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
3	26.6	26.6	26.6	26.5	26.5	26.5	26.4	26.4	26.4	26.3	26.3	26.3	26.3	26.2	26.2	26.2	26.2	26.2	26.2	26.1
4	14.0	13.9	13.9	13.9	13.8	13.8	13.7	13.7	13.7	13.7	13.6	13.6	13.6	13.6	13.6	13.5	13.5	13.5	13.5	13.5
5	9.51	9.47	9.43	9.40	9.38	9.33	9.29	9.26	9.24	9.20	9.18	9.16	9.14	9.13	9.11	9.10	9.09	9.08	9.08	9.02
6	7.35	7.31	7.28	7.25	7.23	7.18	7.14	7.11	7.09	7.06	7.03	7.01	7.00	6.99	6.97	6.96	6.95	6.94	6.93	6.88
7	6.11	6.07	6.04	6.02	5.99	5.94	5.91	5.88	5.86	5.82	5.80	5.78	5.77	5.75	5.74	5.72	5.72	5.71	5.70	5.65
8	5.32	5.28	5.25	5.22	5.20	5.15	5.12	5.09	5.07	5.03	5.01	4.99	4.97	4.96	4.95	4.93	4.92	4.92	4.91	4.86
9	4.77	4.73	4.70	4.67	4.65	4.60	4.57	4.54	4.52	4.48	4.46	4.44	4.43	4.41	4.40	4.39	4.38	4.37	4.36	4.31
10	4.36	4.33	4.30	4.27	4.25	4.20	4.17	4.14	4.12	4.08	4.06	4.04	4.03	4.01	4.00	3.98	3.97	3.97	3.96	3.91
11	4.06	4.02	3.99	3.96	3.94	3.89	3.86	3.83	3.81	3.78	3.75	3.73	3.72	3.71	3.69	3.68	3.67	3.66	3.66	3.60
12	3.82	3.78	3.75	3.72	3.70	3.65	3.62	3.59	3.57	3.54	3.51	3.49	3.48	3.47	3.45	3.44	3.43	3.42	3.41	3.36
13	3.62	3.59	3.56	3.53	3.51	3.46	3.43	3.40	3.38	3.34	3.32	3.30	3.28	3.27	3.25	3.24	3.23	3.23	3.22	3.17
14	3.46	3.43	3.40	3.37	3.35	3.30	3.27	3.24	3.22	3.18	3.16	3.14	3.12	3.11	3.09	3.08	3.07	3.06	3.06	3.01
15	3.33	3.29	3.26	3.24	3.21	3.17	3.13	3.10	3.08	3.05	3.02	3.00	2.99	2.98	2.96	2.95	2.94	2.93	2.92	2.87
16	3.22	3.18	3.15	3.12	3.10	3.05	3.02	2.99	2.97	2.93	2.91	2.89	2.87	2.86	2.84	2.83	2.82	2.81	2.81	2.75
17	3.12	3.08	3.05	3.03	3.00	2.96	2.92	2.89	2.87	2.83	2.81	2.79	2.78	2.76	2.75	2.73	2.72	2.72	2.71	2.65
18	3.03	3.00	2.97	2.94	2.92	2.87	2.84	2.81	2.78	2.75	2.72	2.70	2.69	2.68	2.66	2.65	2.64	2.63	2.62	2.57
19	2.96	2.92	2.89	2.87	2.84	2.80	2.76	2.73	2.71	2.67	2.65	2.63	2.61	2.60	2.58	2.57	2.56	2.55	2.55	2.49
20	2.90	2.86	2.83	2.80	2.78	2.73	2.69	2.67	2.64	2.61	2.58	2.56	2.55	2.54	2.52	2.50	2.49	2.49	2.48	2.42
22	2.78	2.75	2.72	2.69	2.67	2.62	2.58	2.55	2.53	2.50	2.47	2.45	2.43	2.42	2.40	2.39	2.38	2.37	2.36	2.31
24	2.70	2.66	2.63	2.60	2.58	2.53	2.49	2.46	2.44	2.40	2.38	2.36	2.34	2.33	2.31	2.30	2.29	2.28	2.27	2.21
26	2.62	2.58	2.55	2.53	2.50	2.45	2.42	2.39	2.36	2.33	2.30	2.28	2.26	2.25	2.23	2.22	2.21	2.20	2.19	2.13
28	2.56	2.52	2.49	2.46	2.44	2.39	2.35	2.32	2.30	2.26	2.24	2.22	2.20	2.19	2.17	2.15	2.14	2.13	2.13	2.07
30	2.51	2.47	2.44	2.41	2.39	2.34	2.30	2.27	2.25	2.21	2.18	2.16	2.14	2.13	2.11	2.10	2.09	2.08	2.07	2.01
35	2.40	2.36	2.33	2.30	2.28	2.23	2.19	2.16	2.14	2.10	2.07	2.05	2.03	2.02	2.00	1.98	1.97	1.96	1.96	1.89
40	2.33	2.29	2.26	2.23	2.20	2.15	2.11	2.08	2.06	2.02	1.99	1.97	1.95	1.94	1.92	1.90	1.89	1.88	1.87	1.81
45	2.27	2.23	2.20	2.17	2.14	2.09	2.05	2.02	2.00	1.96	1.93	1.91	1.89	1.88	1.85	1.84	1.83	1.82	1.81	1.74
50	2.22	2.18	2.15	2.12	2.10	2.05	2.01	1.97	1.95	1.91	1.88	1.86	1.84	1.82	1.80	1.79	1.77	1.76	1.76	1.68
60	2.15	2.12	2.08	2.05	2.03	1.98	1.94	1.90	1.88	1.84	1.81	1.78	1.76	1.75	1.73	1.71	1.70	1.69	1.68	1.60
70	2.11	2.07	2.03	2.01	1.98	1.93	1.89	1.85	1.83	1.78	1.75	1.73	1.71	1.70	1.67	1.65	1.64	1.63	1.62	1.54
80	2.07	2.03	2.00	1.97	1.94	1.89	1.85	1.82	1.79	1.75	1.71	1.69	1.67	1.65	1.63	1.61	1.60	1.59	1.58	1.50
90	2.04	2.00	1.97	1.94	1.92	1.86	1.82	1.79	1.76	1.72	1.68	1.66	1.64	1.62	1.60	1.58	1.57	1.55	1.55	1.46
100	2.02	1.98	1.95	1.92	1.89	1.84	1.80	1.76	1.74	1.69	1.66	1.63	1.61	1.60	1.57	1.55	1.54	1.53	1.52	1.43
120	1.99	1.95	1.92	1.89	1.86	1.81	1.76	1.73	1.70	1.66	1.62	1.60	1.58	1.56	1.53	1.51	1.50	1.49	1.48	1.38
140	1.97	1.93	1.89	1.86	1.84	1.78	1.74	1.70	1.67	1.63	1.60	1.57	1.55	1.53	1.50	1.48	1.47	1.46	1.45	1.35
160	1.95	1.91	1.88	1.85	1.82	1.76	1.72	1.68	1.66	1.61	1.58	1.55	1.53	1.51	1.48	1.46	1.45	1.43	1.42	1.32
180	1.94	1.90	1.86	1.83	1.81	1.75	1.71	1.67	1.64	1.60	1.56	1.53	1.51	1.49	1.47	1.45	1.43	1.42	1.41	1.30
200	1.93	1.89	1.85	1.82	1.79	1.74	1.69	1.66	1.63	1.58	1.55	1.52	1.50	1.48	1.45	1.43	1.42	1.40	1.39	1.28
∞	1.83	1.79	1.76	1.73	1.70	1.64	1.59	1.56	1.53	1.48	1.44	1.41	1.38	1.36	1.33	1.30	1.28	1.26	1.25	1.00

TABLE 6(d) Values of the F-Distribution: $A = .005$



NUMERATOR DEGREES OF FREEDOM

ν_2 \ ν_1 (DENOMINATOR DEGREES OF FREEDOM)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	16211	19999	21615	22500	23056	23437	23715	23925	24091	24224	24334	24426	24505	24572	24630	24681	24727	24767	24803	24836
2	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199
3	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	43.5	43.4	43.3	43.2	43.1	43.0	42.9	42.9	42.8	42.8
4	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.8	20.7	20.6	20.5	20.4	20.4	20.3	20.3	20.2	20.2
5	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.5	13.4	13.3	13.2	13.1	13.1	13.0	13.0	12.9	12.9
6	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.3	10.1	10.0	9.95	9.88	9.81	9.76	9.71	9.66	9.62	9.59
7	16.2	12.4	10.9	10.1	9.52	9.16	8.89	8.68	8.51	8.38	8.27	8.18	8.10	8.03	7.97	7.91	7.87	7.83	7.79	7.75
8	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.10	7.01	6.94	6.87	6.81	6.76	6.72	6.68	6.64	6.61
9	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.31	6.23	6.15	6.09	6.03	5.98	5.94	5.90	5.86	5.83
10	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.75	5.66	5.59	5.53	5.47	5.42	5.38	5.34	5.31	5.27
11	12.2	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.32	5.24	5.16	5.10	5.05	5.00	4.96	4.92	4.89	4.86
12	11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.99	4.91	4.84	4.77	4.72	4.67	4.63	4.59	4.56	4.53
13	11.4	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.72	4.64	4.57	4.51	4.46	4.41	4.37	4.33	4.30	4.27
14	11.1	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.51	4.43	4.36	4.30	4.25	4.20	4.16	4.12	4.09	4.06
15	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.33	4.25	4.18	4.12	4.07	4.02	3.98	3.95	3.91	3.88
16	10.6	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.18	4.10	4.03	3.97	3.92	3.87	3.83	3.80	3.76	3.73
17	10.4	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	4.05	3.97	3.90	3.84	3.79	3.75	3.71	3.67	3.64	3.61
18	10.2	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.94	3.86	3.79	3.73	3.68	3.64	3.60	3.56	3.53	3.50
19	10.1	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.84	3.76	3.70	3.64	3.59	3.54	3.50	3.46	3.43	3.40
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.76	3.68	3.61	3.55	3.50	3.46	3.42	3.38	3.35	3.32
22	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.61	3.54	3.47	3.41	3.36	3.31	3.27	3.24	3.21	3.18
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.50	3.42	3.35	3.30	3.25	3.20	3.16	3.12	3.09	3.06
26	9.41	6.54	5.41	4.79	4.38	4.10	3.89	3.73	3.60	3.49	3.40	3.33	3.26	3.20	3.15	3.11	3.07	3.03	3.00	2.97
28	9.28	6.44	5.32	4.70	4.30	4.02	3.81	3.65	3.52	3.41	3.32	3.25	3.18	3.12	3.07	3.03	2.99	2.95	2.92	2.89
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.25	3.18	3.11	3.06	3.01	2.96	2.92	2.89	2.85	2.82
35	8.98	6.19	5.09	4.48	4.09	3.81	3.61	3.45	3.32	3.21	3.12	3.05	2.98	2.93	2.88	2.83	2.79	2.76	2.72	2.69
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	3.03	2.95	2.89	2.83	2.78	2.74	2.70	2.66	2.63	2.60
45	8.71	5.97	4.89	4.29	3.91	3.64	3.43	3.28	3.15	3.04	2.96	2.88	2.82	2.76	2.71	2.66	2.62	2.59	2.56	2.53
50	8.63	5.90	4.83	4.23	3.85	3.58	3.38	3.22	3.09	2.99	2.90	2.82	2.76	2.70	2.65	2.61	2.57	2.53	2.50	2.47
60	8.49	5.79	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.82	2.74	2.68	2.62	2.57	2.53	2.49	2.45	2.42	2.39
70	8.40	5.72	4.66	4.08	3.70	3.43	3.23	3.08	2.95	2.85	2.76	2.68	2.62	2.56	2.51	2.47	2.43	2.39	2.36	2.33
80	8.33	5.67	4.61	4.03	3.65	3.39	3.19	3.03	2.91	2.80	2.72	2.64	2.58	2.52	2.47	2.43	2.39	2.35	2.32	2.29
90	8.28	5.62	4.57	3.99	3.62	3.35	3.15	3.00	2.87	2.77	2.68	2.61	2.54	2.49	2.44	2.39	2.35	2.32	2.28	2.25
100	8.24	5.59	4.54	3.96	3.59	3.33	3.13	2.97	2.85	2.74	2.66	2.58	2.52	2.46	2.41	2.37	2.33	2.29	2.26	2.23
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.62	2.54	2.48	2.42	2.37	2.33	2.29	2.25	2.22	2.19
140	8.14	5.50	4.47	3.89	3.52	3.26	3.06	2.91	2.78	2.68	2.59	2.52	2.45	2.40	2.35	2.30	2.26	2.22	2.19	2.16
160	8.10	5.48	4.44	3.87	3.50	3.24	3.04	2.88	2.76	2.66	2.57	2.50	2.43	2.38	2.33	2.28	2.24	2.20	2.17	2.14
180	8.08	5.46	4.42	3.85	3.48	3.22	3.02	2.87	2.74	2.64	2.56	2.48	2.42	2.36	2.31	2.26	2.22	2.19	2.15	2.12
200	8.06	5.44	4.41	3.84	3.47	3.21	3.01	2.86	2.73	2.63	2.54	2.47	2.40	2.35	2.30	2.25	2.21	2.18	2.14	2.11
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.75	2.62	2.52	2.43	2.36	2.30	2.24	2.19	2.14	2.10	2.07	2.03	2.00

NUMERATOR DEGREES OF FREEDOM (ν_1); DENOMINATOR DEGREES OF FREEDOM (ν_2)

