significance with a smaller sample than a less powerful test. In addition, in choosing a test, one can consider how the sample is drawn, the nature of the population, and (importantly) the type of measurement scale used. For instance, some tests are useful only when the sequence of scores is known or when observations are paired; other tests are appropriate only if the population has certain characteristics; still other tests are useful only if the measurement scale is interval or ratio. More attention is given to test selection later in the chapter.

3. *Select the desired level of significance.* The choice of the **level of significance** should be made before we collect the data. The most common level is .05, although .01 is also widely used. Other levels such as .10, .025, or .001 are sometimes chosen. The exact level to choose is largely determined by how much risk one is willing to accept and the effect that this choice has on β risk. The larger the α, the lower is the β.
4. *Compute the calculated difference value.* After the data are collected, use the formula for the appropriate significance test to obtain the calculated value. Although the computation typically results from a software program, we illustrate the procedures in this chapter to help you visualize what is being done.
5. *Obtain the critical test value.* After we compute the calculated t, χ^2, or other measure, we must look up the critical value in the appropriate table for that distribution (or it is provided with the software calculation). The critical value is the criterion that defines the region of rejection from the region of acceptance of the null hypothesis.
6. *Interpret the test.* For most tests if the calculated value is larger than the critical value, we reject the null hypothesis and conclude that the alternative hypothesis is supported (although it is by no means proved). If the critical value is larger, we conclude we have failed to reject the null.[3]

Probability Values (*p* Values)

According to the "interpret the test" step of the statistical test procedure, the conclusion is stated in terms of rejecting or not rejecting the null hypothesis based on a reject region selected before the test is conducted. A second method of presenting the results of a statistical test reports the extent to which the test statistic disagrees with the null hypothesis. This method has become popular because analysts want to know what percentage of the sampling distribution lies beyond the sample statistic on the curve, and most statistical computer programs report the results of statistical tests as probability values (p values). The ***p* value** is the probability of observing a sample value as extreme as, or more extreme than, the value actually observed, given that the null hypothesis is true. This area represents the probability of a Type I error that must be assumed if the null hypothesis is rejected. The p value is compared to the significance level (α), and on this basis the null hypothesis is either rejected or not rejected. If the p value is less than the significance level, the null hypothesis is rejected (if p value $< \alpha$, reject the null). If p is greater than or equal to the significance level, the null hypothesis is not rejected (if p value $> \alpha$, don't reject the null).

Statistical data analysis programs commonly compute the p value during the execution of a hypothesis test. The following example will help illustrate the correct way to interpret a p value.

In part B of Exhibit 18-4 the critical value was shown for the situation where the manufacturer was interested in determining whether the average mpg had increased. The critical value of 53.29 was computed based on a standard deviation of 10, sample size of 25, and the manufacturer's willingness to accept a 5 percent α risk. Suppose that the sample mean equaled 55. Is there enough evidence to reject the null hypothesis? If the p value is less than .05, the null hypothesis will be rejected. The p value is computed as follows.

The standard deviation of the distribution of sample means is 2. The appropriate Z value is

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

$$Z = \frac{55 - 50}{2}$$

$$Z = 2.5$$

The p value is determined using the standard normal table. The area between the mean and a Z value of 2.5 is .4938. For this one-tailed test, the p value is the area above the Z value. The probability of observing a Z value at least as large as 2.5 is only .0062 (.5000 − .4938 = .0062) if the null hypothesis is true.

This small p value represents the risk of rejecting a true null hypothesis. It is the probability of a Type I error if the null hypothesis is rejected. Since the p value ($p = .0062$) is smaller than $\alpha = .05$, the null hypothesis is rejected. The manufacturer can conclude that the average mpg has increased. The probability that this conclusion is wrong is .0062.

> Tests of Significance

This section provides an overview of statistical tests that are representative of the vast array available to the researcher. After a review of the general types of tests and their assumptions, the procedures for selecting an appropriate test are discussed. The remainder of the section contains examples of parametric and nonparametric tests for one-sample, two-sample, and k-sample cases. Readers needing a comprehensive treatment of significance tests are referred to the suggested readings for this chapter.

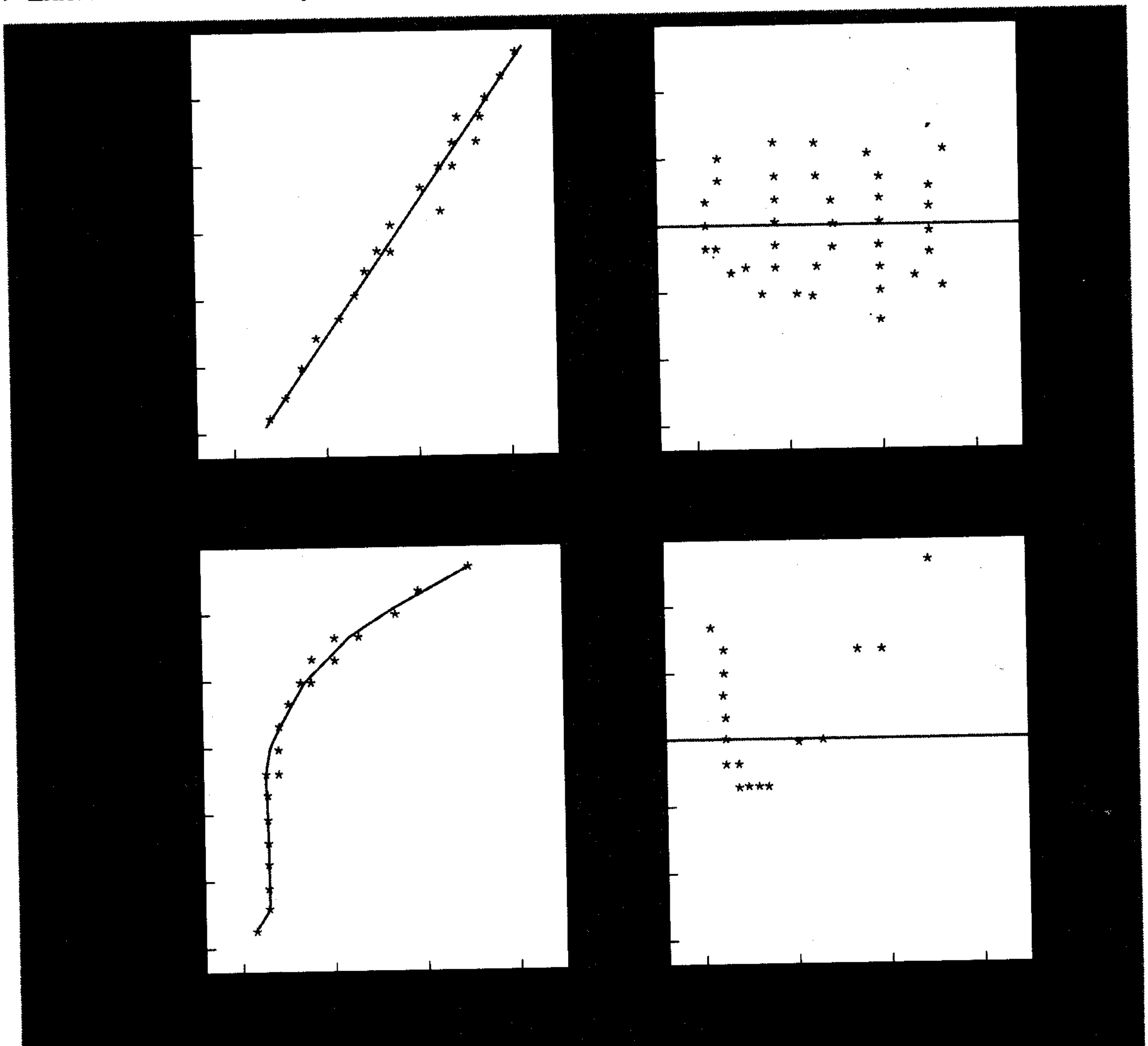
Types of Tests

There are two general classes of significance tests: parametric and nonparametric. **Parametric tests** are more powerful because their data are derived from interval and ratio measurements. **Nonparametric tests** are used to test hypotheses with nominal and ordinal data. Parametric techniques are the tests of choice if their assumptions are met. Assumptions for parametric tests include the following:

- The observations must be independent—that is, the selection of any one case should not affect the chances for any other case to be included in the sample.
- The observations should be drawn from normally distributed populations.
- These populations should have equal variances.
- The measurement scales should be at least interval so that arithmetic operations can be used with them.