ν_2 \ ν_1	22	24	26	28	30	35	40	45	50	60	70	80	90	100	120	140	160	180	200	∞
1	24892	24940	24980	25014	25044	25103	25148	25183	25211	25253	25283	25306	25323	25337	25359	25374	25385	25394	25401	25464
2	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199	199
3	42.7	42.6	42.6	42.5	42.5	42.4	42.3	42.3	42.2	42.1	42.1	42.1	42.0	42.0	42.0	42.0	41.9	41.9	41.9	41.8
4	20.1	20.0	20.0	19.9	19.9	19.8	19.8	19.7	19.7	19.6	19.6	19.5	19.5	19.5	19.5	19.4	19.4	19.4	19.4	19.3
5	12.8	12.8	12.7	12.7	12.7	12.6	12.5	12.5	12.5	12.4	12.4	12.3	12.3	12.3	12.3	12.3	12.2	12.2	12.2	12.1
6	9.53	9.47	9.43	9.39	9.36	9.29	9.24	9.20	9.17	9.12	9.09	9.06	9.04	9.03	9.00	8.98	8.97	8.96	8.95	8.88
7	7.69	7.64	7.60	7.57	7.53	7.47	7.42	7.38	7.35	7.31	7.28	7.25	7.23	7.22	7.19	7.18	7.16	7.15	7.15	7.08
8	6.55	6.50	6.46	6.43	6.40	6.33	6.29	6.25	6.22	6.18	6.15	6.12	6.10	6.09	6.06	6.05	6.04	6.03	6.02	5.95
9	5.78	5.73	5.69	5.65	5.62	5.56	5.52	5.48	5.45	5.41	5.38	5.36	5.34	5.32	5.30	5.28	5.27	5.26	5.26	5.19
10	5.22	5.17	5.13	5.10	5.07	5.01	4.97	4.93	4.90	4.86	4.83	4.80	4.79	4.77	4.75	4.73	4.72	4.71	4.71	4.64
11	4.80	4.76	4.72	4.68	4.65	4.60	4.55	4.52	4.49	4.45	4.41	4.39	4.37	4.36	4.34	4.32	4.31	4.30	4.29	4.23
12	4.48	4.43	4.39	4.36	4.33	4.27	4.23	4.19	4.17	4.12	4.09	4.07	4.05	4.04	4.01	4.00	3.99	3.98	3.97	3.91
13	4.22	4.17	4.13	4.10	4.07	4.01	3.97	3.94	3.91	3.87	3.84	3.81	3.79	3.78	3.76	3.74	3.73	3.72	3.71	3.65
14	4.01	3.96	3.92	3.89	3.86	3.80	3.76	3.73	3.70	3.66	3.62	3.60	3.58	3.57	3.55	3.53	3.52	3.51	3.50	3.44
15	3.83	3.79	3.75	3.72	3.69	3.63	3.58	3.55	3.52	3.48	3.45	3.43	3.41	3.39	3.37	3.36	3.34	3.34	3.33	3.26
16	3.68	3.64	3.60	3.57	3.54	3.48	3.44	3.40	3.37	3.33	3.30	3.28	3.26	3.25	3.22	3.21	3.20	3.19	3.18	3.11
17	3.56	3.51	3.47	3.44	3.41	3.35	3.31	3.28	3.25	3.21	3.18	3.15	3.13	3.12	3.10	3.08	3.07	3.06	3.05	2.99
18	3.45	3.40	3.36	3.33	3.30	3.25	3.20	3.17	3.14	3.10	3.07	3.04	3.02	3.01	2.99	2.97	2.96	2.95	2.94	2.87
19	3.35	3.31	3.27	3.24	3.21	3.15	3.11	3.07	3.04	3.00	2.97	2.95	2.93	2.91	2.89	2.87	2.86	2.85	2.85	2.78
20	3.27	3.22	3.18	3.15	3.12	3.07	3.02	2.99	2.96	2.92	2.88	2.86	2.84	2.83	2.81	2.79	2.78	2.77	2.76	2.69
22	3.12	3.08	3.04	3.01	2.98	2.92	2.88	2.84	2.82	2.77	2.74	2.72	2.70	2.69	2.66	2.65	2.63	2.62	2.62	2.55
24	3.01	2.97	2.93	2.90	2.87	2.81	2.77	2.73	2.70	2.66	2.63	2.60	2.58	2.57	2.55	2.53	2.52	2.51	2.50	2.43
26	2.92	2.87	2.84	2.80	2.77	2.72	2.67	2.64	2.61	2.56	2.53	2.51	2.49	2.47	2.45	2.43	2.42	2.41	2.40	2.33
28	2.84	2.79	2.76	2.72	2.69	2.64	2.59	2.56	2.53	2.48	2.45	2.43	2.41	2.39	2.37	2.35	2.34	2.33	2.32	2.25
30	2.77	2.73	2.69	2.66	2.63	2.57	2.52	2.49	2.46	2.42	2.38	2.36	2.34	2.32	2.30	2.28	2.27	2.26	2.25	2.18
35	2.64	2.60	2.56	2.53	2.50	2.44	2.39	2.36	2.33	2.28	2.25	2.22	2.20	2.19	2.16	2.15	2.13	2.12	2.11	2.04
40	2.55	2.50	2.46	2.43	2.40	2.34	2.30	2.26	2.23	2.18	2.15	2.12	2.10	2.09	2.06	2.05	2.03	2.02	2.01	1.93
45	2.47	2.43	2.39	2.36	2.33	2.27	2.22	2.19	2.16	2.11	2.08	2.05	2.03	2.01	1.99	1.97	1.95	1.94	1.93	1.85
50	2.42	2.37	2.33	2.30	2.27	2.21	2.16	2.13	2.10	2.05	2.02	1.99	1.97	1.95	1.93	1.91	1.89	1.88	1.87	1.79
60	2.33	2.29	2.25	2.22	2.19	2.13	2.08	2.04	2.01	1.96	1.93	1.90	1.88	1.86	1.83	1.81	1.80	1.79	1.78	1.69
70	2.28	2.23	2.19	2.16	2.13	2.07	2.02	1.98	1.95	1.90	1.86	1.84	1.81	1.80	1.77	1.75	1.73	1.72	1.71	1.62
80	2.23	2.19	2.15	2.11	2.08	2.02	1.97	1.94	1.90	1.85	1.82	1.79	1.77	1.75	1.72	1.70	1.68	1.67	1.66	1.57
90	2.20	2.15	2.12	2.08	2.05	1.99	1.94	1.90	1.87	1.82	1.78	1.75	1.73	1.71	1.68	1.66	1.64	1.63	1.62	1.52
100	2.17	2.13	2.09	2.05	2.02	1.96	1.91	1.87	1.84	1.79	1.75	1.72	1.70	1.68	1.65	1.63	1.61	1.60	1.59	1.49
120	2.13	2.09	2.05	2.01	1.98	1.92	1.87	1.83	1.80	1.75	1.71	1.68	1.66	1.64	1.61	1.58	1.57	1.55	1.54	1.43
140	2.11	2.06	2.02	1.99	1.96	1.89	1.84	1.80	1.77	1.72	1.68	1.65	1.62	1.60	1.57	1.55	1.53	1.52	1.51	1.39
160	2.09	2.04	2.00	1.97	1.93	1.87	1.82	1.78	1.75	1.69	1.65	1.62	1.60	1.58	1.55	1.52	1.51	1.49	1.48	1.36
180	2.07	2.02	1.98	1.95	1.92	1.85	1.80	1.76	1.73	1.68	1.64	1.61	1.58	1.56	1.53	1.50	1.49	1.47	1.46	1.34
200	2.06	2.01	1.97	1.94	1.91	1.84	1.79	1.75	1.71	1.66	1.62	1.59	1.56	1.54	1.51	1.49	1.47	1.45	1.44	1.32
∞	1.95	1.90	1.86	1.82	1.79	1.72	1.67	1.63	1.59	1.54	1.49	1.46	1.43	1.40	1.37	1.34	1.31	1.30	1.28	1.00

TABLE **7(a)** Critical Values of the Studentized Range, $\alpha = .05$

	k																		
ν	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	18.0	27.0	32.8	37.1	40.4	43.1	45.4	47.4	49.1	50.6	52.0	53.2	54.3	55.4	56.3	57.2	58.0	58.8	59.6
2	6.08	8.33	9.80	10.9	11.7	12.4	13.0	13.5	14.0	14.4	14.7	15.1	15.4	15.7	15.9	16.1	16.4	16.6	16.8
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72	9.95	10.2	10.3	10.5	10.7	10.8	11.0	11.1	11.2
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.81	5.90	5.98	6.06	6.13	6.20	6.27	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61	5.71	5.80	5.88	5.95	6.02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.54	5.61	5.67	5.73	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

TABLE **7(b)** Critical Values of the Studentized Range, $\alpha = .01$

ν	k = 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	90.0	135	164	186	202	216	227	237	246	253	260	266	272	277	282	286	290	294	298
2	14.0	19.0	22.3	24.7	26.6	28.2	29.5	30.7	31.7	32.6	33.4	34.1	34.8	35.4	36.0	36.5	37.0	37.5	37.9
3	8.26	10.6	12.2	13.3	14.2	15.0	15.6	16.2	16.7	17.1	17.5	17.9	18.2	18.5	18.8	19.1	19.3	19.5	19.8
4	6.51	8.12	9.17	9.96	10.6	11.1	11.5	11.9	12.3	12.6	12.8	13.1	13.3	13.5	13.7	13.9	14.1	14.2	14.4
5	5.70	6.97	7.80	8.42	8.91	9.32	9.67	9.97	10.2	10.5	10.7	10.9	11.1	11.2	11.4	11.6	11.7	11.8	11.9
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.49	9.65	9.81	9.95	10.1	10.2	10.3	10.4	10.5
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	4.74	5.63	6.20	6.63	6.96	7.24	7.47	7.68	7.87	8.03	8.18	8.31	8.44	8.55	8.66	8.76	8.85	8.94	9.03
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.32	7.49	7.65	7.78	7.91	8.03	8.13	8.23	8.32	8.41	8.49	8.57
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.48	7.60	7.71	7.81	7.91	7.99	8.07	8.15	8.22
11	4.39	5.14	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	4.32	5.04	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
13	4.26	4.96	5.40	5.73	5.98	6.19	6.37	6.53	6.67	6.79	6.90	7.01	7.10	7.19	7.27	7.34	7.42	7.48	7.55
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77	6.87	6.96	7.05	7.12	7.20	7.27	7.33	7.39
15	4.17	4.83	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66	6.76	6.84	6.93	7.00	7.07	7.14	7.20	7.26
16	4.13	4.78	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48	6.57	6.66	6.73	6.80	6.87	6.94	7.00	7.05
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41	6.50	6.58	6.65	6.72	6.79	6.85	6.91	6.96
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.29	6.37	6.45	6.52	6.59	6.65	6.71	6.76	6.82
24	3.96	4.54	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	3.82	4.37	4.70	4.93	5.11	5.27	5.39	5.50	5.60	5.69	5.77	5.84	5.90	5.96	6.02	6.07	6.12	6.17	6.21
60	3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.79	5.84	5.89	5.93	5.98	6.02
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.38	5.44	5.51	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

Source: From E. S. Pearson and H. O. Hartley, *Biometrika Tables for Statisticians*, 1: 176–77. Reproduced by permission of the Biometrika Trustees.

TABLE **8(a)** Critical Values for the Durbin-Watson Statistic, $\alpha = .05$

	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	.95	1.54	.82	1.75	.69	1.97	.56	2.21
16	1.10	1.37	.98	1.54	.86	1.73	.74	1.93	.62	2.15
17	1.13	1.38	1.02	1.54	.90	1.71	.78	1.90	.67	2.10
18	1.16	1.39	1.05	1.53	.93	1.69	.82	1.87	.71	2.06
19	1.18	1.40	1.08	1.53	.97	1.68	.86	1.85	.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	.90	1.83	.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	.93	1.81	.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	.96	1.80	.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	.99	1.79	.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Source: From J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression, II," *Biometrika* 30 (1951): 159–78. Reproduced by permission of the Biometrika Trustees.

TABLE **8(b)** Critical Values for the Durbin-Watson Statistic, $\alpha = .01$

	$k = 1$		$k = 2$		$k = 3$		$k = 4$		$k = 5$	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	.81	1.07	.70	1.25	.59	1.46	.49	1.70	.39	1.96
16	.84	1.09	.74	1.25	.63	1.44	.53	1.66	.44	1.90
17	.87	1.10	.77	1.25	.67	1.43	.57	1.63	.48	1.85
18	.90	1.12	.80	1.26	.71	1.42	.61	1.60	.52	1.80
19	.93	1.13	.83	1.26	.74	1.41	.65	1.58	.56	1.77
20	.95	1.15	.86	1.27	.77	1.41	.68	1.57	.60	1.74
21	.97	1.16	.89	1.27	.80	1.41	.72	1.55	.63	1.71
22	1.00	1.17	.91	1.28	.83	1.40	.75	1.54	.66	1.69
23	1.02	1.19	.94	1.29	.86	1.40	.77	1.53	.70	1.67
24	1.04	1.20	.96	1.30	.88	1.41	.80	1.53	.72	1.66
25	1.05	1.21	.98	1.30	.90	1.41	.83	1.52	.75	1.65
26	1.07	1.22	1.00	1.31	.93	1.41	.85	1.52	.78	1.64
27	1.09	1.23	1.02	1.32	.95	1.41	.88	1.51	.81	1.63
28	1.10	1.24	1.04	1.32	.97	1.41	.90	1.51	.83	1.62
29	1.12	1.25	1.05	1.33	.99	1.42	.92	1.51	.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	.94	1.51	.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	.96	1.51	.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	.98	1.51	.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

Source: From J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least Squares Regression, II," *Biometrika* 30 (1951): 159–78. Reproduced by permission of the Biometrika Trustees.

TABLE 9 Critical Values for the Wilcoxon Rank Sum Test

(a) α = .025 one-tail; α = .05 two-tail

n_1	3		4		5		6		7		8		9		10	
n_2	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
4	6	18	11	25	17	33	23	43	31	53	40	64	50	76	61	89
5	6	11	12	28	18	37	25	47	33	58	42	70	52	83	64	96
6	7	23	12	32	19	41	26	52	35	63	44	76	55	89	66	104
7	7	26	13	35	20	45	28	56	37	68	47	81	58	95	70	110
8	8	28	14	38	21	49	29	61	39	63	49	87	60	102	73	117
9	8	31	15	41	22	53	31	65	41	78	51	93	63	108	76	124
10	9	33	16	44	24	56	32	70	43	83	54	98	66	114	79	131

(b) α = .05 one-tail; α = .10 two-tail

n_1	3		4		5		6		7		8		9		10	
n_2	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U	T_L	T_U
3	6	15	11	21	16	29	23	37	31	46	39	57	49	68	60	80
4	7	17	12	24	18	32	25	41	33	51	42	62	52	74	63	87
5	7	20	13	27	19	37	26	46	35	56	45	67	55	80	66	94
6	8	22	14	30	20	40	28	50	37	61	47	73	57	87	69	101
7	9	24	15	33	22	43	30	54	39	66	49	79	60	93	73	107
8	9	27	16	36	24	46	32	58	41	71	52	84	63	99	76	114
9	10	29	17	39	25	50	33	63	43	76	54	90	66	105	79	121
10	11	31	18	42	26	54	35	67	46	80	57	95	69	111	83	127

Source: From F. Wilcoxon and R. A. Wilcox, "Some Rapid Approximate Statistical Procedures" (1964), p. 28. Reproduced with the permission of American Cyanamid Company.

TABLE 10
Critical Values for the Wilcoxon Signed Rank Sum Test

	(a) $\alpha = .025$ one-tail; $\alpha = .05$ two-tail		(b) $\alpha = .05$ one-tail; $\alpha = .10$ two-tail	
n	T_L	T_U	T_L	T_U
6	1	20	2	19
7	2	26	4	24
8	4	32	6	30
9	6	39	8	37
10	8	47	11	44
11	11	55	14	52
12	14	64	17	61
13	17	74	21	70
14	21	84	26	79
15	25	95	30	90
16	30	106	36	100
17	35	118	41	112
18	40	131	47	124
19	46	144	54	136
20	52	158	60	150
21	59	172	68	163
22	66	187	75	178
23	73	203	83	193
24	81	219	92	208
25	90	235	101	224
26	98	253	110	241
27	107	271	120	258
28	117	289	130	276
29	127	308	141	294
30	137	328	152	313

Source: From F. Wilcoxon and R. A. Wilcox, "Some Rapid Approximate Statistical Procedures" (1964), p. 28. Reproduced with the permission of American Cyanamid Company.

TABLE 11 Critical Values for the Spearman Rank Correlation Coefficient

The α values correspond to a one-tail test of H_0: $\rho_s = 0$.
The value should be doubled for two-tail tests.

n	$\alpha = .05$	$\alpha = .025$	$\alpha = .01$
5	.900	–	–
6	.829	.886	.943
7	.714	.786	.893
8	.643	.738	.833
9	.600	.683	.783
10	.564	.648	.745
11	.523	.623	.736
12	.497	.591	.703
13	.475	.566	.673
14	.457	.545	.646
15	.441	.525	.623
16	.425	.507	.601
17	.412	.490	.582
18	.399	.476	.564
19	.388	.462	.549
20	.377	.450	.534
21	.368	.438	.521
22	.359	.428	.508
23	.351	.418	.496
24	.343	.409	.485
25	.336	.400	.475
26	.329	.392	.465
27	.323	.385	.456
28	.317	.377	.448
29	.311	.370	.440
30	.305	.364	.432

Source: From E. G. Olds, "Distribution of Sums of Squares of Rank Differences for Small Samples," *Annals of Mathematical Statistics* 9 (1938). Reproduced with the permission of the Institute of Mathematical Statistics.