The researcher is responsible for reviewing the assumptions pertinent to the chosen test. Performing diagnostic checks on the data allows the researcher to select the most appropriate technique. The normality of a distribution may be checked in several ways. We have previously discussed the measures of location, shape, and spread for preliminary analysis and considered graphic techniques for exploring data patterns and examining distributions. Another diagnostic tool is the **normal probability plot.** This plot compares the observed values with those expected from a normal distribution.[4] If the data display the characteristics of normality, the points will fall within a narrow band along a straight line. An example is shown in the upper left panel of Exhibit 18-6.

> Exhibit 18-6 Probability Plots and Tests of Normality



	Statistic	d.f.	Significance
Shapiro-Wilks	.6802	50	< .0100
K-S (Lilliefors)	.2558	50	.0000

An alternative way to look at this is to plot the deviations from the straight line. These are shown in a "detrended" plot in the upper right panel of the figure. Here we would expect the points to cluster without pattern around a straight line passing horizontally through 0. In the bottom two panels of Exhibit 18-6, there is neither a straight line in the normal probability plot nor a random distribution of points about 0 in the detrended plot. Visually, the bottom two plots tell us the variable is not normally distributed. In addition, two separate tests of the hypothesis that the data come from normal distributions are rejected at a significance level of less than .01.[5]

If we wished to check another assumption—say, one of equal variance—a spread-and-level plot would be appropriate. Statistical software programs often provide diagnostic tools for checking assumptions. These

>snapshot

Testing a Hypothesis of Unrealistic Drug Use in Movies

Are American teens exposed to unrealistic drug usage or to unrealistic consequences from such use? Mediascope, a nonprofit organization concerned with responsible depictions of social and health issues in the media, recently completed for the Office of National Drug Control Policy a content analysis of the top 200 rental movies to determine their depiction of substance use. The researchers used the Video Software Dealers Association's most popular (top 100) home video titles based on rental income during two sequential years. Movies were categorized as follows: action adventure, comedy, or drama. Data were also collected on each title's Motion Picture Association of America (MPAA) rating (G, PG, PG-13, or R). Although technically teens should have been excluded from R-rated titles (which made up 48 percent of the overall sample), the study included all 20 of the most popular teen movies as identified in a prior independent study.

Trained coders watched all 200 movies, paying particular attention to alcohol, tobacco, illicit drugs, over-the-counter medicines, prescription medicines, inhalants, and unidentified pills. Coders ignored substances administered by medical personnel in a hospital or health-related scenario. Substance use included explicit portrayals of consumption. Substance appearance was noted when evidence of materials or paraphernalia was noted without any indication of use. Coders identified dominant messages about substance use and the consequences of use. Coders also noted scenes depicting illicit drug use or those depicting use by characters known to be under 18. Prevalence of use was determined by counting the characters in each movie and determining not only the percentage of characters using drugs but also whether the character had a major or minor role. Coders profiled characters by age, gender, and ethnicity, as well as other characteristics. Frequency of substance abuse was determined for each five-minute interval of each movie, with the presence or absence of various substances noted, starting with the completion of the title credits and ending when the final credits began. How would the last movie you watched have fared under this scrutiny?

www.mediacampaign.org; www.americacares.org; www.vsda.org

may be nested within a specific statistical procedure, such as analysis of variance or regression, or provided as a general set of tools for examining assumptions.

Parametric tests place different emphasis on the importance of assumptions. Some tests are quite robust and hold up well despite violations. For others, a departure from linearity or equality of variance may threaten the validity of the results.

Nonparametric tests have fewer and less stringent assumptions. They do not specify normally distributed populations or equality of variance. Some tests require independence of cases; others are expressly designed for situations with related cases. Nonparametric tests are the only ones usable with nominal data; they are the only technically correct tests to use with ordinal data, although parametric tests are sometimes employed in this case. Nonparametric tests may also be used for interval and ratio data, although they waste much of the information available. Nonparametric tests are also easy to understand and use. Parametric tests have greater efficiency when their use is appropriate, but even in such cases nonparametric tests often achieve an efficiency as high as 95 percent. This means the nonparametric test with a sample of 100 will provide the same statistical testing power as a parametric test with a sample of 95.

How to Select a Test

In attempting to choose a particular significance test, the researcher should consider at least three questions:

- Does the test involve one sample, two samples, or k (more than two) samples?
- If two samples or k samples are involved, are the individual cases independent or related?
- Is the measurement scale nominal, ordinal, interval, or ratio?

Additional questions may arise once answers to these are known: What is the sample size? If there are several samples, are they of equal size? Have the data been weighted? Have the data been transformed? Often such questions are unique to the selected technique. The answers can complicate the selection, but once a tentative choice is made, standard statistics textbooks will provide further details.

Decision trees provide a more systematic means of selecting techniques. One widely used guide from the Institute for Social Research starts with questions about the number of variables, nature of the variables (continuous, discrete, dichotomous, independent, dependent, and so forth), and level of measurement. It goes through a tree structure asking detailed questions about the nature of the relationships being searched, compared, or tested. Over 130 solutions to data analysis problems are paired with commonly asked questions.[6]

An expert system offers another approach to choosing appropriate statistics. Capitalizing on the power and convenience of personal computers, expert system programs provide a comprehensive search of the statistical terrain just as a computer search of secondary sources does. Most programs ask about your research objectives, the nature of your data, and the intended audience for your final report. When you are not 100 percent confident of your answers, you can bracket them with an estimate of the degree of your certainty. One such program, Statistical Navigator, covers various categories of statistics from exploratory data analysis through reliability testing and multivariate data analysis. In response to your answers, a report is printed containing recommendations, rationale for selections, references, and the statistical packages that offer the suggested procedure.[7] SPSS and SAS include coaching and help modules with their software.

Selecting Tests Using the Choice Criteria

In the next section, we use the three questions discussed in the last section (see bullets) to develop a classification of the major parametric and nonparametric tests and measures. Because parametric tests are preferred for their power when their assumptions are met, we discuss them first in each of the subsections: one-sample tests, two-sample tests, *k*- (more-than-two) sample tests. This is shown in Exhibit 18-7.[8] To illustrate the application of the criteria to test selection, consider that your testing situation involves two samples, the samples are indepen-

> Exhibit 18-7 Recommended Statistical Techniques by Measurement Level and Testing Situation

		Two-Samples Tests		*k*-Samples Tests	
Measurement Scale	**One-Sample Case**	**Related Samples**	**Independent Samples**	**Related Samples**	**Independent Samples**
Nominal	• Binomial • χ^2 one-sample test	• McNemar	• Fisher exact test • χ^2 two-samples test	• Cochran Q	• χ^2 for *k* samples
Ordinal	• Kolmogorov-Smirnov one-sample test • Runs test	• Sign test • Wilcoxon matched-pairs test	• Median test • Mann-Whitney U • Kolmogorov-Smirnov • Wald-Wolfowitz	• Friedman two-way ANOVA	• Median extension • Kruskal-Wallis one-way ANOVA
Interval and Ratio	• *t*-test • *Z* test	• *t*-test for paired samples	• *t*-test • *Z* test	• Repeated-measures ANOVA	• One-way ANOVA • *n*-way ANOVA

dent, and the data are interval. The figure suggests the *t*-test of differences as the appropriate choice. The most frequently used of the tests listed in Exhibit 18-7 are covered next. For additional examples see Appendix B.

One-Sample Tests

One-sample tests are used when we have a single sample and wish to test the hypothesis that it comes from a specified population. In this case we encounter questions such as these:

- Is there a difference between observed frequencies and the frequencies we would expect, based on some theory?
- Is there a difference between observed and expected proportions?
- Is it reasonable to conclude that a sample is drawn from a population with some specified distribution (normal, Poisson, and so forth)?
- Is there a significant difference between some measures of central tendency ($\bar{X}$) and its population parameter (μ)?