TABLE 12 Control Chart Constants

SAMPLE SIZE n	A_2	d_2	d_3	D_3	D_4
2	1.880	1.128	.853	.000	3.267
3	1.023	1.693	.888	.000	2.575
4	.729	2.059	.880	.000	2.282
5	.577	2.326	.864	.000	2.115
6	.483	2.534	.848	.000	2.004
7	.419	2.704	.833	.076	1.924
8	.373	2.847	.820	.136	1.864
9	.337	2.970	.808	.184	1.816
10	.308	3.078	.797	.223	1.777
11	.285	3.173	.787	.256	1.744
12	.266	3.258	.778	.284	1.716
13	.249	3.336	.770	.308	1.692
14	.235	3.407	.762	.329	1.671
15	.223	3.472	.755	.348	1.652
16	.212	3.532	.749	.364	1.636
17	.203	3.588	.743	.379	1.621
18	.194	3.640	.738	.392	1.608
19	.187	3.689	.733	.404	1.596
20	.180	3.735	.729	.414	1.586
21	.173	3.778	.724	.425	1.575
22	.167	3.819	.720	.434	1.566
23	.162	3.858	.716	.443	1.557
24	.157	3.895	.712	.452	1.548
25	.153	3.931	.709	.459	1.541

Source: From E. S. Pearson, "The Percentage Limits for the Distribution of Range in Samples from a Normal Population," *Biometrika* 24 (1932): 416. Reproduced by permission of the Biometrika Trustees.

APPENDIX C

Answers to Selected Even-Numbered Exercises

All answers have been double-checked for accuracy. However, we cannot be absolutely certain that there are no errors. Students should not automatically assume that answers that don't match ours are wrong. When and if we discover mistakes we will post corrected answers on our web page. (See page 10 for the address.) If you find any errors, please email the author (address on web page). We will be happy to acknowledge you with the discovery.

Chapter 1

1.2 Descriptive statistics summarizes a set of data. Inferential statistics makes inferences about populations from samples.

1.4 **a.** The complete production run
b. 1,000 chips
c. Proportion defective
d. Proportion of sample chips that are defective (7.5%)
e. Parameter
f. Statistic
g. Because the sample proportion is less than 10%, we can conclude that the claim is true.

1.6 **a.** Flip the coin 100 times and count the number of heads and tails.
b. Outcomes of flips
c. Outcomes of the 100 flips
d. Proportion of heads
e. Proportion of heads in the 100 flips

1.8 **a.** Fuel mileage of all the taxis in the fleet.
b. Mean mileage
c. The 50 observations
d. Mean of the 50 observations
e. The statistic would be used to estimate the parameter from which the owner can calculate total costs.

We computed the sample mean to be 19.8 mpg.

Chapter 2

2.2 **a.** Interval **b.** Interval
c. Nominal **d.** Ordinal

2.4 **a.** Nominal **b.** Interval
c. Nominal **d.** Interval **e.** Ordinal

2.6 **a.** Interval **b.** Interval
c. Nominal **d.** Ordinal **e.** Interval

2.8 **a.** Interval **b.** Ordinal
c. Nominal **d.** Ordinal

2.10 **a.** Ordinal **b.** Ordinal **c.** Ordinal

2.34 Three out of four Americans are White. Note that the survey did not separate Hispanics.

2.36 Almost half the sample is married and about one out of four were never married.

2.38 The "Less than high school" category has remained constant, while the number of college graduates has increased.

2.40 The dominant source in Australia is coal. In New Zealand it is oil.

2.42 Universities 1 and 2 are similar and quite dissimilar from universities 3 and 4, which also differ. The two nominal variables appear to be related.

2.44 The two variables are related.

2.46 The number of prescriptions filled by independent drug stores has decreased while the others remained constant or increased slightly.

2.48 More than 40% rate the food as less than good.

2.50 There are considerable differences between the two countries.

2.52 Customers with children rated the restaurant more highly than did customers with no children.

2.54 **a.** Males and females differ in their areas of employment. Females tend to choose accounting, marketing, or sales and males opt for finance.
b. Area and job satisfaction are related. Graduates who work in finance and general management appear to be more satisfied than those in accounting, marketing, sales, and others.

Chapter 3

3.2 10 or 11

3.4 **a.** 7 to 9
b. 5.25, 5.40, 5.55, 5.70, 5.85, 6.00, 6.15

3.6 **c.** The number of pages is bimodal and slightly positively skewed.

3.8 The histogram is bimodal.

3.10 **c.** The number of stores is bimodal and positively skewed.

3.12 **d.** The histogram is symmetric (approximately) and bimodal.

3.14 **d.** The histogram is slightly positively skewed, unimodal, and not bell-shaped.

3.16 **a.** The histogram should contain 9 or 10 bins.
c. The histogram is positively skewed.
d. The histogram is not bell shaped.

3.18 The histogram is unimodal, bell shaped, and roughly symmetric. Most of the lengths lie between 18 and 23 inches.

3.20 The histogram is unimodal, symmetric, and bell shaped. Most tomatoes weigh between 2 and 7 ounces with a small fraction weighing less than 2 ounces or more than 7 ounces.

3.22 The histogram of the number of books shipped daily is negatively skewed. It appears that there is a maximum number that the company can ship.

3.24 **c. and d.** This scorecard is a much better predictor.

3.26 The histogram is highly positively skewed indicating that most people watch 4 or less hours per day with some watching considerably more.

3.28 Many people work more than 40 hours per week.

3.32 The numbers of females and males are both increasing with the number of females increasing faster.

3.34 The per capita number of property crimes decreased faster than did the absolute number of property crimes.

3.36 Consumption is increasing and production is falling.

3.38 **c.** Over the last 28 years, both receipts and outlays increased rapidly. There was a 5-year period where receipts were higher than outlays. Between 2004 and 2007, the deficit has decreased.

3.40 The inflation adjusted deficits are not large.

3.42 Imports from Canada has greatly exceeded exports to Canada.

3.44 In the early 1970s, the Canadian dollar was worth more than the U.S. dollar. By the late 1970s, the Canadian dollar lost ground but has recently recovered.

3.46 The index grew slowly until month 400 and then grew quickly until month 600. It then fell sharply and recently recovered.

3.48 There does not appear to be a linear relationship between the two variables.

3.50 **b.** There is a positive linear relationship between calculus and statistics marks.

3.52 **b.** There is a moderately strong positive linear relationship. In general, those with more education use the Internet more frequently.

3.54 **b.** There is a moderately strong positive linear relationship.

3.56 **b.** There is a very weak positive linear relationship.

3.58 There is a moderately strong positive linear relationship.

3.60 There is moderately strong positive linear relationship.
3.62 There does not appear to be any relationship between the two variables.
3.64 There does not appear to be a linear relationship.
3.66 There does not appear to be a linear relationship between the two variables.
3.68 There is a moderately strong positive linear relationship between the education levels of spouses.
3.70 There is a weak positive linear relationship between the amount of education of mothers and their children.
3.76 **c.** The accident rate generally decreases as the ages increase. The fatal accident rate decreases until the age of 64.
3.84 There has been a long-term decline in the value of the Australian dollar.
3.86 There is a very strong positive linear relationship.
3.88 **b.** The slope is positive.
c. There is a moderately strong linear relationship.
3.90 The value of the British pound has fluctuated quite a bit but the current exchange rate is close to the value in 1987.
3.92 **d.** The United States imports more products from Mexico than it exports to Mexico. Moreover, the trade imbalance is worsening (only interrupted by the recession in 2008–2009).
3.96 The number of fatal accidents and the number of deaths have been decreasing.
3.98 The histogram tells us that about 70% of gallery visitors stay for 60 minutes or less, and most of the remainder leave within 120 minutes.
3.100 The relationship between midterm marks and final marks appear to be similar for both statistics courses; that is, there is a weak positive linear relationship.

Chapter 4

4.2 $\bar{x} = 6$, median $= 5$, mode $= 5$
4.4 **a.** $\bar{x} = 39.3$, median $= 38$, mode $=$ all
4.6 $R_g = .19$
4.8 **a.** $\bar{x} = .106$, median $= .10$
b. $R_g = .102$ **c.** Geometric mean
4.10 **a.** .20, 0, .25, .33
b. $\bar{x} = .195$, median $= .225$
c. $R_g = .188$
d. Geometric mean
4.12 **a.** $\bar{x} = 75{,}750$, median $= 76{,}410$
4.14 **a.** $\bar{x} = 117.08$; median $= 124.00$
4.16 **a.** $\bar{x} = .81$; median $= .83$
4.18 **a.** $\bar{x} = 592.04$; median $= 591.00$
4.20 $s^2 = 1.14$
4.22 $s^2 = 15.12$, $s = 3.89$
4.24 **a.** $s^2 = 51.5$ **b.** $s^2 = 6.5$
c. $s^2 = 174.5$
4.26 6, 6, 6, 6, 6
4.28 **a.** 16% **b.** 97.5% **c.** 16%
4.30 **a.** Nothing
b. At least 75% lie between 60 and 180
c. At least 88.9% lie between 30 and 210
4.32 $s^2 = 40.73$ mph^2, and $s = 6.38$ mph; at least 75% of the speeds lie within 12.76 mph of the mean; at least 88.9% of the speeds lie within 19.14 mph of the mean.
4.34 $s^2 = .0858$ cm^2, and $s = .2929$cm; at least 75% of the lengths lie within .5858 of the mean; at least 88.9% of the rods will lie within .8787 cm of the mean.
4.36 **a.** $s = 15.01$
4.38 **a.** $\bar{x} = 77.86$ and $s = 85.35$
c. The histogram is positively skewed. At least 75% of American adults watch between 0 and 249 minutes of television.
4.40 3, 5, 7
4.42 44.6, 55.2
4.44 6.6, 17.6
4.46 4
4.50 **a.** 2, 4, 8
b. Most executives spend little time reading resumes. Keep it short.
4.52 50, 125, 260. The amounts are positively skewed.
4.54 **b.** 145.11, 164.17, 175.18
c. There are no outliers.
d. The data are positively skewed. One-quarter of the times are below 145.11, and one-quarter are above 175.18.
4.56 **a.** 26, 28.5, 32
b. the times are positively skewed.
4.58 Americans spend more time watching news on television than reading news on the Internet.
4.60 The two sets of numbers are quite similar.
4.62 1, 2, 4; The number of hours of television watching is highly positively skewed.
4.64 **a.** $-.7813$; there is a moderately strong negative linear relationship.
b. 61.04% of the variation in y is explained by the variation in x.
4.66 **a.** 98.52 **b.** .8811 **c.** .7763
d. $\hat{y} = 5.917 + 1.705x$
e. There is a strong positive linear relationship between marks and study time. For each additional hour of study time, marks increased on average by 1.705.
4.68 40.09% of the variation in the employment rate is explained by the variation in the unemployment rate.
4.70 Only 5.93% of the variation in the number of houses sold is explained by the variation in interest rates.
4.72 $R^2 = .0069$. There is a very weak positive relationship between the two variables.
4.74 $\hat{y} = 263.4 + 71.65x$. Estimated fixed costs $= \$263.40$, estimated variable costs $= \$71.65$.
4.76 **a.** $R^2 = .0915$; there is a very weak relationship between the two variables.
b. The slope coefficient is 58.59; away attendance increases on average by 58.59 for each win. However, the relationship is very weak.
4.78 **a.** The slope coefficient is .0428; for each million dollars in payroll, the number of wins increases on average by .0428. Thus, the cost of winning one additional game is 1/.0428 million $=$ \$23.364 million.
b. The coefficient of determination $= .0866$, which reveals that the linear relationship is very weak.
4.80 **a.** For each additional win, home attendance increases on average by 84.391. The coefficient of determination is .2468; there is a weak relationship between the number of wins and home attendance.
b. For each additional win, away attendance increases on average by 31.151. The coefficient of determination is .4407; there is a moderately strong relationship between the number of wins and away attendance.
4.82 For each additional win, home attendance increases on average by 947.38. The coefficient of determination is .1108; there is a very weak linear relationship between the number of wins and home attendance.
For each additional win, away attendance increases on average by 216.74. The coefficient of determination is .0322; there is a very weak linear relationship between the number of wins and away attendance.
4.84 **a.** There is a weak negative linear relationship between education and television watching.
b. $R^2 = .0572$; 5.72% of the variation in the amount of television is explained by the variation in education.
4.86 $r = .2107$; there is a weak positive linear relationship between the two variables.
4.90 **b.** We can see that among those who repaid the mean score is larger than that of those who did not and the standard deviation is smaller. This information is similar but more precise than that obtained in Exercise 3.23.
4.92 46.03% of the variation in statistics marks is explained by the variation in calculus marks. The coefficient of determination provides a more precise indication of the strength of the linear relationship.
4.94 **a.** $\hat{y} = 17.933 + .6041x$
b. The coefficient of determination is .0505, which indicates that

only 5.05% of the variation in incomes is explained by the variation in heights.

4.96 **a.** $\hat{y} = 103.44 + .07x$
b. The slope coefficient is .07. For each additional square foot, the price increases an average of $.07 thousand. More simply, for each additional square foot the price increases on average by $70.
c. From the least squares line, we can more precisely measure the relationship between the two variables.

4.100 **a.** $\bar{x} = 29{,}913$, median = 30,660
b. $s^2 = 148{,}213{,}791$; $s = 12{,}174$
d. The number of coffees sold varies considerably.

4.102 **a. & b.** $R^2 = .5489$ and the least squares line is $\hat{y} = 49{,}337 - 553.7x$.
c. 54.8% of the variation in the number of coffees sold is explained by the variation in temperature. For each additional degree of temperature, the number of coffees sold decreases on average by 553.7 cups. Alternatively for each 1-degree drop in temperature, the number of coffees increases, on average, by 553.7 cups.
d. We can measure the strength of the linear relationship accurately, and the slope coefficient gives information about how temperature and the number of coffees sold are related.

4.104 **a.** $\bar{x} = 26.32$ and median = 26
b. $s^2 = 88.57$, $s = 9.41$
d. The times are positively skewed. Half the times are above 26 hours.

4.106 **a. & b.** $R^2 = .412$, and the least squares line is $\hat{y} = -8.2897 + 3.146x$.
c. 41.2% of the variation in Internet use is explained by the variation in education. For each additional year of education, Internet use increases on average by 3.146 hours.
d. We can measure the strength of the linear relationship accurately and the slope coefficient gives information about how education and Internet use are related.

4.108 **a. & b.** $R^2 = .369$, and the least squares line is $\hat{y} = 89.543 + .128$ rainfall.
c. 36.92% of the variation in yield is explained by the variation in rainfall. For each additional inch of rainfall, yield increases on average by .128 bushels.
d. We can measure the strength of the linear relationship accurately, and the slope coefficient gives information about how rainfall and crop yield are related.

4.110 **b.** The mean debt is $12,067. Half the sample incurred debts below $12,047 and half incurred debts above. The mode is $11,621.

Chapter 6

6.2 **a.** Subjective approach
b. If all the teams in major league baseball have exactly the same players, the New York Yankees will win 25% of all World Series.
6.4 **a.** Subjective approach
b. The Dow Jones Industrial Index will increase on 60% of the days if economic conditions remain unchanged.
6.6 {Adams wins. Brown wins, Collins wins, Dalton wins}
6.8 **a.** {0, 1, 2, 3, 4, 5} **b.** {4, 5} **c.** .10 **d.** .65 **e.** 0
6.10 2/6, 3/6, 1/6
6.12 **a.** .40 **b.** .90
6.14 **a.** P(single) = .15, P(married) = .50, P(divorced) = .25, P(widowed) = .10
b. Relative frequency approach
6.16 $P(A_1) = .3$, $P(A_2) = .4$, $P(A_3) = .3$. $P(B_1) = .6$, $P(B_2) = .4$.
6.18 **a.** .57 **b.** .43
c. It is not a coincidence.
6.20 The events are not independent.
6.22 The events are independent.
6.24 $P(A_1) = .40$, $P(A_2) = .45$, $P(A_3) = .15$. $P(B_1) = .45$, $P(B_2) = .55$.
6.26 **a.** .85. **b.** .75 **c.** .50
6.28 **a.** .36 **b.** .49 **c.** .83
6.30 **a.** .31 **b.** .85 **c.** .387 **d.** .043
6.32 **a.** .390 **b.** .66 **c.** No
6.34 **a.** .11 **b.** .043 **c.** .091 **d.** .909
6.36 **a.** .33 **b.** 30
c. Yes, the events are dependent.
6.38 **a.** .778 **b.** .128 **c.** .385
6.40 **a.** .636 **b.** .205
6.42 **a.** .848 **b.** .277 **c.** .077
6.44 No
6.46 **a.** .201 **b.** .199 **c.** .364 **d.** .636
6.52 **a.** .81 **b.** .01 **c.** .18 **d.** .99
6.54 **b.** .8091 **c.** .0091 **d.** .1818 **e.** .9909
6.56 **a.** .28 **b.** .30 **c.** .42
6.58 .038
6.60 .335
6.62 .698
6.64 .2520
6.66 .033
6.68 .00000001
6.70 .6125
6.72 **a.** .696 **b.** .304 **c.** .889 **d.** .111
6.74 .526
6.76 .327
6.78 .661
6.80 .593
6.82 .843
6.84 .920, .973, .1460, .9996
6.86 **a.** .290 **b.** .290 **c.** Yes
6.88 **a.** .19 **b.** .517 **c.** No
6.90 .295
6.92 .825
6.94 **a.** .3285 **b.** .2403
6.96 .9710
6.98 2/3
6.100 .2214
6.102 .3333

Chapter 7

7.2 **a.** any value between 0 and several hundred miles
b. No **c.** No **d.** continuous
7.4 **a.** 0, 1, 2, . . . , 100 **b.** Yes
c. Yes, 101 values **d.** discrete.
7.6 $P(x) = 1/6$, for $x = 1, 2, \ldots, 6$
7.8 **a.** .950 .020 .680
b. 3.066
c. 1.085
7.10 **a.** .8 **b.** .8 **c.** .8 **d.** .3
7.12 .0156
7.14 **a.** .25 **b.** .25 **c.** .25 **d.** .25
7.18 **a.** 1.40, 17.04 **c.** 7.00, 426.00
d. 7.00, 426.00
7.20 **a.** .6 **b.** 1.7, .81
7.22 **a.** .40 **b.** .95
7.24 1.025, .168
7.26 **a.** .06 **b.** 0 **c.** .35 **d.** .65
7.28 **a.** .21 **b.** .31 **c.** .26
7.30 2.76, 1.517
7.32 3.86, 2.60
7.34 E(value of coin) = $460; take the $500
7.36 $18
7.38 4.00, 2.40
7.40 1.85
7.42 3,409
7.44 .14, .58
7.46 **b.** 2.8, .76
7.48 0, 0
7.50 **b.** 2.9, .45, **c.** yes
7.54 **c.** 1.07, .505 **d.** .93, .605
e. −.045, −.081
7.56 **a.** .412 **b.** .286 **c.** .148
7.58 145, 31
7.60 168, 574
7.62 **a.** .211, .1081 **b.** .211, .1064
c. .211 .1052
7.64 .1060, .1456
7.68 Coca-Cola and McDonalds: .01180, .04469
7.70 .00720, .04355
7.72 .00884, .07593
7.74 Fortis and RIM: .01895, .08421
7.78 .00913, .05313
7.84 **a.** .2668 **b.** .1029 **c.** .0014
7.86 **a.** .26683 **b.** .10292 **c.** .00145
7.88 **a.** .2457 **b.** .0819 **c.** .0015
7.90 **a.** .1711 **b.** .0916 **c.** .9095
d. .8106
7.92 **a.** .4219 **b.** .3114 **c.** .25810
7.94 **a.** .0646 **b.** .9666 **c.** .9282
d. 22.5
7.96 .0081
7.98 .1244
7.100 .00317
7.102 **a.** .3369 **b.** .75763
7.104 **a.** .2990 **b.** .91967
7.106 **a.** .69185 **b.** .12519 **c.** .44069
7.108 **a.** .05692 **b.** .47015
7.110 **a.** .1353 **b.** .1804 **c.** .0361
7.112 **a.** .0302 **b.** .2746 **c.** .3033
7.114 **a.** .1353 **b.** .0663

7.116 **a.** .20269 **b.** .26761
7.118 .6703
7.120 **a.** .4422 **b.** .1512
7.122 **a.** .2231 **b.** .7029 **c.** .5768
7.124 **a.** .8 **b.** .4457
7.126 **a.** .0993 **b.** .8088 **c.** .8881
7.128 .0473
7.130 .0064
7.132 **a.** .00793 **b.** 56 **c.** 4.10
7.134 **a.** .1612 **b.** .0095 **c.** .0132
7.136 **a.** 1.46, 1.49 **b.** 2.22, 1.45
7.138 .08755
7.140 .95099, .04803, .00097, .00001, 0, 0

Chapter 8

8.2 **a.** .1200 **b.** .4800 **c.** .6667 **d.** .1867
8.4 **b.** 0 **c.** .25 **d.** 005
8.6 **a.** .1667 **b.** .3333 **c.** 0
8.8 57 minutes
8.10 123 tons
8.12 **b.** .5 **c.** .25
8.14 **b.** .25 **c.** .33
8.16 .9345
8.18 .0559
8.20 .0107
8.22 .9251
8.24 .0475
8.26 .1196
8.28 .0010
8.30 0
8.32 1.70
8.34 .0122
8.36 .4435
8.38 **a.** .6759 **b.** .3745 **c.** .1469
8.40 .6915
8.42 **a.** .2023 **b.** .3372
8.44 **a.** .1056 **b.** .1056 **c.** .8882
8.46 Top 5%: 34.4675. Bottom 5%: 29.5325
8.48 .1151
8.50 **a.** .1170 **b.** .3559 **c.** .0162 **d.** 4.05 hours
8.52 9,636 pages
8.54 **a.** .3336 **b.** .0314 **c.** .0436 **d.** $32.88
8.56 **a.** .0099 **b.** $12.88
8.58 132.80 (rounded to 133)
8.60 .5948
8.62 .0465
8.64 171
8.66 873
8.68 .8159
8.70 **a.** .2327 **b.** .2578
8.74 **a.** .5488 **b.** .6988 **c.** .1920 **d.** 0
8.76 .1353
8.78 .8647
8.80 .4857
8.82 .1889
8.84 **a.** 2.750 **b.** 1.282 **c.** 2.132 **d.** 2.528
8.86 **a.** 1.6556 **b.** 2.6810 **c.** 1.9600 **d.** 1.6602
8.88 **a.** .1744 **b.** .0231 **c.** .0251 **d.** .0267
8.90 **a.** 17.3 **b.** 50.9 **c.** 2.71 **d.** 53.5
8.92 **a.** 33.5705 **b.** 866.911 **c.** 24.3976 **d.** 261.058
8.94 **a.** .4881 **b.** .9158 **c.** .9988 **d.** .9077
8.96 **a.** 2.84 **b.** 1.93 **c.** 3.60 **d.** 3.37
8.98 **a.** 1.5204 **b.** 1.5943 **c.** 2.8397 **d.** 1.1670
8.100 **a.** .1050 **b.** .1576 **c.** .0001 **d.** .0044

Chapter 9

9.2 **a.** 1/36 **b.** 1/36
9.4 The variance of $\bar{X}$ is smaller than the variance of X.
9.6 No, because the sample mean is approximately normally distributed.
9.8 **a.** .1056 **b.** .1587 **c.** .0062
9.10 **a.** .4435 **b.** .7333 **c.** .8185
9.12 **a.** .1191 **b.** .2347 **c.** .2902
9.14 **a.** 15.00 **b.** 21.80 **c.** 49.75
9.18 **a.** .0918 **b.** .0104 **c.** .00077
9.20 **a.** .3085 **b.** 0
9.22 **a.** .0038 **b.** It appears to be false.
9.26 .1170
9.28 .9319
9.30 **a.** 0 **b.** .0409 **c.** .5
9.32 .1056
9.34 .0035
9.36 **a.** .1151 **b.** .0287
9.38 .0096; the commercial is dishonest.
9.40 **a.** .0071 **b.** The claim appears to be false.
9.42 .0066
9.44 The claim appears to be false.
9.46 .0033
9.48 .8413
9.50 .8413
9.52 .3050
9.54 1

Chapter 10

10.10 **a.** 200 ± 19.60 **b.** 200 ± 9.80 **c.** 200 ± 3.92 **d.** The interval narrows.
10.12 **a.** 500 ± 3.95 **b.** 500 ± 3.33 **c.** 500 ± 2.79 **d.** The interval narrows.
10.14 **a.** 10 ± .82 **b.** 10 ± 1.64 **c.** 10 ± 2.60 **d.** The interval widens.
10.16 **a.** 400 ± 1.29 **b.** 200 ± 1.29 **c.** 100 ± 1.29 **d.** The width of the interval is unchanged.
10.18 Yes, because the variance decreases as the sample size increases.
10.20 **a.** 500 ± 3.50
10.22 LCL = 36.82, UCL = 50.68
10.24 LCL = 6.91, UCL = 12.79
10.26 LCL = 12.83, UCL = 20.97
10.28 LCL = 10.41, UCL = 15.89
10.30 LCL = 249.44, UCL = 255.32
10.32 LCL = 11.86, UCL = 12.34
10.34 LCL = .494, UCL = .526
10.36 LCL = 18.66, UCL = 19.90
10.38 LCL = 579,545, UCL = 590,581
10.40 LCL = 25.62, UCL = 28.76
10.48 **a.** 1,537 **b.** 500 ± 10
10.52 2,149
10.54 1,083
10.56 217

Chapter 11

11.2 H_0: I will complete the Ph.D. H_1: I will not be able to complete the Ph.D.
11.4 H_0: Risky investment is more successful H_1: Risky investment is not more successful
11.6 O. J. Simpson

All p-values and probabilities of Type II errors were calculated manually using Table 3 in Appendix B.

11.8 $z = .60$; rejection region: $z > 1.88$; p-value = .2743; not enough evidence that $\mu > 50$.
11.10 $z = 0$; rejection region: $z < -1.96$ or $z > 1.96$; p-value = 1.0; not enough evidence that $\mu \neq 100$.
11.12 $z = -1.33$; rejection region: $z < -1.645$; p-value = .0918; not enough evidence that $\mu < 50$
11.14 **a.** .2743 **b.** .1587 **c.** .0013 **d.** The test statistics decreases and the p-value decreases.
11.16 **a.** .2112 **b.** .3768 **c.** .5764 **d.** The test statistic increases and the p-value increases.
11.18 **a.** .0013 **b.** .0228 **c.** .1587 **d.** The test statistic decreases and the p-value increases.
11.20 **a.** $z = 4.57$, p-value = 0 **b.** $z = 1.60$, p-value = .0548.
11.22 **a.** $z = -.62$, p-value = .2676 **b.** $z = -1.38$, p-value = .0838
11.24 p-values: .5, .3121, .1611, .0694, .0239, .0062, .0015, 0, 0
11.26 **a.** $z = 2.30$, p-value = .0214 **b.** $z = .46$, p-value = .6456
11.28 $z = 2.11$, p-value = .0174; yes
11.30 $z = -1.29$, p-value = .0985; yes
11.32 $z = .95$, p-value = .1711; no
11.34 $z = 1.85$, p-value = .0322; no
11.36 $z = -2.06$, p-value = .0197; yes
11.38 **a.** $z = 1.65$, p-value = .0495; yes
11.40 $z = 2.26$, p-value = .0119; no
11.42 $z = -1.22$, p-value = .1112; no
11.44 $z = 3.33$, p-value = 0; yes
11.46 $z = -2.73$, p-value = .0032; yes
11.48 .1492
11.50 .6480
11.52 **a.** .6103 **b.** .8554 **c.** β increases.
11.56 **a.** .4404 **b.** .6736 **c.** β increases.
11.60 p-value = .9931; no evidence that the new system will not be cost effective.
11.62 .1170
11.64 .1635 (with $\alpha = .05$)

The answers for the exercises in Chapters 12 through 19 were produced in the following way. In exercises where the statistics are provided in the question or in Appendix A, the solutions were produced manually. The solutions to exercises requiring the use of a computer were produced using Excel. When the test result is calculated manually and the test statistic is normally distributed (z statistic) the p-value was computed manually using the normal table (Table 3 in Appendix B). The p-value for all other test statistics was determined using Excel.