A number of tests may be appropriate in this situation. The parametric test is discussed first.

Parametric Tests

The ***Z* test** or ***t*-test** is used to determine the statistical significance between a sample distribution mean and a parameter.

The ***Z* distribution** and ***t* distribution** differ. The *t* has more tail area than that found in the normal distribution. This is a compensation for the lack of information about the population standard deviation. Although the sample standard deviation is used as a proxy figure, the imprecision makes it necessary to go farther away from 0 to include the percentage of values in the *t* distribution necessarily found in the standard normal.

When sample sizes approach 120, the sample standard deviation becomes a very good estimate of the population standard deviation (σ); beyond 120, the *t* and *Z* distributions are virtually identical.

Some typical real-world applications of the one-sample test are:

- Finding the average monthly balance of credit card holders compared to the average monthly balance five years ago.
- Comparing the failure rate of computers in a 20-hour test of quality specifications.
- Discovering the proportion of people who would shop in a new district compared to the assumed population proportion.
- Comparing the average product revenues this year to last year's revenues.

Example To illustrate the application of the *t*-test in the one-sample case, consider again the hybrid-vehicle problem mentioned earlier. With a sample of 100 vehicles, the researchers find that the mean miles per gallon for the car is 52.5 mpg, with a standard deviation of 14. Do these results indicate the population mean might still be 50?

In this problem, we have only the sample standard deviation (s). This must be used in place of the population standard deviation (σ). When we substitute s for σ, we use the t distribution, especially if the sample size is less than 30. We define t as

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

This significance test is conducted by following the six-step procedure recommended earlier:

1. *Null hypothesis.*

H_0: = 50 miles per gallon (mpg)

H_A: > 50 mpg (one-tailed test)

2. *Statistical test.* Choose the t-test because the data are ratio measurements. Assume the underlying population is normal and we have randomly selected the sample from the population of production vehicles.
3. *Significance level.* Let $\alpha = .05$, with $n = 100$.
4. *Calculated value.*

$$t = \frac{52.5 - 50}{14\sqrt{100}} - \frac{2.5}{1.4} = 1.786 \qquad \text{d.f.} = n - 1 = 99$$

5. *Critical test value.* We obtain this by entering the table of critical values of t (see Appendix C, Exhibit C-2, at the back of the book), with 99 degrees of freedom (d.f.) and a level of significance value of .05. We secure a critical value of about 1.66 (interpolated between d.f. = 60 and d.f. = 120 in Exhibit C-2).
6. *Interpretation.* In this case, the calculated value is greater than the critical value (1.786 > 1.66), so we reject the null hypothesis and conclude that the average mpg has increased.

Nonparametric Tests

In a one-sample situation, a variety of nonparametric tests may be used, depending on the measurement scale and other conditions. If the measurement scale is nominal (classificatory only), it is possible to use either the binomial test or the chi-square (χ^2) one-sample test. The binomial test is appropriate when the population is viewed as only two classes, such as male and female, buyer and nonbuyer, and successful and unsuccessful, and all observations fall into one or the other of these categories. The binomial test is particularly useful when the size of the sample is so small that the χ^2 test cannot be used.

Chi-Square Test

Probably the most widely used nonparametric test of significance is the **chi-square (χ^2) test.** It is particularly useful in tests involving nominal data but can be used for higher scales. Typical are cases where persons, events, or objects are grouped in two or more nominal categories such as "yes-no," "favor-undecided-against," or class "A, B, C, or D."

Using this technique, we test for significant differences between the *observed* distribution of data among categories and the *expected* distribution based on the null hypothesis. Chi-square is useful in cases of one-sample analysis, two independent samples, or k independent samples. It must be calculated with actual counts rather than percentages.

In the one-sample case, we establish a null hypothesis based on the expected frequency of objects in each category. Then the deviations of the actual frequencies in each category are compared with the hypothesized frequencies. The greater the difference between them, the less is the probability that these differences can be attributed to chance. The value of χ^2 is the measure that expresses the extent of this difference. The larger the divergence, the larger is the χ^2 value.

The formula by which the χ^2 test is calculated is

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

in which

O_i = observed number of cases categorized in the ith category

E_i = expected number of cases in the ith category under H_0

k = the number of categories

There is a different distribution for χ^2 for each number of degrees of freedom (d.f.), defined as $(k - 1)$ or the number of categories in the classification minus 1:

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

With chi-square contingency tables of the two-samples or k-samples variety, we have both rows and columns in the cross-classification table. In that instance, d.f. is defined as rows minus 1 $(r - 1)$ times columns minus 1 $(c - 1)$:

$$\text{d.f.} = (r - 1)(c - 1)$$

In a 2×2 table there is 1 d.f., and in a 3×2 table there are 2 d.f. Depending on the number of degrees of freedom, we must be certain the numbers in each cell are large enough to make the χ^2 test appropriate. When d.f. = 1, each expected frequency should be at least 5 in size. If d.f. > 1, then the χ^2 test should not be used if more than 20 percent of the expected frequencies are smaller than 5 or when any expected frequency is less than 1. Expected frequencies can often be increased by combining adjacent categories. Four categories of freshmen, sophomores, juniors, and seniors might be classified into upper class and lower class. If there are only two categories and still there are too few in a given class, it is better to use the binomial test.

Assume a survey of student interest in the Metro University Dining Club (discussed in Chapter 15) is taken. We have interviewed 200 students and learned of their intentions to join the club. We would like to analyze the results by living arrangement (type and location of student housing and eating arrangements). The 200 responses are classified into the four categories shown in the accompanying table.

Living Arrangement	O Intend to Join	Number Interviewed	Percent (no. interviewed/200)	E = Expected Frequencies (percent × 60)
Dorm/fraternity	16	90	45	27
Apartment/rooming house, nearby	13	40	20	12
Apartment/rooming house, distant	16	40	20	12
Live at home	15	30	15	9
Total	60	200	100	60

Do these variations indicate there is a significant difference among these students, or are they sampling variations only? Proceed as follows:

1. *Null hypothesis.* H_0: $O_i = E_i$. The proportion in the population who intend to join the club is independent of living arrangement. In H_A: $O_i \neq E_i$, the proportion in the population who intend to join the club is dependent on living arrangement.
2. *Statistical test.* Use the one-sample χ^2 to compare the observed distribution to a hypothesized distribution. The χ^2 test is used because the responses are classified into nominal categories and there are sufficient observations.
3. *Significance level.* Let $\alpha = .05$.
4. *Calculated value.*

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

>snapshot

Research beyond the Clip

You're McDonald's, and you just announced that you are closing 300 stores in the United States. How is this playing in newspapers across America? Is the reporting balanced? Or maybe you're BP (formerly British Petroleum), and you've just invested millions to change your corporate identification and logo and to reposition your firm as the "environmentally friendly" energy conglomerate through executive presentations and advertising. How is the press treating the story? Is the spin positive or negative? Where is the story appearing? Is your key message getting through? The Burrelle's Information Services division of Burrelle's/Luce (B/L) offers one of the longest established "clipping" services used by public relations managers to answer questions like these.

"The most basic research we provide clients is the ad-equivalent value of the news clips," shared Sharon Miller, account executive for B/L. The client notifies B/L that it plans to distribute a press release. Staff at B/L scan the desired print and Internet sources for news of the client—"clips." "For print and online publications, the actual space that the story occupies is physically measured. Then, that space is multiplied by the ad rate for identical space in that medium—the ad equivalent." Burrelle's also delivers an assessment of the coverage of key messages the client tried to convey, as well as the tone of the story and the firm's prominence in any story—did the firm get mentioned in the headline or the lead paragraph, or was it the focus of more than half of the article? Managers can obtain comparative analysis evaluating their firm's news coverage against that of other firms in their industry or against ROI investment criteria through online reporting via B/L's secure Insight platform. If you were a public relations professional, how would you test the hypothesis that your coverage for any given event was more positive than that of your competition?

www.burrellesluce.com

Calculate the expected distribution by determining what proportion of the 200 students interviewed were in each group. Then apply these proportions to the number who intend to join the club. Then calculate the following:

$$\chi^2 = \frac{(16-27)^2}{27} + \frac{(13-12)^2}{12} + \frac{(16-12)^2}{12} + \frac{(15-9)^2}{9}$$

$$= 4.48 + 0.08 + 1.33 + 4.0$$

$$= 9.89$$

$$\text{d.f.} = (4-1)(2-1) = 3$$

5. *Critical test value.* Enter the table of critical values of χ^2 (see Exhibit C-3), with 3 d.f., and secure a value of 7.82 for $\alpha = .05$.
6. *Interpretation.* The calculated value (9.89) is greater than the critical value (7.82), so the null hypothesis is rejected and we conclude that intending to join is dependent on living arrangement.