Chapter 12

12.4 **a.** 1500 ± 59.52 **b.** 1500 ± 39.68 **c.** 1500 ± 19.84 **d.** Interval narrows
12.6 **a.** $10 \pm .20$ **b.** $10 \pm .79$ **c.** 10 ± 1.98 **d.** Interval widens
12.8 **a.** 63 ± 1.77 **b.** 63 ± 2.00 **c.** 63 ± 2.71 **d.** Interval widens
12.10 **a.** $t = -3.21$, p-value = .0015 **b.** $t = -1.57$, p-value = .1177 **c.** $t = -1.18$, p-value = .2400 **d.** t decreases and p-value increases
12.12 **a.** $t = .67$, p-value = .5113 **b.** $t = .52$, p-value = .6136 **c.** $t = .30$, p-value = .7804 **d.** t decreases and p-value increases
12.14 **a.** $t = 1.71$, p-value = .0448 **b.** $t = 2.40$, p-value = .0091 **c.** $t = 4.00$, p-value = .0001 **d.** t increase and p-value decreases
12.16 **a.** 175 ± 28.60 **b.** 175 ± 22.07 **c.** Because the distribution of Z is narrower than that of the Student t
12.18 **a.** 350 ± 11.56 **b.** 350 ± 11.52 **c.** When n is large the distribution of Z is virtually identical to that of the Student t
12.20 **a.** $t = -1.30$, p-value = .1126 **b.** $z = -1.30$, p-value = .0968 **c.** Because the distribution of Z is narrower than that of the Student t
12.22 **a.** $t = 1.58$, p-value = .0569 **b.** $z = 1.58$, p-value = .0571 **c.** When n is large the distribution of Z is virtually identical to that of the Student t
12.24 LCL = 14,422, UCL = 33,680
12.26 $t = -4.49$, p-value = .0002; yes
12.28 LCL = 18.11, UCL = 35.23
12.30 $t = -2.45$, p-value = .0185; yes
12.32 LCL = 427 million, UCL = 505 million
12.34 LCL = \$727,350 million, UCL = \$786,350 million
12.36 LCL = 2.31, UCL = 3.03
12.38 LCL = \$51,725 million, UCL = \$56,399 million
12.40 $t = .51$, p-value = .3061; no
12.42 $t = 2.28$, p-value = .0127; yes
12.44 LCL = 650,958 million, UCL = 694,442 million
12.46 $t = 20.89$, p-value = 0; yes
12.48 $t = 4.80$, p-value = 0; yes
12.50 LCL = 2.85, UCL = 3.02
12.52 LCL = 4.80, UCL = 5.12
12.56 **a.** $X^2 = 72.60$, p-value = .0427 **b.** $X^2 = 35.93$, p-value = .1643
12.58 **a.** LCL = 7.09, UCL = 25.57 **b.** LCL = 8.17, UCL = 19.66
12.60 $\chi^2 = 7.57$, p-value = .4218; no
12.62 LCL = 7.31, UCL = 51.43
12.64 $\chi^2 = 305.81$, p-value = .0044; yes
12.66 $\chi^2 = 86.36$, p-value = .1863; no
12.70 **a.** $.48 \pm .0438$ **b.** $.48 \pm .0692$ **c.** $.48 \pm .0310$
12.72 **a.** $z = .61$, p-value = .2709 **b.** $z = .87$, p-value = .1922 **c.** $z = 1.22$, p-value = .1112
12.74 752
12.76 **a.** $.75 \pm .0260$
12.78 **a.** $.75 \pm .03$
12.80 **a.** $.5 \pm .0346$
12.82 $z = -1.47$, p-value = .0708; yes
12.84 $z = .33$, p-value = .3707; no
12.86 LCL = .1332, UCL = .2068
12.88 LCL = 0, UCL = .0312
12.90 LCL = 0, UCL = .0191
12.92 LCL = 5,940, UCL = 9,900
12.94 $z = 1.58$, p-value = .0571; no
12.96 LCL = 3.45 million, UCL = 3.75 million
12.98 $z = 1.40$, p-value = .0808; yes
12.100 LCL = 4.945 million, UCL = 6.325 million
12.102 LCL = .861 million, UCL = 1.17 million
12.104 **a.** LCL = .4780, UCL = .5146 **b.** LCL = .0284, UCL = .0448
12.106 LCL = .1647, UCL = .1935
12.108 $z = 6.00$, p-value = 0; yes
12.110 $z = 3.87$, p-value = 0; yes
12.112 $z = 5.63$, p-value = 0; yes
12.114 $z = 15.08$, p-value = 0; yes
12.116 $z = 7.27$, p-value = 0; yes
12.118 $z = 5.05$, p-value = 0; yes
12.120 LCL = 35,121,043, UCL = 43,130,297
12.122 $z = -.539$, p-value = .5898.
12.124 LCL = 13,195,985, UCL = 14,720,803
12.126 **a.** LCL = .2711, UCL = .3127 **b.** LCL = 29,060,293, UCL = 33,519,564
12.128 LCL = 26.928 million, UCL = 38.447 million
12.130 **a.** $t = 3.04$, p-value = .0015; yes **b.** LCL = 30.68, UCL = 33.23 **c.** The costs are required to be normally distributed.
12.132 $\chi^2 = 30.71$, p-value = .0435; yes
12.134 **a.** LCL = 69.03, UCL = 74.73 **b.** $t = 2.74$, p-value = .0043; yes
12.136 LCL = .582, UCL = .682
12.138 LCL = 6.05, UCL = 6.65
12.140 LCL = .558, UCL = .776
12.142 $z = -1,33$, p-value = .0912; yes
12.144 **a.** $t = -2.97$, p-value = .0018; yes **b.** $\chi^2 = 101.58$, p-value = .0011; yes
12.146 LCL = 49,800, UCL = 72,880
12.148 **a.** LCL = −5.54%, UCL = 29.61% **b.** $t = -.47$, p-value = .3210; no
12.150 $t = .908$, p-value = .1823; no
12.152 $t = .959$, p-value = .1693; no
12.154 $t = 2.44$, p-value = .0083; yes

For all exercises in Chapter 13 and all chapter appendixes, we employed the F-test of two variances at the 5% significance level to decide which one of the equal-variances or unequal-variances t-test and estimator of the difference between two means to use to solve the problem. In addition, for exercises that compare two populations and are accompanied by data files, our answers were derived by defining the sample from population 1 as the data stored in the first column (often column A). The data stored in the second column represent the sample from population 2. Paired differences were defined as the difference between the variable in the first column minus the variable in the second column.

Chapter 13

13.6 **a.** $t = .43$, p-value = .6703; no **b.** $t = .04$, p-value = .9716; no **c.** The t-statistic decreases and the p-value increases. **d.** $t = 1.53$, p-value = .1282; no **e.** The t-statistic increases and the p-value decreases. **f.** $t = .72$, p-value = .4796; no **g.** The t-statistic increases and the p-value decreases.
13.8 **a.** $t = .62$, p-value = .2689; no **b.** $t = 2.46$, p-value = .0074; yes **c.** The t-statistic increases and the p-value decreases. **d.** $t = .23$, p-value = .4118 **e.** The t-statistic decreases and the p-value increases. **f.** $t = .35$, p-value = .3624 **g.** The t-statistic decreases and the p-value increases.
13.12 $t = -2.04$, p-value = .0283; yes
13.14 $t = -1.59$, p-value = .1368; no
13.16 $t = 1.12$, p-value = .2761; no
13.18 $t = 1.55$, p-value = .1204; no
13.20 **a.** $t = 2.88$ p-value = .0021; yes **b.** LCL = .25, UCL = 4.57
13.22 $t = .94$, p-value = .1753; switch to supplier B.
13.24 **a.** $t = 2.94$, p-value = .0060; yes **b.** LCL = 4.31, UCL = 23.65 **c.** The times are required to be normally distributed.
13.26 $t = 7.54$, p-value = 0; yes
13.28 $t = .90$, p-value = .1858; no
13.30 $t = -2.05$, p-value = .0412; yes
13.32 $t = 1.16$, p-value = .2467; no
13.34 $t = -2.09$, p-value = .0189; yes
13.36 $t = 6.28$, p-value = 0; yes
13.38 LCL = 13,282, UCL = 21,823
13.42 $t = -4.65$, p-value = 0; yes
13.44 $t = 9.20$, p-value = 0; yes
13.46 Experimental
13.52 $t = -3.22$, p-value = .0073; yes
13.54 $t = 1.98$, p-value = .0473; yes
13.56 **a.** $t = 1.82$, p-value = .0484; yes **b.** LCL = −.66, UCL = 6.82
13.58 $t = -3.70$, p-value = .0006; yes
13.60 **a.** $t = 16.92$, p-value = 0; yes **b.** LCL = 50.12, UCL = 64.48 **c.** Differences are required to be normally distributed.

13.62 $t = -1.52$, p-value $= .0647$; no
13.64 $t = 2.08$, p-value $= .0210$; yes
13.70 $t = 23.35$, p-value $= 0$; yes
13.72 $t = 2.22$, p-value $= .0132$; yes
13.76 **a.** $F = .50$, p-value $= .0669$; yes
b. $F = .50$, p-value $= .2071$; no
c. The value of the test statistic is unchanged but the conclusion did change.
13.78 $F = .50$, p-value $= .3179$; no
13.80 $F = 3.23$, p-value $= .0784$; no
13.82 $F = 2.08$, p-value $= .0003$; yes
13.84 $F = .31$, p-value $= 0$; yes
13.88 **a.** $z = 1.07$, p-value $= .2846$
b. $z = 2.01$, p-value $= .0444$
c. The p-value decreases.
13.90 $z = 1.70$, p-value $= .0446$; yes
13.92 $z = 1.74$, p-value $= .0409$; yes.
13.94 $z = -2.85$, p-value $= .0022$; yes
13.96 **a.** $z = -4.04$, p-value $= 0$; yes
13.98 $z = 2.00$, p-value $= .0228$; yes
13.100 $z = -1.19$, p-value $= .1170$; no
13.102 **a.** $z = 3.35$, p-value $= 0$; yes
b. LCL $= .0668$, UCL $= .3114$
13.104 $z = -4.24$, p-value $= 0$; yes
13.106 $z = 1.50$, p-value $= .0664$; no
13.108 Canada: $z = 2.82$, p-value $= .0024$; yes. United States: $z = .98$, p-value $= .1634$; no. Britain: $z = 1.00$, p-value $= .1587$; no
13.110 $z = 2.04$, p-value $= .0207$; yes
13.112 $z = -1.25$, p-value $= .2112$; no
13.114 $z = 4.61$, p-value $= 0$; yes
13.116 $z = 1.45$, p-value $= .1478$; no
13.118 $z = 5.13$, p-value $= 0$; yes
13.120 $z = .40$, p-value $= .6894$; no
13.122 2002: $z = 2.40$, p-value $= .0164$; yes.
2004: $z = .29$, p-value $= .7716$; no.
2006: $z = 2.24$, p-value $= .0250$.
2008: $z = .99$, p-value $= .3202$
13.124 $z = -3.69$, p-value $= .0002$; yes
13.126 **a.** $z = 2.49$, p-value $= .0065$; yes
b. $z = .89$, p-value $= .1859$; no
13.128 $t = .88$, p-value $= .1931$; no
13.130 $t = -6.09$, p-value $= 0$; yes
13.132 $z = -2.30$, p-value $= .0106$; yes
13.134 **a.** $t = -1.06$, p-value $= .2980$; no
b. $t = -2.87$, p-value $= .0040$; yes
13.136 $z = 2.26$, p-value $= .0119$; yes
13.138 $z = -4.28$, p-value $= 0$; yes
13.140 $t = -4.53$, p-value $= 0$; yes
13.142 **a.** $t = 4.14$, p-value $= .0001$; yes
b. LCL $= 1.84$, UCL $= 5.36$
13.144 $t = -2.40$, p-value $= .0100$; yes
13.146 $z = 1.20$, p-value $= .1141$; no
13.148 $t = 14.07$, p-value $= 0$; yes
13.150 $t = -2.40$, p-value $= .0092$; yes
13.152 F-Test: $F = 1.43$, p-value $= 0$.
t-Test: $t = .71$, p-value $= .4763$
13.154 $t = 2.85$, p-value $= .0025$; yes
13.156 $z = -3.54$, p-value $= .0002$; yes
13.158 $t = -2.13$, p-value $= .0171$; yes
13.160 $z = -.45$, p-value $= .6512$; no

Chapter 14

14.4 $F = 4.82$, p-value $= .0377$; yes
14.6 $F = 3.91$, p-value $= .0493$; yes
14.8 $F = .81$, p-value $= .5224$; no
14.10 **a.** $F = 2.94$, p-value $= .0363$; evidence of differences
14.12 $F = 3.32$, p-value $= .0129$; yes
14.14 $F = 1.17$, p-value $= .3162$; no
14.16 $F = 1.33$, p-value $= .2675$; no
14.18 **a.** $F = 25.60$, p-value $= 0$; yes
b. $F = 7.37$, p-value $= .0001$; yes
c. $F = 1.82$, p-value $= .1428$; no
14.20 $F = .26$, p-value $= .7730$; no
14.22 $F = 31.86$, p-value $= 0$; yes
14.24 $F = .33$, p-value $= .8005$; no
14.26 $F = .50$, p-value $= .6852$; no
14.28 $F = 11.59$, p-value $= 0$; yes
14.30 $F = 17.10$, p-value $= 0$; yes
14.32 $F = 37.47$, p-value $= 0$; yes
14.34 **a.** μ_1 and μ_2, μ_1 and μ_4, μ_1 and μ_5, μ_2 and μ_4, μ_3 and μ_4, μ_3 and μ_5, and μ_4 and μ_5 differ.
b. μ_1 and μ_5, μ_2 and μ_4, μ_3 and μ_4, and μ_4 and μ_5 differ.
c. μ_1 and μ_2, μ_1 and μ_5, μ_2 and μ_4, μ_3 and μ_4, and μ_4 and μ_5 differ.
14.36 **a.** BA and BBA differ.
b. BA and BBA differ.
14.38 **a.** The means for Forms 1 and 4 differ.
b. No means differ.
14.40 **a.** Lacquers 2 and 3 differ.
b. Lacquers 2 and 3 differ.
14.42 No fertilizers differ.
14.44 Blacks differ from Whites and others.
14.46 Married and separated, married and never married, and divorced and single differ.
14.48 Democrats and Republicans and Republicans and Independents differ.
14.50 All three groups differ.
14.52 **a.** $F = 16.50$, p-value $= 0$; treatment means differ
b. $F = 4.00$, p-value $= .0005$; block means differ
14.54 **a.** $F = 7.00$, p-value $= .0078$; treatment means differ
b. $F = 10.50$, p-value $= .0016$; treatment means differ
c. $F = 21.00$, p-value $= .0001$; treatment means differ
d. F-statistic increases and p-value decreases.
14.56 **a.** SS(Total) 14.9, SST $= 8.9$, SSB $= 4.2$, SSE $= 1.8$
b. SS(Total) 14.9, SST $= 8.9$, SSE $= 6.0$
14.58 $F = 1.65$, p-value $= .2296$; no
14.60 **a.** $F = 123.36$, p-value $= 0$; yes
b. $F = 323.16$, p-value $= 0$; yes
14.62 **a.** $F = 21.16$, p-value $= 0$; yes
b. $F = 66.02$, p-value $= 0$; randomized block design is best
14.64 **a.** $F = 10.72$, p-value $= 0$; yes
b. $F = 6.36$, p-value $= 0$; yes
14.66 $F = 44.74$, p-value $= 0$; yes
14.68 **b.** $F = 8.23$; Treatment means differ
c. $F = 9.53$; evidence that factors A and B interact
14.70 **a.** $F = .31$, p-value $= .5943$; no evidence that factors A and B interact.
b. $F = 1.23$, p-value $= .2995$; no evidence of differences between the levels of factor A.
c. $F = 13.00$, p-value $= .0069$; evidence of differences between the levels of factor B.
14.72 $F = .21$, p-value $= .8915$; no evidence that educational level and gender interact. $F = 4.49$, p-value $= .0060$; evidence of differences between educational levels. $F = 15.00$, p-value $= .0002$; evidence of a difference between men and women.
14.74 **d.** $F = 4.11$, p-value $= .0190$; yes
e. $F = 1.04$, p-value $= .4030$; no
f. $F = 2.56$, p-value $= .0586$; no
14.76 **d.** $F = 7.27$, p-value $= .0007$; evidence that the schedules and drug mixtures interact.
14.78 Both machines and alloys are sources of variation.
14.80 The only source of variation is skill level.
14.82 **a.** $F = 7.67$, p-value $= .0001$; yes
14.84 $F = 13.79$, p-value $= 0$; use the typeface that was read the most quickly.
14.86 $F = 7.72$, p-value $= 0.0070$; yes
14.88 **a.** $F = 136.58$, p-value $= 0$; yes
b. All three means differ from one another. Pure method is best.
14.90 $F = 14.47$, p-value $= 0$; yes
14.92 $F = 13.84$, p-value $= 0$; yes
14.94 $F = 1.62$, p-value $= .2022$; no
14.96 $F = 45.49$, p-value $= 0$; yes
14.98 $F = 211.61$, p-value $= 0$; yes

Chapter 15

15.2 $\chi^2 = 2.26$, p-value $= .6868$; no evidence that at least one p_i is not equal to its specified value.
15.6 $\chi^2 = 9.96$, p-value $= .0189$; evidence that at least one p_i is not equal to its specified value.
15.8 $\chi^2 = 6.85$, p-value $= .0769$; not enough evidence that at least one p_i is not equal to its specified value.
15.10 $\chi^2 = 14.07$, p-value $= .0071$; yes
15.12 $\chi^2 = 33.85$, p-value $= 0$; yes
15.14 $\chi^2 = 6.35$, p-value $= .0419$; yes
15.16 $\chi^2 = 5.70$, p-value $= .1272$; no
15.18 $\chi^2 = 4.97$, p-value $= .0833$; no
15.20 $\chi^2 = 46.36$, p-value $= 0$; yes
15.22 $\chi^2 = 19.10$, p-value $= 0$; yes
15.24 $\chi^2 = 4.77$, p-value $= .0289$; yes
15.26 $\chi^2 = 4.41$, p-value $= .1110$; no
15.28 $\chi^2 = 2.36$, p-value $= .3087$; no
15.30 $\chi^2 = 19.71$, p-value $= .0001$; yes
15.32 **a.** $\chi^2 = .64$, p-value $= .4225$; no
15.34 $\chi^2 = 41.77$, p-value $= 0$; yes
15.36 $\chi^2 = 43.36$, p-value $= 0$; yes
15.38 $\chi^2 = 20.89$, p-value $= .0019$; yes
15.40 $\chi^2 = 36.57$, p-value $= .0003$; yes
15.42 $\chi^2 = 110.3$, p-value $= 0$; yes
15.44 $\chi^2 = 5.89$, p-value $= .0525$; no
15.46 $\chi^2 = 35.21$, p-value $= 0$; yes
15.48 $\chi^2 = 9.87$, p-value $= .0017$; yes
15.50 $\chi^2 = 506.76$, p-value $= 0$; yes
15.52 Phone: $\chi^2 = .2351$, p-value $= .8891$; no.
Not on phone: $\chi^2 = 3.18$, p-value $= .2044$; no
15.54 $\chi^2 = 3.20$, p-value $= .2019$; no

15.56 $\chi^2 = 5.41$, p-value = .2465; no
15.58 $\chi^2 = 20.38$, p-value = .0004; yes
15.60 $\chi^2 = 86.62$, p-value = 0; yes
15.62 $\chi^2 = 4.13$, p-value = .5310; no
15.64 $\chi^2 = 9.73$, p-value = .0452; yes
15.66 $\chi^2 = 4.57$, p-value = .1016; no
15.68 **a.** $\chi^2 = .648$, p-value = .4207; no
b. $\chi^2 = 7.72$, p-value = .0521; no
c. $\chi^2 = 23.11$, p-value = 0; yes
15.70 $\chi^2 = 4.51$, p-value = .3411; no

Chapter 16

16.2 $\hat{y} = 9.107 + .0582x$
16.4 **b.** $\hat{y} = -24.72 + .9675x$
16.6 **b.** $\hat{y} = 3.635 + .2675x$
16.8 $\hat{y} = 7.460 + 0.899x$
16.10 $\hat{y} = 7.286 + .1898x$
16.12 $\hat{y} = 4{,}040 + 44.97x$
16.14 $\hat{y} = 458.4 + 64.05x$
16.16 $\hat{y} = 20.64 - .3039x$
16.18 $\hat{y} = 89.81 + .0514x$
16.22 $t = 10.09$, p-value = 0; evidence of linear relationship
16.24 **a.** 1.347
b. $t = 3.93$, p-value = .0028; yes
c. LCL = .0252, UCL = .0912
d. .6067
16.26 $t = 6.55$, p-value = 0; yes
16.28 **a.** 5.888 **b.** .2892
c. $t = 4.86$, p-value = 0; yes
d. LCL = .1756, UCL = .3594
16.30 $t = 2.17$, p-value = .0305; yes
16.32 $t = 7.50$, p-value = 0; yes
16.34 **a.** 3,287 **b.** $t = 2.24$, p-value = .0309 **c.** .1167
16.36 $s_\varepsilon = 191.1$; $R^2 = .3500$; $t = 10.39$, p-value = 0
16.38 $t = -3.39$, p-value = .0021; yes
16.40 **a.** .0331
b. $t = 1.21$, p-value = .2319; no
16.42 $t = 4.86$, p-value = 0; yes
16.44 $t = 7.49$, p-value = 0; yes
16.46 $\hat{y} = -29{,}984 + 4905x$; $t = 15.37$, p-value = 0.
16.48 $t = 6.58$, p-value = 0; yes
16.50 $t = 7.80$, p-value = 0; yes
16.52 $t = -8.95$, p-value = 0; yes
16.56 141.8, 181.8
16.58 13,516, 27,260
16.60 **a.** 186.8, 267.2 **b.** 200.5, 215.5
16.62 24.01, 31.43
16.64 **a.** 27.62, 72.06 **b.** 29.66, 37.92
16.66 23.30, 34.10
16.68 190.4, 313.4
16.70 **a.** 60.00, 62.86 **b.** 41.51, 74.09
16.72 92.01, 95.83
16.74 16,466, 21,657
16.76 0 (increased from −83.98), 204.8
16.78 3.15, 3.40
16.80 0(increased from −.15), 8.38
16.100 **a.** $\hat{y} = 115.24 + 2.47x$ **c.** .5659
d. $t = 4.84$, p-value = .0001; yes
e. Lower prediction limit = 318.1, upper prediction limit = 505.2
16.102 **a.** $t = 21.78$, p-value = 0; yes
b. $t = 11.76$, p-value = 0; yes
16.104 $t = 3.01$, p-value = .0042; yes
16.106 $t = 1.67$, p-value = .0522; no
16.108 $r = t = -9.88$, p-value = 0; yes

Chapter 17

17.2 **a.** $\hat{y} = 13.01 + .194x_1 + 1.11x_2$
b. 3.75 **c.** .7629
d. $F = 43.43$, p-value = 0; evidence that the model is valid.
f. $t = .97$, p-value = .3417; no
g. $t = 9.12$, p-value = 0; yes
h. 23, 39 **i.** 49, 65
17.4 **c.** $s_\varepsilon = 6.99$, $R^2 = .3511$; model is not very good.
d. $F = 22.01$, p-value = 0; evidence that the model is valid.
e. Minor league home runs: $t = 7.64$, p-value = 0; Age: $t = .26$, p-value = .7961
Years professional: $t = 1.75$, p-value = .0819
Only the number of minor league home runs is linearly related to the number of major league home runs.
f. 9.86 (rounded to 10), 38.76 (rounded to 39)
g. 14.66, 24.47
17.6 **b.** .2882
c. .$F = 12.96$, p-value = 0; evidence that the model is valid.
d. High school GPA: $t = 6.06$, p-value = 0; SAT: $t = .94$, p-value = .3485
Activities: $t = .72$, p-value = .4720
e. 4.45, 12.00 (actual value = 12.65; 12 is the maximum)
f. 6.90, 8.22
17.8 **b.** $F = 29.80$, p-value = 0; evidence to conclude that the model is valid.
d. House size : $t = 3.21$, p-value = .0006; Number of children: $t = 7.84$ p-value = 0
Number of adults at home: $t = 4.48$, p-value = 0
17.10 **b.** $F = 67.97$, p-value = 0; evidence that the model is valid.
d. 65.54, 77.31
e. 68.75, 74.66
17.12 **a.** $\hat{y} = -28.43 + .604x_1 + .374x_2$
b. $s_\varepsilon = 7.07$ and $R^2 = .8072$; the model fits well.
d. 35.16, 66.24 **e.** 44.43, 56.96
17.14 **b.** $F = 24.48$, p-value = 0; yes
c.

Variable	t	p-value
UnderGPA	.52	.6017
GMAT	8.16	0
Work	3.00	.0036

17.16 **a.** 9.09 + .219 PAEDUC + .197 MAEDUC
b. $F = 234.9$, p-value = 0
c. PAEDUC: $t = 9.73$, p-value = 0
MAEDUC: $t = 7.69$, p-value = 0
17.18 **a.** $F = 9.09$, p-value = 0
b.

Variable	t	p-value
AGE	2.34	.0194
EDUC	−3.11	.0019
HRS	−2.35	.0189
PRESTG80	−3.47	.0005
CHILDS	−.84	.4021
EARNRS	−.98	.3299

c. $R^2 = 0659$
17.20 **a.** $F = 35.06$, p-value = 0
b.

Variable	t	p-value
AGE	.40	.6864
EDUC	7.89	0
HRS	7.10	0
CHILDS	1.61	.1084
AGEKDBRN	4.90	0
YEARSJOB	5.85	0
MOREDAYS	1.36	.1754
NUMORG	1.37	.1713

17.22 **a.** $\hat{y} = 6.36 + .135$ DAYS1 + .036 DAYS2 + .060 DAYS3 + .107 DAYS4 + .142 DAYS5 + .134 DAYS6
b. $F = 11.72$, p-value = 0
c.

Variable	t	p-value
DAYS1	3.33	.0009
DAYS2	.81	.4183
DAYS3	1.41	.1582
DAYS4	3.00	.0027
DAYS5	3.05	.0024
DAYS6	3.71	.0002

17.40 $d_L = 1.16$, $d_U = 1.59$; $4 - d_L = 2.84$, $4 - d_U = 2.41$; evidence of negative first-order autocorrelation.
17.42 $d_L = 1.46$, $d_U = 1.63$. There is evidence of positive first-order autocorrelation.
17.44 $4 - d_U = 4 - 1.73 = 2.27$, $4 - d_L = 4 - 1.19 = 2.81$. There is no evidence of negative first-order autocorrelation.
17.46 **a.** The regression equation is $\hat{y} = 2260 + .423x$
c. $d = .7859$. There is evidence of first-order autocorrelation.
17.48 $d = 2.2003$; $d_L = 1.30$, $d_U = 1.46$, $4 - d_U = 2.70$, $4 - d_L = 2.54$. There is no evidence of first-order autocorrelation.
17.50 **a.** $\hat{y} = 164.01 + .140x_1 + .0313x_2$
b. $t = 1.72$, p-value = .0974; no
c. $t = 4.64$, p-value = .0001; yes
d. $s_\varepsilon = 63.08$ and $R^2 = .4752$; the model fits moderately well.
f. 69.2, 349.3
17.52 **a.** $\hat{y} = 29.60 - .309x_1 - 1.11x_2$
b. $R_2 = .6123$; the model fits moderately well.
c. $F = 21.32$, p-value = 0; evidence to conclude that the model is valid.
d. Vacancy rate: $t = -4.58$, p-value = .0001; yes
Unemployment rate: $t = -4.73$, p-value = .0001; yes
e. The error is approximately normally distributed with a constant variance.
f. d = 2.0687; no evidence of first-order autocorrelation.
g. $14.18, $23.27

INDEX

Acute otitis media (ear infections), 588
Addition rule, 193–195
Advertising applications, 353
Alternative hypothesis (research hypothesis), 361–364, 374
  determining, 391–392
American National Election Survey (ANES), 7
Analysis of variance (ANOVA)
  for complete factorial experiments, 574
  experimental designs for, 553–554
  for multiple comparisons, 543–551
  for multiple regression analysis, 705
  one-way, 526–539
  randomized block (two-way), 554–561
  two-factor, 563–575
Analysis of variance (ANOVA) tables, 531–532
  for randomized block ANOVA, 556
Applets, 8
  analysis of regression deviations, 657
  for χ^2 (chi-square) distribution, 300
  for confidence interval estimators of a mean, 345
  distribution of difference between means, 329
  fair dice, 314
  for *F* distribution, 304
  for fitting regression lines, 639
  loaded dice, 315
  normal approximation to binomial probabilities, 324
  normal distribution areas, 277
  normal distribution parameters, 271–272
  for plots of two-way ANOVA effects, 575
  for power of *z*-tests, 391
  sampling, 173
  for scatter diagrams and correlation, 131–132
  skewed dice, 315
  for Student *t* distribution, 296–297
Arithmetic means. *See* Means
Asset allocation, 236–241
Auditing taxes, 175, 202
Autocorrelated (serially correlated) error variables, 675
Autocorrelation, first-order, 716–719, 722–724
Averages. *See* Means

Balanced factorial design, 567
Bar charts, 21–24
  deception in, 87
Barmonic means, 549
Baseball applications
  bunting decisions, 260
  cost of one more win, 97, 140–141
  of numerical descriptive techniques, 144–147
  of probability, 213–214
Bayes's Law, 199–208, 210
Bell-shaped histograms, 51
Bernoulli process, 243
β (beta)
  operating characteristic curve of, 390–391
  for probability of Type II error, 361, 385–387
Beta (β) coefficient, 148–149
Between-treatments variation (SST), for one-way ANOVA, 528–529
Bias, selection, 174
Bimodal histograms, 51
Binomial distributions, 242–248
  normal approximation of, 321–323
  Poisson distributions and, 250
Binomial experiment, 242–243
  multinomial experiment and, 597
Binomial random variables, 243, 244
Binomial table, 246–248
Bivariate distributions, 228–233
  normal, 649
Bivariate techniques, 32
Blocks (randomized block design), 554
  criteria for, 559–560
Bonferroni adjustment to LSD method, 547–548, 551
Box plots, 120–124
Breakeven analysis, 132–133

Calculations, for types of data, 15–16
Cause-and-effect relationships, 659–660
Census, 161–162
  sampling and, 171–172
Central limit theorem, 312, 339
Central location, measures of, 2, 98–106
  arithmetic means, 98–100
  comparisons of, 103–104
  medians, 100–101
  modes, 101–103
  for ordinal and nominal data, 104
Chebysheff's Theorem, 114–115
χ^2 chi-square density function, 297
χ^2 chi-squared goodness-of-fit test, 598–601
  for nominal data, 616
χ^2 chi-square distribution, 297–300
  Excel and Minitab for, 416–419
  table for, 299
χ^2 chi-squared statistic, for estimator of population variance, 414–419
χ^2 chi-squared tests
  of contingency tables, 604–612
  for goodness-of-fit, 61, 597–601
  for normality, 617–620
Classes, in histograms, 46, 48–50
Climate change. *See* Global warming
Cluster sampling, 171–172
Coefficient of correlation, 128–129
  for bivariate distributions, 231
  compared with other measures of linear relationship, 130–132
  testing, 660–662
Coefficient of determination, 139
  in linear regression analysis, 655–659
  in multiple regression analysis, 698–699
Coefficient of variation, 115
Coefficients, estimating
  in linear regression analysis, 637–644
  in multiple regression analysis, 694–706
Collinearity (multicollinearity; intercorrelation), 714–715
Complement rule, 191
Complete factorial experiments, 566–567
  ANOVA for, 574
Completely randomized design, 534
Conditional probabilities, 183–185
  Bayes's Law and, 199–202
  multiplication rule for, 191–192
Confidence interval estimators, 340–345
  for difference between two population means, 451–452
  for difference between two proportions, 498
  Excel and Minitab for, 343–345
  hypothesis testing and, 380
  ignoring for large populations, 407
  interpretation of, 345–348
  for linear regression model, 654
  for population variance, 414
  for ratio of two variances, 490
  for regression equation, 667
  for standard error in proportions, 422
  for *t*-statistic, 400
  width of interval for, 348–349
  Wilson estimate used for, 430
Confidence levels, 4–5, 340
Consistency, 338
Consumer Price Index (CPI), 68
Contingency tables, 607
  χ^2 (chi-squared) tests of, 604–612
Continuity correction factor, 323
  omitting, 324–325
Continuous random variables, 218, 264
Correction factor for continuity, 323
  omitting, 324–325
Correlation
  cause-and-effect relationships and, 659–660
  coefficient of, 128–129
  interpreting, 141
Correlation analysis, 634
Costs, fixed and variable, 133–136
Covariance, 127–128, 230–231
  compared with other measures of linear relationship, 130–132
Credit scorecard, 63
Critical Path Method (CPM), 234–236, 287
Cross-classification tables (cross-tabulation tables), 32–34
Cross-sectional data, 64
Cumulative probabilities, 245–246
Cumulative relative frequency distributions, 59

Data
  collection methods for, 162–165
  definition of, 13–14
  formats for, 38
  guidelines for exploring, 153–154
  hierarchy of, 16–17
  missing, 426
  nonsampling errors in collection of, 173–174
  observational and experimental, 472–474, 484
  sampling for, 165–166
  types of, 13–17, 394–395