Two-Independent-Samples Tests

The need to use **two-independent-samples tests** is often encountered in business research. We might compare the purchasing predispositions of a sample of subscribers from two magazines to discover if they are from the same population. Similarly, a test of distribution methods from two channels or the market share movements from two competing products could be compared.

Parametric Tests

The Z and *t*-tests are frequently used parametric tests for independent samples, although the *F* test also can be used.

The Z test is used with large sample sizes (exceeding 30 for both independent samples) or with smaller samples when the data are normally distributed and population variances are known. The formula for the Z test is

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)0}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$$

With small sample sizes, normally distributed populations, and the assumption of equal population variances, the t-test is appropriate:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)0}{\sqrt{S_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$$

where

$(\mu_1 - \mu_2)$ is the difference between the two population means.

S_p^2 is associated with the pooled variance estimate:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

To illustrate this application, consider a problem that might face a manager at KDL, a media firm that is evaluating account executive trainees. The manager wishes to test the effectiveness of two methods for training new account executives. The company selects 22 trainees, who are randomly divided into two experimental groups. One receives type A and the other type B training. The trainees are then assigned and managed without regard to the training they have received. At the year's end, the manager reviews the performances of employees in these groups and finds the following results:

	A Group	B Group
Average hourly sales	$X_1 = \$1,500$	$X_2 = \$1,300$
Standard deviation	$s_1 = 225$	$s_2 = 251$

Following the standard testing procedure, we will determine whether one training method is superior to the other:

1. *Null hypothesis.*

   H_0: There is no difference in sales results produced by the two training methods.

   H_A: Training method A produces sales results superior to those of method B.
2. *Statistical test.* The t-test is chosen because the data are at least interval and the samples are independent.
3. *Significance level.* $\alpha = .05$ (one-tailed test).
4. *Calculated value.*

$$t = \frac{(1{,}500 - 1{,}300) - 0}{\sqrt{\dfrac{(10)(225)^2 + (10)(251)^2}{20}\left(\dfrac{1}{11} + \dfrac{1}{11}\right)}}$$

$$= \frac{200}{101.63} = 1.97$$

There are $n - 1$ degrees of freedom in each sample, so total d.f. is

$$\text{d.f.} = (11 - 1) + (11 - 1) = 20$$

5. *Critical test value.* Enter Appendix C, Exhibit C-2 with d.f. = 20, one-tailed test, $\alpha = .05$. The critical value is 1.725.
6. *Interpretation.* Since the calculated value is larger than the critical value ($1.97 > 1.725$), reject the null hypothesis and conclude that training method A is superior.

Nonparametric Tests

The chi-square (χ^2) test is appropriate for situations in which a test for differences between samples is required. It is especially valuable for nominal data but can be used with ordinal measurements. When parametric data have been reduced to categories, they are frequently treated with χ^2 although this results in a loss of information. Preparing to solve this problem is the same as presented earlier although the formula differs slightly:

$$\chi^2 = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

in which

O_{ij} = observed number of cases categorized in the ijth cell

E_{ij} = expected number of cases under H_0 to be categorized in the ijth cell

Suppose MindWriter is implementing a smoke-free workplace policy and is interested in whether smoking affects worker accidents. Since the company has complete reports on on-the-job accidents, a sample of names of workers is drawn from those who were involved in accidents during the last year. A similar sample from among workers who had no reported accidents in the last year is drawn. Members of both groups are interviewed to determine if each is a nonsmoker or smoker and, if a smoker, whether the person classifies himself or herself as a heavy or moderate smoker. The results appear in the table below, with expected values calculated as shown.

	Cell Designation **Count** **Expected Values**	**On-the-Job Accident** **Yes**	**No**	**Total**
Smoker	Heavy smoker	[redacted] 127 [redacted]	[redacted] 4 7.75	16
	Moderate	[redacted] 9 [redacted]	[redacted] 6 [redacted]	15
	Nonsmoker	3,1 13 18.03	3,2 22 16.97	35
	Column Total	34	32	66

The testing procedure is:

1. *Null hypothesis.*

   H_0: There is no relationship in on-the-job accident occurrences between smokers and nonsmokers.

   H_A: There is a relationship in on-the-job accident occurences between smokers and nonsmokers.

2. *Statistical test.* χ^2 is appropriate, but it may waste some of the data because the measurement appears to be ordinal.
3. *Significance level.* $\alpha = .05$, with d.f. $= (3 - 1)(2 - 1) = 2$
4. *Calculated value.* The expected distribution is provided by the marginal totals of the table. If there is no relationship between accidents and smoking, there will be the same proportion of smokers in both accident and nonaccident groups. The numbers of expected observations in each cell are calculated by multiplying the two marginal totals common to a particular cell and dividing this product by n. For example,

$$\frac{34 \times 16}{66} = 8.24, \text{ the expected value in cell (1,1)}$$

$$\chi^2 = \frac{(12 - 8.24)^2}{8.24} + \frac{(4 - 7.75)^2}{7.75} + \frac{(9 - 7.73)^2}{7.73} + \frac{(6 - 7.72)^2}{7.72} + \frac{(13 - 18.03)^2}{18.03} + \frac{(22 - 16.97)^2}{16.97}$$

$$= 7.01$$

5. *Critical test value.* Turn to Appendix C, Exhibit C-3, and find the critical value 7.01 with $\alpha = .05$ and d.f. $= 2$.
6. *Interpretation.* Since the calculated value is greater than the critical value, the null hypothesis is rejected.

For chi-square to operate properly, data must come from random samples of multinomial distributions, and the expected frequencies should not be too small. We previously noted the traditional cautions that expected frequencies (E_i) below 5 should not compose more than 20 percent of the cells, and that no cell should have an E_i of less than 1. Some research has argued that these restrictions are too severe.[9]

In another type of χ^2, the 2×2 table, a correction known as *Yates' correction for continuity* is applied when sample sizes are greater than 40 or when the sample is between 20 and 40 and the values of E_i are 5 or more. (We use this correction because a continuous distribution is approximating a discrete distribution in this table. When the E_i's are small, the approximation is not necessarily a good one.) The formula for this correction is

$$\chi^2 = \frac{n\left(|AD - BC| - \frac{n}{2}\right)^2}{(A + B)(C + D)(A + C)(B + D)}$$

where the letters represent the cells designated as

A	*B*
C	*D*

When the continuity correction is applied to the data shown in Exhibit 18-8, a χ^2 value of 5.25 is obtained. The observed level of significance for this value is .02192. If the level of significance had been set at .01, we

> Exhibit 18-8 Comparison of Corrected and Noncorrected Chi-Square Results Using SPSS Procedure Cross-Tab

INCOME BY POSSESSION OF MBA

INCOME	Count	MBA Yes 1	No 2	Row Total
	High 1	30	30	60 60.0
	Low 2	10	30	40 40.0
	Column Total	40 40.0	60 60.0	100 100.0

Chi-Square	Value	D.F.	Significance
Pearson	[redacted]	1	[redacted]
Continuity Correction		1	
Likelihood Ratio	6.43786	1	.01117
Mantel-Haenszel	6.18750	1	.01287
Minimum Expected Frequency: 16.000			

would accept the null hypothesis. However, had we calculated χ^2 without correction, the value would have been [redacted] which has an observed level of significance of [redacted] Some researchers may be tempted to reject the null at this level. (But note that the critical value of χ^2 at .01 with 1 d.f. is 6.64. See Appendix C, Exhibit C-3.) The literature is in conflict regarding the merits of Yates' correction, but if nothing else, this example suggests one should take care when interpreting 2 × 2 tables.[10] To err on the conservative side would be in keeping with our prior discussion of Type I errors.