Data formats
for χ^2 (chi-squared) test of contingency tables, 610
for difference between two population means, 465–466
Deception, graphical, 84–88
Degrees of freedom, 407
for χ^2 (chi-square) distribution, 297, 414
for *F* distribution, 301, 303
for matched pairs experiments, 477
for ratio of two variances, 490
for Student *t* distribution, 292–294
for *t*-statistic, 400
Density functions, 264–269
χ^2 (chi-square), 297
F, 301
normal, 270
Student *t* density function, 292
Dependent variables, 634
in multiple regression analysis, 693
Descriptive statistics, 2–3
describing relationship between two nominal variables, 32–38
graphical, 12–13
for interval data, 44–61
for nominal data, 18–27
for relationship between two interval variables, 74–80
for time-series data, 64–68
types of data for, 13–17
Deterministic models, 635–636
Direct observation, 162–163
Discrete bivariate distributions, 229
Discrete probability distributions, 219
continuous distributions to approximate, 269
Discrete random variables, 218, 219
Distributions
binomial distributions, 242–248
bivariate distributions, 228–233
χ^2 (chi-square) distribution, 297–300
exponential distribution, 287–290
F distribution, 301–304
normal distribution, 270–284
Poisson distributions, 250–254
probability distributions, 217–224
Student *t* distribution, 291–296
Diversification, 236–241
Double-blind experiments, 508
Down syndrome, 214–215
Durbin-Watson test, 716–719
Excel and Minitab for, 721–724

Ear infections (acute otitis media), 588
Elections. *See* Voting and elections
Equal-variances test statistic, 451, 453
Errors
calculating probability of Type II errors, 385–392
of estimation, 354
false-positive and false-negative test results, 203–207
multiple tests increasing chance of Type I errors, 535, 547
in polls, 166
in sampling, 172–174
Type I and Type II, 361–362
See also Type I errors; Type II errors
Error variables (ε epsilon), 636
heteroscedasticity of, 674
in multiple regression analysis, 694
nonindependence of, 675
required conditions for, 647–649
Estimates, standard errors of. *See* Standard error of estimate
Estimation
confidence interval estimators, 340–341
errors of, 354
point and interval estimators, 336–339
pooled variance estimators for, 451
of standard error in proportions, 422
Wilson estimators, 430–431
Events
assigning probabilities to, 176–179
independence of, 185
intersection of, 181
union of, 186
Excel, 7–8
for analysis of variance, 663
Analysis ToolPak in, 341
for ANOVA for multiple comparisons, 545, 549
for arithmetic means, 100
for bar and pie charts, 22
for binomial distributions, 247
for box plots, 121
for χ^2 (chi-squared) goodness-of-fit test, 600
for χ^2 (chi-square) distribution, 300, 416, 418
for χ^2 (chi-squared) test for normality, 619–620
for χ^2 (chi-squared) test of contingency tables, 609
for coefficient of correlation, 662
for coefficient of determination, 139, 658
to compute coefficients in multiple regression analysis, 696
for confidence interval estimators, 343–344
for cross-classification tables, 33–34
for difference between two population means, 454, 456–458, 460, 461, 463
for difference between two proportions, 500, 502, 504, 505
for Durbin-Watson test, 721, 722, 724
for exponential distribution, 289
for *F* distribution, 304
for frequency distributions, 20
for geometric means, 105
for histograms, 47
for interactions, 573
for least squares method, 135, 140
for linear regression model, 654
for line charts, 66–67
for market segmentation problem, 438
for matched pairs experiments, 477, 480, 482
for measures of central location, 102–103
for measuring strength of linear relationships, 137–138
for medians, 101
for medical screening, 205
missing data problem in, 426
for modes, 102
for normal distribution, 282
for observational data, 473
for ogives, 60
for one-way analysis of variance, 533, 537, 538
for Poisson distributions, 254
for portfolio management, 239–240
for power of statistical tests, 389
for prediction intervals in linear multiple regression analysis, 669
for prediction intervals in multiple regression analysis, 706
for *p*-values, 372
for quartiles, 119
for randomized block ANOVA, 558
random samples generated by, 167–168
for ratio of two variances, 491–493
for regression lines, 642–643
for residuals in linear regression analysis, 672
for scatter diagrams, 75
for standard deviation, 112
for standard error of estimate, 651, 698
for stem-and-leaf displays, 58
for Student *t* distribution, 296
for testing population means, 378
for testing validity of multiple regression model, 700
for time-series analysis, 719
for *t*-statistic, 402, 405
for *t*-tests, 408
for two-factor ANOVA, 565, 570–571
for two-way ANOVA, 582
for variance, 111
for *z* scores for population proportions, 424
Exit polls, 4, 423
Expected values
Law of, 224
for population means, 222
Experimental data, 474
error variables for, 648–649
observational data and, 484
Experimental units, 528
Experiments, 163
analysis of variance and design of, 553–554
completely randomized design for, 534
factorial, 563
for inference about difference between two means, with matched pairs, 475–486
matched pairs compared with independent samples in, 483–484
pharmaceutical and medical experiments, 508–509
random, 176–177
Taguchi methods and design of, 582
Exponential distribution, 287–290
Exponential probability density function, 287
Exponential random variables, 288

Factorial experiments, 563
complete, 566–567
complete, ANOVA for, 574
sum of squares for factors and interactions in, 567–570
False-negative test results, 203
False-positive test results, 203–207
F density function, 301
F distribution, 301–304
for difference between two population means, 454
table for, 302
Financial applications
measuring risk for, 277
mutual funds, 181–187
negative return on investment, 277–282
on numerical descriptive techniques, 147–149
portfolio diversification and asset allocation, 236–241
return on investment, 52–54
stock and bond valuation, 51–52
Finite population correction factor, 313
Firm-specific (nonsystematic) risk, 149
First-order autocorrelation, 716–719
First-order linear model (simple linear regression model), 636
assessing models for, 650–664
diagnosing violations in, 671–678
error variables in, 647–649
estimating coefficients for, 637–644
estimators and sampling distributions for, 653
F-test and *t*-tests used in, 705
model for, 635–644

regression equation for, 666–670
testing slope in, 652–653
Fisher's least significant difference (LSD) method, 546–547, 551
Fixed and variable costs, 133
estimating, 134–136
Fixed-effects analysis of variance, 554
Frequency distributions, 18, 20
F-statistic, *t*-statistic compared and, 536–537
F-test
for difference between two population means, 459–462
for multiple regression analysis, 705
for one-way ANOVA, 530–531, 534
for randomized block ANOVA, 559
for ratio of two variances, 489–493
for two-factor ANOVA, 569

General Social Survey (GSS), 7
Geometric means, 104–105
Global warming, 95–96, 157
public opinion on, 510
Goodness-of-fit, chi-squared (χ^2) tests for, 597–601
Gosset, William S., 291, 400
Graphical descriptive techniques, 12–13
bar and pie charts, 21–25
deception in, 84–88
excellence in, 82–84
histograms, 46–57
for interval data, 44–61
line charts, 65–67
numerical descriptive techniques compared with, 150–152
ogives, 59–61
probability trees, 195–197
for relationship between two nominal variables, 35–36, 605
scatter diagrams, 74–80
stem-and-leaf displays, 57–59
for time-series data, 64–68
Graphical excellence, 82–84
Grouped data, approximating mean and variance for, 115

Heteroscedasticity, 674
Histograms, 44, 46–57
Chebysheff's Theorem for, 114–115
Holmes, Oliver Wendell, 362
Homoscedasticity, 674
Human resources management applications
retention of workers, 645–646
severance pay, 708
testing job applicants, 647
Hypothesis testing, 361–364
calculating probability of Type II errors in, 385–392
determining alternative hypothesis for null hypothesis, 391–392
testing population means with known standard deviation, 365–381

Independence, of events, 185
multiplication rule for, 192
Independent samples, 553–554
Independent variables, 634
multicollinearity among, 714–715
in multiple regression analysis, 693, 695–696
Inferences, 336
about difference between two means, using independent samples, 449–467
about difference between two means, using matched pairs, 475–486
about difference between two proportions, 495–506
about population proportions, 421–431
about populations, with standard deviation unknown, 399–408
about population variance, 413–419
about ratio of two variances, 489–493
definition of, 4–5
sampling distribution used for, 317–319, 330–331
for Student *t* distribution used for, 293
Inferential statistics, 34
Influential observations, 677, 714
Information
types of, 13–17
See also Data
Interactions (between variables), 565, 573–574
sum of squares for factors and, 567–570
Intercorrelation (multicollinearity; collinearity), 714–715
Interrquartile range, 120–121
Intersections, of events, 181
Interval data, 14, 395
analysis of variance on, 527
calculations for, 15
graphical techniques for, 44–61
relationship between two interval variables, 74–80
Interval estimators, 336–339
for population variance, 413–414
Intervals
prediction intervals, 666, 670
width of, for confidence interval estimators, 348–349
Interval variables
in multiple regression analysis, 695–696
relationship between two, 74–80
Interviews, 163–164
Inventory management, 283, 342
Investments
comparing returns on, 150–151
management of, 51–52
measuring risk for, 277
mutual funds, 181–187, 727–728
negative return on, 277–282
portfolio diversification and asset allocation for, 236–241
returns on, 52–54
stock market indexes for, 148

Joint probabilities, 181
selecting correct methods for, 209–210

Laws
Bayes's Law, 199–208, 210
of expected value, 224, 232
of variance, 224, 232
Lead time, 283
Least significant difference (LSD) method
Bonferroni adjustment to, 547–548
Fisher's, 546–547
Tukey's, 548–549
Least squares line coefficients, 637–638
Least squares method, 77, 132–136, 637
Likelihood probabilities, 200
Linearity, in for scatter diagrams, 76–77
Linear programming, 241
Linear relationships, 126–141
coefficient of correlation for, 128–129
coefficient of determination for, 139–141
comparisons among, 130–132
covariance for, 127–128
least squares method for, 132
measuring strength of, 136–139
in scatter diagrams, 76–78
Line charts, 65–67
deception in, 84–88
Logistic regression, 63
Lower confidence limit (LCL), 340

Macroeconomics, 23
Marginal probabilities, 183
Marketing applications
in advertising, 353
market segmentation, 435–438, 511, 517–518, 542, 603, 624–625
test marketing, 499–504, 542
Market models, 148–149
Market-related (systematic) risk, 149
Market segmentation, 435–438, 511, 517–518, 542, 603, 624–625
Markowitz, Harry, 236
Mass marketing, 435–436
Matched pairs, 553–554
compared with independent samples, 483–484
for inference about difference between two population means, 475–486
Mean of population of differences, 479
Means, 2
approximating, for grouped data, 115
arithmetic, 98–100
of binomial distributions, 248
compared with medians, 103–104
expected values for, 222
geometric, 104–105
for normal distribution, 271
sampling distribution of, 308–319
sampling distributions of difference between two means, 327–329
See also Population means; Sample means
Mean square for treatments (mean squares; MSE), 530
for randomized block experiments, 556
Measurements, descriptive, 2
Medians, 100–101
compared with means, 103–104
used in estimate of population mean, 349–350
Medical applications
comparing treatments for child-hood ear infections, 588
estimating number of Alzheimer's cases, 447
estimating total medical costs, 446
pharmaceutical and medical experiments, 508–509
of probability, 203–207, 214–215
Microsoft Excel. *See* Excel
Minitab, 7–8
for analysis of variance, 663, 673
for ANOVA for multiple comparisons, 545, 550–551
for arithmetic means, 100
for bar and pie charts, 22–23
for binomial distributions, 248
for box plots, 122
for χ^2 (chi-squared) goodness-of-fit test, 601
for χ^2 (chi-square) distribution, 300, 417–419
for χ^2 (chi-squared) test of contingency tables, 609–610
for coefficient of correlation, 662
for coefficient of determination, 139, 658
to compute coefficients in multiple regression analysis, 697
for confidence interval estimators, 344–345
for cross-classification tables, 34

for difference between two population means, 455–458, 461–463
for difference between two proportions, 501, 502, 504, 505
for Durbin–Watson test, 721, 722, 724
for exponential distribution, 290
for *F* distribution, 304
for frequency distributions, 20
for histograms, 48
for interactions, 573
for least squares method, 136, 140
for linear regression model, 655
for line charts, 67
for market segmentation problem, 438
for matched pairs experiments, 477, 481, 482
for measures of central location, 102–103
for measuring strength of linear relationships, 138
for medians, 101
missing data problem in, 426
for modes, 102
for normal distribution, 282
for ogives, 61
for one-way analysis of variance, 533–534, 538
for Poisson distributions, 254
for power of statistical tests, 389
for prediction intervals in linear multiple regression analysis, 669
for prediction intervals in multiple regression analysis, 706
for *p*-values, 373
for quartiles, 120
for randomized block ANOVA, 558
random samples generated by, 168
for ratio of two variances, 492
for regression lines, 643
for scatter diagrams, 76
for standard deviation, 113
for standard error of estimate, 652, 698
for stem-and-leaf displays, 58–59
for Student *t* distribution, 296
for testing population means, 379
for testing validity of multiple regression model, 700
for time-series analysis, 720
for *t*-statistic, 403, 405
for two-factor ANOVA, 565, 571
for two-way ANOVA, 582
for variance, 111
for *z* scores for population proportions, 425
Missing data, 426
Mitofsky, Warren, 423*n*
Modal classes, 50, 102
Models, 635–644
deterministic and probabilistic, 635–636
in linear regression, assessing, 650–664
in multiple regression, 693–694
in multiple regression, assessing, 694–706
Modern portfolio theory (MPT), 236
Modes, 101–103
in histograms, 50–51
Multicollinearity (collinearity; intercorrelation), 696, 714–715
Multifactor experimental design, 553
Multinomial experiment, 597–598
Multiple comparisons
ANOVA for, 543–551
Tukey's method for, 548–549
Multiple regression analysis
diagnosing violations in, 713–715
estimating coefficients and assessing models in, 694–706
models and required conditions for, 693–694
time-series data, 716–724
Multiple regression equation, 694–695
Multiplication rule, 191–192
Mutual funds, 181–187, 727–728
Mutually exclusive events, addition rule for, 194

Negative linear relationships, 77
Nominal data, 14, 18–27, 394
calculations for, 15–16
χ^2 (chi-squared) test of contingency table for, 604–612
describing relationship between two nominal variables, 32–38
inferences about difference between two population proportions, 495–506
inferences about population proportions, 421–431
measures of central location for, 104
measures of variability for, 115
tests on, 615–617
Nonindependence
of error variables, 675
of time series, 714
Nonnormal populations (nonnormality), 406
in linear regression analysis, 673–674
in multiple regression analysis, 713, 714
nonparametric statistics for, 465, 485
test of, 419
Nonparametric statistics
Spearman rank correlation coefficient, 664
Wilcoxon rank sum test, 465, 485
Nonresponse errors, 174
Nonsampling errors, 173–174
Nonsystematic (firm-specific) risk, 149
Normal density functions, 270
Normal distribution, 270–284
approximation of binomial distribution to, 321–323
bivariate, 649
Student *t* distribution as, 292
test of, 419
Normality, χ^2 (chi-squared) test for, 617–620
Normal random variables, 270
Null hypothesis, 361–364
calculating probability of Type II errors and, 385–392
determining alternative hypothesis for, 391–392
Numerical descriptive techniques
baseball applications of, 144–147
financial applications of, 147–149
graphical descriptive techniques compared with, 150–152
for measures of central location, 98–106
for measures of linear relationship, 126–141
for measures of relative standing and box plots, 117–125
for measures of variability, 108–115

Observation, 162–163
Observational data, 472–474
error variables for, 648–649
experimental data and, 484
influential observations, 677, 714
Observed frequencies, 599
Ogives, 59–61
One-sided confidence interval estimators, 380
One-tailed tests, 376–377, 379–380
for linear regression model, 655
One-way analysis of variance, 526–539
Operating characteristic (OC) curve, 390–391
Operations management applications
finding and reducing variation, 578–582
inventory management in, 283, 342
location analysis, 711–712
pharmaceutical and medical experiments, 508–509
Project Evaluation and Review Technique and Critical Path Method in, 234–236, 287
quality of production in, 415
waiting lines in, 255–256, 290
Ordinal data, 14–15, 394–395
calculations for, 16
describing, 27
measures of central location for, 104
measures of relative standing for, 124
measures of variability for, 115
Outliers, 121
in linear regression analysis, 676–677
in multiple regression analysis, 714

Parameters, 162
definition of, 4, 98
Paths (in operations management), 234–235
Pearson coefficient of correlation, 660
Percentiles, 117–119
definition of, 117
Personal interviews, 163–164
Pharmaceutical and medical experiments, 508–509
Pictograms, 87–88
Pie charts, 21–25
Point estimators, 336–339
Point prediction, 666
Poisson distributions, 250–254
Poisson experiment, 250
Poisson probability distributions, 251
Poisson random variables, 250, 251
Poisson table, 252–254
Polls
errors in, 166
exit polls, 8
Pooled proportion estimate, 497
Pooled variance estimators, 451
Population means
analysis of variance test of differences in, 526
estimating, with standard deviation known, 339–350
estimating, using sample median, 349–350
expected values for, 222
inferences about differences between two, using independent samples, 449–467
inferences about differences between two, using matched pairs, 475–486
testing, when population standard deviation is known, 365–381
Populations, 395
coefficient of correlation for, 128
covariance for, 127
definition of, 4, 13
inferences about, with standard deviation unknown, 399–408
inferences about population proportions, 421–431
large but finite, 407
nonnormal, 406
probability distributions and, 221–224

in sampling distribution of mean, 308–309
target and sampled, 166
variance for, 108
Populations standard deviations, 222
Population variance, 222
inferences about, 413–419
Portfolio diversification, 236–241
Positive linear relationships, 77
Posterior probabilities (revised probabilities), 200
Power of statistical tests, 388
Excel and Minitab for, 389
of *z*-tests, 391
Prediction intervals
in linear regression analysis, 666, 669
in multiple regression analysis, 705–706
Prior probabilities, 200
Probabilistic models, 635–636
Probability
assigning for events, 176–179
Bayes's Law for, 199–208
joint, marginal, and conditional, 180–187
in normal distribution, calculating, 272
rules of, 191–195
selecting correct methods for, 209–210
trees to represent, 195–197
Probability density functions, 264–269
exponential, 287
Probability distributions, 217–224
binomial, 244
definition of, 218
Poisson, 251
populations and, 221–224
Probability trees, 195–197
Process capability index, 579
Project Evaluation and Review Technique (PERT), 234–236, 287
Proportions
inferences about difference between two population proportions, 495–506
inferences about population proportions, 421–431
sampling distribution of, 321–326
Prostate cancer, 203–207
p-values, 368–369
definition of, 369
Excel and Minitab for, 371–373
interpreting, 369–371

Quadratic relationships, 653
Quartiles, 118–121
Questionnaires, design of, 164–165

Random-effects analysis of variance, 554
Random experiments, 176–177
Randomized block design, 554
Randomized block (two-way) analysis of variance, 554–561
Random sampling
cluster sampling, 171–172
simple, 167–169
stratified, 169–171
Random variables, 217–224
binomial, 243, 244
definition of, 218
exponential, 288
exponential probability density function for, 287
normal, 270
Poisson, 250, 251
standard normal random variables, 272
Range, 2, 108
interrquartile range, 120–121
Ratios, of two variances, 489–493
Rectangular probability distributions (uniform probability density functions), 266–269
Regression analysis, 634–635
diagnosing violations in, 671–678
equation for, 666–670
estimation of coefficients in, 637–644
fitting regression lines in, 639
models in, 635–644
multiple, 693–694
time-series data, 716–724
See also First-order linear model; Multiple regression analysis
Regression equation, 666–670, 705–706
Regression lines
applet for, 640
Excel and Minitab for, 642–643
Rejection region, 365–367
for χ^2 (chi-squared) test of contingency tables, 608
definition of, 366
one- and two-tailed tests, 376–377, 379–380
p-values and, 371
z scores for, 367–368
Relative efficiency, 338
Relative frequency approach, in assigning probabilities, 178
Relative frequency distributions, 18, 59
Relative standing, measures of, 117–125
Reorder points, 283–284
Repeated measures, 554
Replicates, 567
Research hypothesis (alternative hypothesis), 361–364, 374
determining, 391–392
Residual analysis, 672–673
Residuals
in linear regression analysis, 672–673
in sum of squares for error, 639
Response rates, to surveys, 163
Responses, 528
Response surfaces, 694
Response variable, 528
Return on investment, 52–54
investing to maximize, 239–240
negative, 277–282
ρ (rho), for coefficient of correlation, 128
Risks
investing to minimize, 239–240
market-related and firm-specific, 149
measuring, 277
Robustness of test statistics, 406
Rule of five, 601, 610

Safety stocks, 283
Sampled populations, 166
Sample means
as estimators, 336
as test statistics, 364
Samples
coefficient of correlation for, 128
covariance for, 127
definition of, 4, 13
exit polls, 4
independent, 553–554
matched pairs compared with independent samples, 483–484
missing data from, 426
size of, 171
variance for, 108–111
Sample size, 171, 353–356
barmonic mean of, 549
to estimate proportions, 428–430
increasing, 387–388
Sample space, 177
Sample variance, 407
Sampling, 165–172
errors in, 172–174
replacement in selection of, 192–194
sample size for, 353–356
simple random sampling for, 167–169
Sampling distributions
of difference between two means, 327–329
for differences between two population means, 450
inferences from, 330–331
for linear regression models, 653
of means, 308–319
of means of any population, 312–313
for one-way ANOVA, 531
of proportions, 321–326
of sample means, 310, 313
of sample proportion, 325–326
Sampling errors, 172–173
Scatter diagrams, 74–80
compared with other measures of linear relationship, 130–132
Screening tests
for Down syndrome, 214–215
for prostate cancer, 203–207
Selection, with and without replacement, 192–194
Selection bias, 174
Self-administered surveys, 164
Self-selected samples, 166
Serially correlated (autocorrelated) error variables, 675
σ^2 (sigma squared)
for population variance, inferences about, 413–414
for sample variance, 108–110
Significance levels, 4–5
p-values and, 371
for Type I errors, 361
Simple events, 178–179
Simple linear regression model. *See* First-order linear model
Simple random sampling, 167–169
cluster sampling, 171–172
definition of, 167
Single-factor experimental design, 553
Six sigma (tolerance goal), 579
Skewness, in histograms, 50
Slope, in linear regression analysis, 652–653
Smith, Adam, 96
Spearman rank correlation coefficient, 664
Spreadsheets, 7–8
See also Excel
Stacked data format, 465–466
Standard deviations, 112–114
Chebysheff's Theorem for, 114–115
estimating population mean, with standard deviation known, 339–350
for normal distribution, 271
populations standard deviations, 222
of residuals, 673
of sampling distribution, 312
testing population mean, when population standard deviation is known, 365–381
t-statistic estimator for, 400
Standard error of estimate
in linear regression analysis, 650–651
in multiple regression analysis, 697–698, 701
Standard errors
of difference between two means, 326
of estimates, 650–651
of mean, 312
of proportions, 325
Standardized test statistics, 367–368
Standard normal random variables, 272
Statistical inference, 308, 336
definition of, 4–5
sampling distribution used for, 317–319, 330–331
for Student *t* distribution used for, 293
Statisticians, 1–2*n*

Statistics
definition of, 1–2, 98
descriptive, 2–3
inferential, 3–5
of samples, 4
Stem-and-leaf displays, 57–59
Stocks and bonds
portfolio diversification and asset allocation for, 236–241
stock market indexes, 148
valuation of, 51–52
Stratified random sampling, 169–171
definition of, 169
Student *t* density function, 292
t-statistic and, 400
Student *t* distribution, 291–296, 407
for difference between two population means, 451
for nonnormal populations, 406
table for, 294–295
t-statistic and, 400
Subjective approach, in assigning probabilities, 178
Sum of squares
for blocks (SSB), 555, 560
for error (SSE), 639, 650
for error (within-treatments variation; SSE) for one-way ANOVA, 529–530
for factors and interactions, 567–570
for treatments (between-treatments variation; SST) for one-way ANOVA, 528–529
Surveys, 163–165
missing data from, 426
Symmetry, in histograms, 50
Systematic (market-related) risk, 149

Taguchi, Genichi, 580
Taguchi loss function, 580–582
Target populations, 166
Taxes, auditing, 175, 202
t distribution. *See* Student *t* distribution
Telephone interviews, 164
Testing, false positive and false negative results in, 203–207
Test marketing, 499–504, 542
Test statistic, 364
standardized, 367–368
t-statistic, 400
Time-series data, 64–68
diagnosing violations in, 716–724
Tolerance, in variation, 578
Taguchi loss function for, 580–581
Treatment means (in ANOVA), 526
t-statistic, 400–402, 407–408
Excel and Minitab for, 402–403, 405
F-statistic and, 536–537
variables in, 408
t-tests
analysis of variance compared with, 535–536
coefficient of correlation and, 660, 661
Excel for, 408
for matched pairs experiment, 476–478
for multiple regression analysis, 705
for observational data, 473
for two samples with equal variances, 466
for two samples with unequal variances, 467
Tufte, Edward, 83
Tukey, John, 57
Tukey's least significant difference (LSD) method, 548–549, 551
Two-factor analysis of variance, 563–575
Two-tailed tests, 376–377, 379–380
Two-way (randomized block) analysis of variance, 554–561
Type I errors, 361–362
determining alternative hypothesis for, 391–392
in multiple regression analysis, 696, 705
multiple tests increasing chance of, 535, 547
relationship between Type II errors and, 387
Type II errors, 361
calculating probability of, 385–392
determining alternative hypothesis for, 391–392

Unbiased estimators, 337
Unequal-variances test statistic, 452
estimating difference between two population means with, 462–463
Uniform probability density functions (rectangular probability distributions), 266–269
Unimodal histograms, 50–51
Union, of events, 186
addition rule for, 193–195
Univariate distributions, 228
Univariate techniques, 32
Unstacked data format, 465
Upper confidence limit (UCL), 340

Validity of model, testing, 699–701
Valuation of stocks and bonds, 51–52
Values, definition of, 13
Variability, measures of, 2, 108–115
coefficient of variation, 115
range, 108
standard deviations, 112–114
variance, 108–112
Variables, 395
definition of, 13
dependent and independent, 634
interactions between, 565, 573–574
nominal, describing relationship between two nominal variables, 32–38
in one-way analysis of variance, 528
random, 217–224
types of, 17
Variance, 108–112
approximating, for grouped data, 115
of binomial distributions, 248
estimating, 337
inferences about ratio of two variances, 489–493
interpretation of, 111–112
Law of, 224
in matched pairs experiments, 483
pooled variance estimators for, 451
population variance, 222
population variance, inferences about, 413–419
in sampling distribution of mean, 309
shortcut for, 110–111
Variation
coefficient of, 115
finding and reducing, 578–582
Voting and elections
electoral fraud in, 158
errors in polls for, 166
exit polls in, 4, 423

Waiting lines, 255–256, 290
Wilcoxon rank sum test, 465, 485
Wilson, Edwin, 430
Wilson estimators, 430–431
Within-treatments variation (SSE; sum of squares for error), for one-way ANOVA, 529–530

z scores (*z* tests), 272
for difference between two proportions, 500–502, 505
finding, 273–276, 278–282
of nominal data, 616
for population proportions, 424–425
power of, 391
for standardized test statistic, 367–368
table of, 274
z-statistic, 408

APPLICATION BOXES

Accounting

Breakeven analysis	Introduction 132
Fixed and variable costs	Least squares line to estimate fixed and variable costs 133

Banking

Credit scorecards	Histograms to compare credit scores of borrowers who repay and those who default 63

Economics

Macroeconomics	Introduction 24
Energy economics	Pie chart of sources of energy in the United States 24
Measuring inflation	Removing the effect of inflation in a time series of prices 68

Finance

Stock and bond valuation	Introduction 51
Return on investment	Histograms of two sets of returns to assess expected returns and risk 52
Geometric mean	Calculating average returns on an investment over time 104
Stock market indexes	Introduction to the market model 148
Mutual funds	Marginal and conditional probability relating mutual fund performance with manager's education 181
Measuring risk	Normal distribution to show why the standard deviation is a measure of risk 277

Human Resource Management

Employee retention	Regression analysis to predict which workers will stay on the job 645
Job applicant testing	Regression analysis to determine whether testing job applicants is effective 647
Severance pay	Multiple regression to judge consistency of severance packages to laid-off workers 708

Marketing

Pricing	Histogram of long-distance telephone bills 44
Advertising	Estimating mean exposure to advertising 353
Test marketing	Inference about the difference between two proportions of product purchases 499
Market segmentation	Inference about two proportions to determine whether market segments differ 511
Market segmentation	Inference about the difference between two means to determine whether two market segments differ 517
Test marketing	Analysis of variance to determine differences between pricing strategies 542
Market segmentation	Analysis of variance to determine differences between segments 542
Market segmentation	Chi-squared goodness-of-fit test to determine relative sizes of market segments 603
Market segmentation	Chi-squared test of a contingency table to determine whether several market segments differ 624

Operations Management

PERT/CPM	Expected value of the completion time of a project 235
Waiting lines	Poisson distribution to compute probabilities of arrivals 256
Inventory management	Normal distribution to determine the reorder point 283
PERT/CPM	Normal distribution to determine the probability of completing a project on time 287
Waiting lines	Exponential distribution to calculate probabilities of service completions 290
Inventory management	Estimating mean demand during lead time 342
Quality	Inference about a variance 415
Pharmaceutical and medical experiments	Inference about the difference between two drugs 508
Location analysis	Multiple regression to predict profitability of new locations 711

Index of Computer Output and Instructions

Techniques	Excel	Minitab
General		
Data input and retrieval	CD App. A1	CD App. B1
Recoding data	CD App. N	CD App. N
Stacking/Unstacking data	CD App. R	CD App. R
Graphical		
Frequency distribution	20	20
Bar chart	22	23
Pie chart	22	23
Histogram	47	48
Stem-and-leaf display	58	58
Ogive	60	—
Line chart	66	67
Pivot table	34	—
Cross-classification table	34	34
Scatter diagram	75	76
Box plot	121	122
Numerical descriptive techniques		
Descriptive statistics	103	103
Least squares	135	136
Correlation	138	138
Covariance	138	138
Determination	139	139
Probability/random variables		
Binomial	248	249
Poisson	255	255
Normal	282	282
Exponential	289	290
Student t	296	296
Chi-squared	300	300
F	304	304
Inference about μ (σ known)		
Interval estimator	343	343
Test statistic	372	373
Probability of Type II error	389	389
Inference about μ (σ unknown)		
Test statistic	402	403
Interval estimator	405	405
Inference about σ^2		
Test statistic	416	417
Interval estimator	418	418
Inference about p		
Test statistic	424	425
Interval estimator	427	427
Inference about $\mu_1 - \mu_2$		
Equal-variances test statistic	456	456
Equal-variances interval estimator	457	458
Unequal-variances test statistic	461	462
Unequal-variances interval estimator	463	463

Techniques	Excel	Minitab
Inference about μ_D		
Test statistic	480	481
Interval estimator	482	482
Inference about σ_1^2/σ_2^2		
Test statistic	491	492
Interval estimator	493	—
Inference about $p_1 - p_2$		
Test statistic	500	501
Interval estimator	504	504
Analysis of variance		
One-way	533	533
Multiple comparison methods	545	545
Two-way	558	558
Two-factor	570	571
Chi-squared tests		
Goodness-of-fit test	600	601
Contingency table	609	609
Test for normality	619	—
Linear regression		
Coefficients and tests	642	643
Correlation (Pearson)	662	662
Prediction interval	669	669
Regression diagnostics	672	673
Multiple regression		
Coefficients and tests	696	697
Prediction interval	706	706
Durbin-Watson test	721	721