The Mantel-Haenszel test and the likelihood ratio also appear in Exhibit 18-8. The former is used with ordinal data; the latter, based on maximum likelihood theory, produces results similar to Pearson's chi-square.

Two-Related-Samples Tests

The **two-related-samples tests** concern those situations in which persons, objects, or events are closely matched or the phenomena are measured twice. One might compare the consumption of husbands and wives, the performance of employees before and after vacations, or the effects of a marketing test stimulus when persons were randomly assigned to groups and given pretests and posttests. Both parametric and nonparametric tests are applicable under these conditions.

Parametric Tests

The *t*-test for independent samples would be inappropriate for this situation because one of its assumptions is that observations are independent. This problem is solved by a formula where the difference is found between each matched pair of observations, thereby reducing the two samples to the equivalent of a one-sample case—that is, there are now several differences, each independent of the other, for which one can compute various statistics.

In the following formula, the average difference $\bar{D}$ corresponds to the normal distribution when the α difference is known and the sample size is sufficient. The statistic t with $(n - 1)$ degrees of freedom is defined as

$$t = \frac{\bar{D}}{S_D/\sqrt{n}}$$

where

$$\bar{D} = \frac{\Sigma D}{n}$$

$$S_D = \sqrt{\frac{\Sigma D^2 - \frac{(\Sigma D)^2}{n}}{n-1}}$$

To illustrate, we use two years of *Forbes* sales data (in millions of dollars) from 10 companies, as listed in Exhibit 18-9.

1. *Null hypothesis.*
   H_0: $\mu = 0$; there is no difference between year 1 and year 2 sales.
   H_A: $\mu \neq 0$; there is a difference between year 1 and year 2 sales.
2. *Statistical test.* The matched- or paired-samples t-test is chosen because there are repeated measures on each company, the data are not independent, and the measurement is ratio.
3. *Significance level.* Let $\alpha = .01$, with $n = 10$ and d.f. $= n - 1$.
4. *Calculated value.*

$$t = \frac{\bar{D}}{S_D/\sqrt{n}} = \frac{3582.10}{570.98} = 6.27 \qquad \text{d.f.} = 9$$

5. *Critical test value.* Enter Appendix C, Exhibit C-2, with d.f. = 9, two-tailed test, $\alpha = .01$. The critical value is 3.25.
6. *Interpretation.* Since the calculated value is greater than the critical value (6.27 > 3.25), reject the null hypothesis and conclude there is a statistically significant difference between the two years of sales.

> Exhibit 18-9 Sales Data for Paired-Samples t-Test (dollars in millions)

Company	Sales Year 2	Sales Year 1	Difference D	D^2
GM	126932	123505	3427	11744329
GE	54574	49662	4912	24127744
Exxon	86656	78944	7712	59474944
IBM	62710	59512	3198	10227204
Ford	96146	92300	3846	14791716
AT&T	36112	35173	939	881721
Mobil	50220	48111	2109	4447881
DuPont	35099	32427	2672	7139584
Sears	53794	49975	3819	14584761
Amoco	23966	20779	3187	10156969
			$\Sigma D = 35821$	$\Sigma D^2 = 157576853$

> Exhibit 18-10 SPSS Output for Paired-Samples *t*-Test (dollars in millions)

---*t*-tests for paired samples---

Variable	Number of Cases	Mean	Standard Deviation	Standard Error
Year 2 Sales	10	62620.9	31777.649	10048.975
Year 1 Sales	10	59038.8	31072.871	9836.104

(Difference Mean)	Standard Deviation	Standard Error	Corr.	2-tail Prob.	*t* Value	Degrees of Freedom	2-tail Prob.
3582.1000	1803.159	570.209	.999	.000	6.28	9	.000

A computer solution to the problem is illustrated in Exhibit 18-10. Notice that an **observed significance level** is printed for the calculated t value (highlighted). With SPSS, this is often rounded and would be interpreted as significant at the .0005 level. The correlation coefficient, to the left of the t value, is a measure of the relationship between the two pairs of scores. In situations where matching has occurred (such as husbands' and wives' scores), it reveals the degree to which the matching has been effective in reducing the variability of the mean difference.

Nonparametric Tests

The *McNemar test* may be used with either nominal or ordinal data and is especially useful with before-after measurement of the same subjects. Test the significance of any observed change by setting up a fourfold table of frequencies to represent the first and second set of responses:

	After	
Before	**Do Not Favor**	**Favor**
Favor	*A*	*B*
Do Not Favor	*C*	*D*

Since $A + D$ represents the total number of people who changed (B and C are no-change responses), the expectation under a null hypothesis is that 1/2 $(A + D)$ cases change in one direction and the same proportion in the other direction. The McNemar test uses a transformation of the χ^2 test:

$$\chi^2 = \frac{(|A - D| - 1)^2}{A + D} \text{ with d.f.} = 1$$

The "minus 1" in the equation is a correction for continuity since the χ^2 is a continuous distribution and the observed frequencies represent a discrete distribution.

To illustrate this test's application, we use survey data from SteelShelf Corporation, whose researchers decided to test a new concept in office seating with employees at the company's headquarters facility. Managers took a random sample of their employees before the test, asking them to complete a questionnaire on their attitudes toward the design concept. On the basis of their responses, the employees were divided into equal groups reflecting their favorable or unfavorable views of the design. After the campaign, the same 200 employees were asked again to complete the questionnaire. They were again classified as to favorable or unfavorable attitudes. The testing process is:

1. *Null hypothesis.*

$$H_0: P(A) = P(D)$$
$$H_A: P(A) \neq P(D)$$

2. *Statistical test.* The McNemar test is chosen because nominal data are used and the study involves before-after measurements of two related samples.
3. *Significance level.* Let $\alpha = .05$, with $n = 200$.
4. *Calculated value.*

$$\chi^2 = \frac{(|10 - 40| - 1)^2}{10 + 40} = \frac{29^2}{50} = 16.82 \qquad \text{d.f.} = 1$$

	After	
Before	Do Not Favor	Favor
Favor	$A = 10$	$B = 90$
Do Not Favor	$C = 60$	$D = 40$

5. *Critical test value.* Enter Appendix C, Exhibit C-3, and find the critical value to be 3.84 with $\alpha = .05$ and d.f. = 1.
6. *Interpretation.* The calculated value is greater than the critical value (16.82 > 3.84), indicating one should reject the null hypothesis, and conclude that the new concept had a significant positive effect on employees' attitudes. In fact, χ^2 is so large that it would have surpassed an α of .001.

k-Independent-Samples Tests

We often use ***k*-independent-samples tests** in research when three or more samples are involved. Under this condition, we are interested in learning whether the samples might have come from the same or identical populations. When the data are measured on an interval-ratio scale and we can meet the necessary assumptions, analysis of variance and the *F* test are used. If preliminary analysis shows the assumptions cannot be met or if the data were measured on an ordinal or nominal scale, a nonparametric test should be selected.

As with the two-samples case, the samples are assumed to be independent. This is the condition of a completely randomized experiment when subjects are randomly assigned to various treatment groups. It is also common for an ex post facto study to require comparison of more than two independent sample means.

< We discussed factors in Chapter 11.

Parametric Tests

The statistical method for testing the null hypothesis that the means of several populations are equal is **analysis of variance (ANOVA)**. *One-way analysis of variance* is described in this section. It uses a single-factor, fixed-effects model to compare the effects of one *treatment* or *factor* (brands of coffee, varieties of residential housing, types of retail stores) on a continuous dependent variable (coffee consumption, hours of TV viewing, shopping expenditures). In a fixed-effects model, the levels of the factor are established in advance, and the results are not generalizable to other levels of treatment. For example, if coffee were Jamaican-grown, Colombian-grown, and Honduran-grown, we could not extend our inferences to coffee grown in Guatemala or Mexico.

To use ANOVA, certain conditions must be met. The samples must be randomly selected from normal populations, and the populations should have equal variances. In addition, the distance from one value to its group's mean should be independent of the distances of other values to that mean (independence of error). ANOVA is reasonably robust, and minor variations from normality and equal variance are tolerable. Nevertheless, the analyst should check the assumptions with the diagnostic techniques previously described.

Analysis of variance, as the name implies, breaks down or partitions total variability into component parts. Unlike the *t*-test, which uses sample standard deviations, ANOVA uses squared deviations of the variance so that computation of distances of the individual data points from their own mean or from the grand mean can be summed (recall that standard deviations sum to zero).

In an ANOVA model, each group has its own mean and values that deviate from that mean. Similarly, all the data points from all of the groups produce an overall *grand mean.* The total deviation is the sum of the squared differences between each data point and the overall grand mean.

The total deviation of any particular data point may be partitioned *into between-groups variance* and *within-groups variance.* The between-groups variance represents the effect of the treatment, or factor. The differences of between-groups means imply that each group was treated differently, and the treatment will appear as deviations of the sample means from the grand mean. Even if this were not so, there would still be some natural variability among subjects and some variability attributable to sampling. The within-groups variance describes the deviations of the data points within each group from the sample mean. This results from variability among subjects and from random variation. It is often called *error.*

Intuitively, we might conclude that when the variability attributable to the treatment exceeds the variability arising from error and random fluctuations, the viability of the null hypothesis begins to diminish. And this is exactly the way the test statistic for analysis of variance works.

The test statistic for ANOVA is the ***F* ratio.** It compares the variance from the last two sources:

$$F = \frac{\text{between-groups variance}}{\text{within-groups variance}} = \frac{\text{mean square}_{\text{between}}}{\text{mean square}_{\text{within}}}$$

where

$$\text{Mean square}_{\text{between}} = \frac{\text{sum of squares}_{\text{between}}}{\text{degrees of freedom}_{\text{between}}}$$

$$\text{Mean square}_{\text{within}} = \frac{\text{sum of squares}_{\text{within}}}{\text{degrees of freedom}_{\text{within}}}$$

To compute the F ratio, the sum of the squared deviations for the numerator and denominator are divided by their respective degrees of freedom. By dividing, we are computing the variance as an average or mean, thus the term **mean square.** The degrees of freedom for the numerator, the mean square between groups, are one less than the number of groups $(k - 1)$. The degrees of freedom for the denominator, the mean square within groups, are the total number of observations minus the number of groups $(n - k)$.

If the null hypothesis is true, there should be no difference between the population means, and the ratio should be close to 1. If the population means are not equal, the numerator should manifest this difference, and the F ratio should be greater than 1. The F distribution determines the size of ratio necessary to reject the null hypothesis for a particular sample size and level of significance.

To illustrate one-way ANOVA, consider *Travel Industry Magazine*'s reports from international travelers about the quality of in-flight service on various carriers from the United States to Europe. Before writing a feature story coinciding with a peak travel period, the magazine decided to retain a researcher to secure a more balanced perspective on the reactions of travelers. The researcher selected passengers who had current impressions of the meal service, comfort, and friendliness of a major carrier. Three airlines were chosen and 20 passengers were randomly selected for each airline. The data, found in Exhibit 18-11,[11] are used for this and the next two examples. For the one-way analysis of variance problem, we are concerned only with the columns labeled "Flight Service Rating 1" and "Airline." The factor, airline, is the grouping variable for three carriers.

> **Exhibit 18-11** Data Table: Analysis of Variance Examples*

	Flight Service					Flight Service			
	Rating 1	Rating 2	Airline†	Seat Selection‡		Rating 1	Rating 2	Airline†	Seat Selection‡
1	40	36	1	1	31	52	65	2	2
2	28	28	1	1	32	70	80	2	2
3	36	30	1	1	33	73	79	2	2
4	32	28	1	1	34	72	88	2	2
5	60	40	1	1	35	73	89	2	2
6	12	14	1	1	36	71	72	2	2
7	32	26	1	1	37	55	58	2	2
8	36	30	1	1	38	68	67	2	2
9	44	38	1	1	39	81	85	2	2
10	36	35	1	1	40	78	80	2	2
11	40	42	1	2	41	92	95	3	1
12	68	49	1	2	42	56	60	3	1
13	20	24	1	2	43	64	70	3	1
14	33	35	1	2	44	72	78	3	1
15	65	40	1	2	45	48	65	3	1
16	40	36	1	2	46	52	70	3	1
17	51	29	1	2	47	64	79	3	1
18	25	24	1	2	48	68	81	3	1
19	37	23	1	2	49	76	69	3	1
20	44	41	1	2	50	56	78	3	1
21	56	67	2	1	51	88	92	3	2
22	48	58	2	1	52	79	85	3	2
23	64	78	2	1	53	92	94	3	2
24	56	68	2	1	54	88	93	3	2
25	28	69	2	1	55	73	90	3	2
26	32	74	2	1	56	68	67	3	2
27	42	55	2	1	57	81	85	3	2
28	40	55	2	1	58	95	95	3	2
29	61	80	2	1	59	68	67	3	2
30	58	78	2	1	60	78	83	3	2

* All data are hypothetical.
† Airline: 1 = Delta; 2 = Lufthansa; 3 = KLM.
‡ Seat selection: 1 = economy; 2 = business.

Again, we follow the procedure:

1. *Null hypothesis.*

   H_0: $\mu_{A1} = \mu_{A2} = \mu_{A3}$

   H_A: $\mu_{A1} \neq \mu_{A2} \neq \mu_{A3}$ (The means are not equal.)
2. *Statistical test.* The F test is chosen because we have k independent samples, accept the assumptions of analysis of variance, and have interval data.
3. *Significance level.* Let $\alpha = .05$, and d.f. = [numerator $(k - 1) = (3 - 1) = 2$], [denominator $(n - k) = (60 - 3) = 57$] = (2, 57).
4. *Calculated value.*

$$F = \frac{MS_b}{MS_w} = \frac{5822.017}{205.695} = 28.304 \qquad \text{d.f. } (2, 57)$$

   See summary in Exhibit 18-12.
5. *Critical test value.* Enter Appendix C, Exhibit C-8, with d.f. (2, 57), α =.05. The critical value is 3.16.
6. *Interpretation.* Since the calculated value is greater than the critical value (28.3 > 3.16), we reject the null hypothesis and conclude there are statistically significant differences between two or more pairs of means. Note in Exhibit 18-12 that the p value equals .0001. Since the p value (.0001) is less than the significance level (.05), we have a second method for rejecting the null hypothesis.

> Exhibit 18-12 Summary Tables for One-Way ANOVA Example*

Model Summary†

Source	d.f.	Sum of Squares	Mean Square	*F* Value	*p* Value
Model (airline)	2	11644.033	5822.017	28.304	0.0001
Residual (error)	57	11724.550	205.694		
Total	59	23368.583			

Means Table

	Count	Mean	Std. Dev.	Std. Error
Delta	20	38.950	14.006	3.132
Lufthansa	20	58.900	15.089	3.374
KLM	20	72.900	13.902	3.108

Scheffè's *S* Multiple Comparison Procedure‡

	Vs.	Diff.	Crit. Diff.	*p* Value	
Delta	Lufthansa	19.950	11.400	0.0002	*S*
	KLM	33.950	11.400	0.0001	*S*
Lufthansa	KLM	14.000	11.400	0.0122	*S*

*All data are hypothetical.

† Factor: airline; dependent: flight service rating 1.

‡ *S* = significantly different at the .05 level; significance level: .05.

The ANOVA model summary in Exhibit 18-12 is a standard way of summarizing the results of analysis of variance. This table contains the sources of variation, degrees of freedom, sum of squares, mean squares, and calculated F value. The probability of rejecting the null hypothesis is computed up to 100 percent α—that is, the probability value column reports the exact significance for the F ratio being tested.

A Priori Contrasts

When we compute a t-test, it is not difficult to discover the reasons why the null is rejected. But with one-way ANOVA, how do we determine which pairs are not equal? We could calculate a series of t-tests, but they would not be independent of each other and the resulting Type I error would increase substantially. This is not recommended. If we decided in advance that a comparison of specific populations was important, a special class of tests known as ***a priori* contrasts** could be used after the null was rejected with the F test (it is *a priori* because the decision was made before the test).[12]

A modification of the F test provides one approach for computing contrasts:

$$F = \frac{\text{MS}_{\text{CON}}}{\text{MS}_W}$$

The denominator, the within-groups mean square, is the same as the error term of the one-way's F ratio (recorded in the summary table, Exhibit 18-12). We have previously referred to the denominator of the F ratio as the error variance estimator. The numerator of the contrast test is defined as

$$\text{MS}_{\text{CON}} = \text{SS}_{\text{CON}} = \frac{\left(\sum_j C_j X_j\right)^2}{\sum_j \frac{C_j^2}{n}}$$

where

C_j = the contrast coefficient for the group j

n_j = the number of observations recorded for group j

A contrast is useful for experimental and quasi-experimental designs when the researcher is interested in answering specific questions about a subset of the factor. For example, in a comparison of coffee products, we have a factor with six levels. The levels, blends of coffee, are meaningfully ordered. Assume we are particularly interested in two Central American–grown blends and one Colombian blend. Rather than looking at all possible combinations, we can channel the power more effectively by stating the comparisons of interest. This increases our likelihood of detecting differences if they really exist.

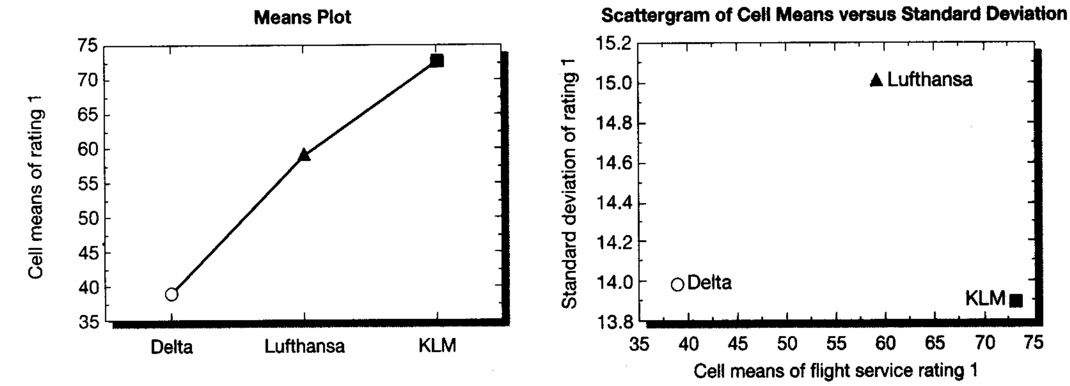
Multiple Comparison Tests

For the probabilities associated with the contrast test to be properly used in the report of our findings, it is important that the contrast strategy be devised ahead of the testing. In the airline study, we had no theoretical reason for an *a priori* contrast. However, when we examine the table of mean ratings (Exhibit 18-12), it is apparent that the airline means were quite different. Comparisons after the results are compared require *post hoc* tests or pairwise **multiple comparison tests** (or *range tests*) to determine which means differ. These tests find homogeneous subsets of means that are not different from each other. Multiple comparisons test the difference between each pair of means and indicate significantly different group means at an α level of .05, or another level that you specify. Multiple comparison tests use group means and incorporate the MS_{error} term of

> **Exhibit 18-13** Selection of Multiple Comparison Procedures

Test	Pairwise Comparisons	Complex Comparisons	Equal *n*'s Only	Unequal *n*'s	Equal Variances Assumed	Unequal Variances Not Assumed
Fisher LSD	X			X	X	
Bonferroni	X		X	X		
Tukey HSD	X		X		X	
Tukey-Kramer	X			X	X	
Games-Howell	X			X		X
Tamhane T2	X			X		X
Scheffé S		X		X	X	
Brown-Forsythe		X		X		X
Newman-Keuls	X		X		X	
Duncan	X		X		X	
Dunnett's T3						X
Dunnett's C						X

> **Exhibit 18-14** One-Way Analysis of Variance Plots



the *F* ratio. Together they produce confidence intervals for the population means and a criterion score. Differences between the mean values may be compared.

There are more than a dozen such tests with different optimization goals: maximum number of comparisons, unequal cell size compensation, cell homogeneity, reduction of Type I or Type II errors, and so forth. The merits of various tests have produced considerable debate among statisticians, leaving the researcher without much guidance for the selection of a test. In Exhibit 18-13, we provide a general guide. For the example in Exhibit 18-12, we chose Scheffé's *S*. It is a conservative test that is robust to violations of assumptions.[13] The computer calculated the critical difference criterion as 11.4; all the differences between the pairs of means exceed this. The null hypothesis for the Scheffé was tested at the .05 level. Therefore, we can conclude that all combinations of flight service mean scores differ from each other.

While the table in Exhibit 18-12 provides information for understanding the rejection of the one-way null hypothesis and the Scheffé null, in Exhibit 18-14 we use plots for the comparisons. The means plot shows relative differences among the three levels of the factor. The means by standard deviations plot reveals lower variability in the opinions recorded by the hypothetical Delta and KLM passengers. Nevertheless, these two groups are sharply divided on the quality of in-flight service, and that is apparent in the upper plot.

Exploring the Findings with Two-Way ANOVA

Is the airline on which the passengers traveled the only factor influencing perceptions of in-flight service? By extending the one-way ANOVA, we can learn more about the service ratings. There are many possible explanations. We have chosen to look at the seat selection of the travelers in the interest of brevity.

Recall that in Exhibit 18-11, data were entered for the variable seat selection: economy and business-class travelers. Adding this factor to the model, we have a *two-way* analysis of variance. Now three questions may be considered with one model:

- Are differences in flight service ratings attributable to airlines?
- Are differences in flight service ratings attributable to seat selection?
- Do the airline and the seat selection interact with respect to flight service ratings?

The third question reveals a distinct advantage of the two-way model. A separate one-way model on airlines averages out the effects of seat selection. Similarly, a single-factor test of seat selection averages out the effects of the airline choice. But an interaction test of airline by seat selection considers them *jointly*.

Exhibit 18-15 reports a test of the hypotheses for these three questions. The significance level was established at the .01 level. We first inspect the interaction effect, airline by seat selection, since the individual *main effects* cannot be considered separately if the factors interact. The interaction was not significant at the .01 level, and the null is accepted. Now the separate main effects, airline and seat selection, can be verified.

> Exhibit 18-15 Summary Table for Two-Way ANOVA Example*

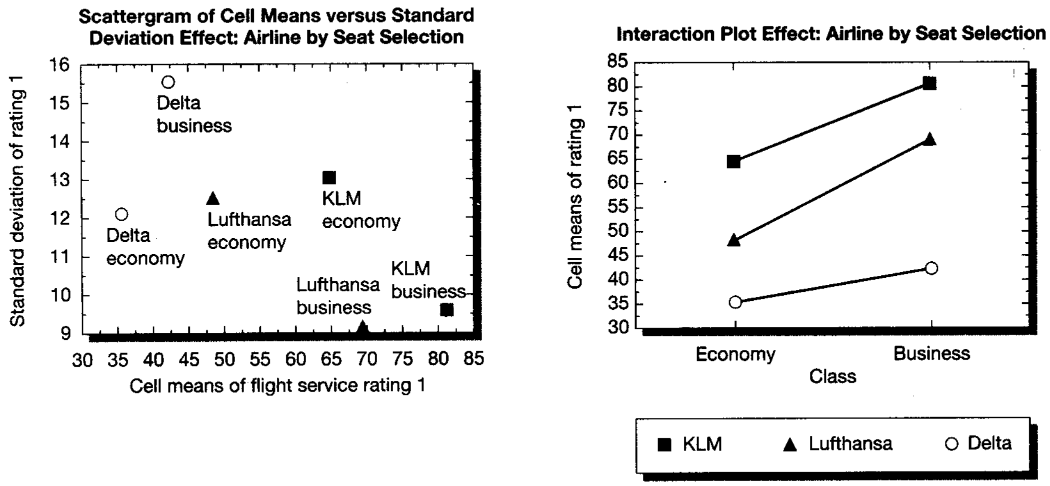
Model Summary†					
Source	**d.f.**	**Sum of Squares**	**Mean Square**	***F* Value**	***p* Value**
Airline	2	11644.033	5822.017	39.178	0.0001
Seat selection	1	3182.817	3182.817	21.418	0.0001
Airline by seat selection	2	517.033	258.517	1.740	0.1853
Residual	54	8024.700	148.606		

Means Table Effect: Airline by Seat Selection				
	Count	**Mean**	**Std. Dev.**	**Std. Error**
Delta economy	10	35.600	12.140	3.839
Delta business	10	42.300	15.550	4.917
Lufthansa economy	10	48.500	12.501	3.953
Lufthansa business	10	69.300	9.166	2.898
KLM economy	10	64.800	13.037	4.123
KLM business	10	81.000	9.603	3.037

*All data are hypothetical.

†Dependent: Flight service rating 1.

> Exhibit 18-16 Two-Way Analysis of Variance Plots



As with the one-way ANOVA, the null hypothesis for the airline factor was rejected, and seat selection was also rejected (statistically significant at .0001).

Means and standard deviations listed in the table are plotted in Exhibit 18-16. We note a band of similar deviations for economy-class travelers and a band of lower variability for business class—with the exception of one carrier. The plot of cell means confirms visually what we already know from the summary table: There is no interaction between airline and seat selection ($p = .185$). If an interaction had occurred, the lines connecting the cell means would have crossed rather than displaying a parallel pattern.

Analysis of variance is an extremely versatile and powerful method that may be adapted to a wide range of testing applications. Discussions of further extensions in n-way and experimental designs may be found in the list of suggested readings.

Nonparametric Tests

When there are k independent samples for which nominal data have been collected, the chi-square test is appropriate. It can also be used to classify data at higher measurement levels, but metric information is lost when reduced. The k-samples χ^2 test is an extension of the two-independent-samples cases treated earlier. It is calculated and interpreted in the same way.

The Kruskal-Wallis test is appropriate for data that are collected on an ordinal scale or for interval data that do not meet F-test assumptions, that cannot be transformed, or that for another reason prove to be unsuitable for a parametric test. Kruskal-Wallis is a one-way analysis of variance by ranks. It assumes random selection and independence of samples and an underlying continuous distribution.

Data are prepared by converting ratings or scores to ranks for each observation being evaluated. The ranks range from the highest to the lowest of all data points in the aggregated samples. The ranks are then tested to decide if they are samples from the same population. An application of this technique is provided in Appendix B.

k-Related-Samples Tests

Parametric Tests

A ***k*-related-samples test** is required for situations where (1) the grouping factor has more than two levels, (2) observations or subjects are matched or the same subject is measured more than once, and (3) the data are

> Exhibit 18-17 Summary Tables for Repeated-Measures ANOVA*

Model Summary†

Source	d.f.	Sum of Squares	Mean Square	*F* Value	*p* Value
Airline	2	35527.550	17763.775	67.199	0.0001
Subject (group)	57	15067.650	264.345		
Ratings	1	625.633	625.633	14.318	0.0004
Ratings by air	2	2061.717	1030.858	23.592	0.0001
Ratings by subj	57	2490.650	43.696		

Means Table Ratings by Airline

	Count	Mean	Std. Dev.	Std. Error
Rating 1, Delta	20	38.950	14.006	3.132
Rating 1, Lufthansa	20	58.900	15.089	3.374
Rating 1, KLM	20	72.900	13.902	3.108
Rating 2, Delta	20	32.400	8.268	1.849
Rating 2, Lufthansa	20	72.250	10.572	2.364
Rating 2, KLM	20	79.800	11.265	2.519

Means Table Effect: Ratings

	Count	Mean	Std. Dev.	Std. Error
Rating 1	60	56.917	19.902	2.569
Rating 2	60	61.483	23.208	2.996

*All data are hypothetical.

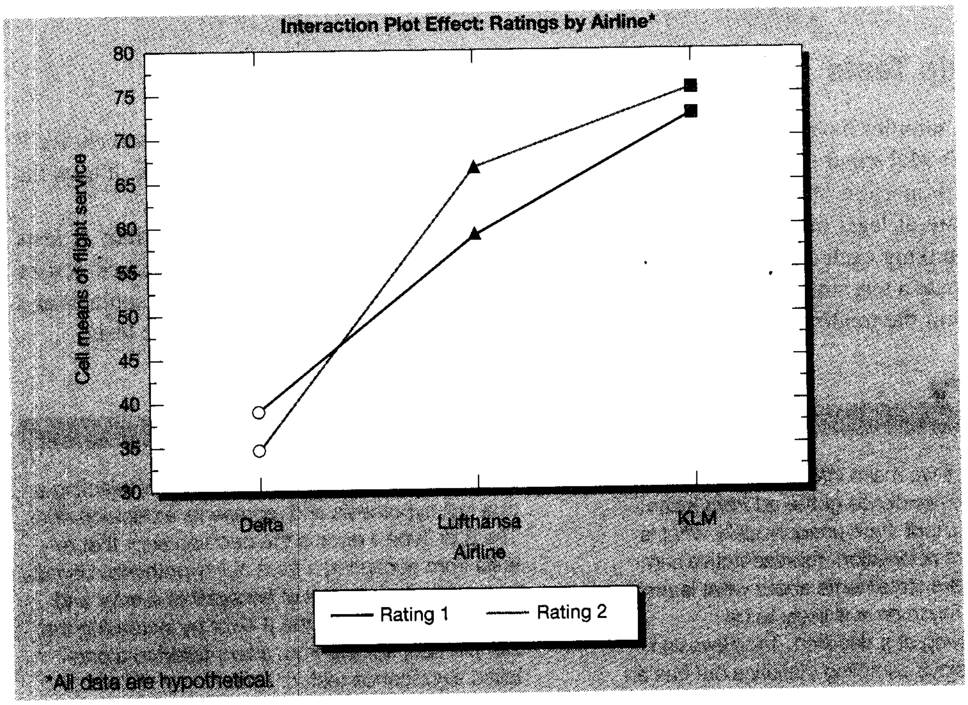
†Dependent: flight service ratings 1 and 2.

at least interval. In test marketing experiments or ex post facto designs with k samples, it is often necessary to measure subjects several times. These repeated measurements are called **trials.** For example, multiple measurements are taken in studies of stock prices, products evaluated by reliability, inventory, sales, and measures of product performance. Hypotheses for these situations may be tested with a univariate or multivariate general linear model. The latter is beyond the scope of this discussion.

The repeated-measures ANOVA is a special type of n-way analysis of variance. In this design, the repeated measures of each subject are related just as they are in the related t-test when only two measures are present. In this sense, each subject serves as its own control requiring a within-subjects variance effect to be assessed differently than the between-groups variance in a factor like airline or seat selection. The effects of the correlated measures are removed before calculation of the F ratio.

This model is an appropriate solution for the data presented in Exhibit 18-11. You will remember that the one-way and two-way examples considered only the first rating of in-flight service. Assume a second rating was obtained after one week by reinterviewing the same respondents. We now have two trials for the dependent variable, and we are interested in the same general question as with the one-way ANOVA, with the addition of how the passage of time affects perceptions of in-flight service.

> Exhibit 18-18 Repeated-Measures ANOVA Plot



Following the testing procedure, we state:

1. *Null hypotheses.*
   (1) Airline: H_0: $\mu_{A1} = \mu_{A2} = \mu_{A3}$
   (2) Ratings: H_0: $\mu_{R1} = \mu_{R2}$
   (3) Ratings × airline: H_0: $(\mu_{R2A1} - \mu_{R2A2} - \mu_{R2A3}) = (\mu_{R1A1} - \mu_{R1A2} - \mu_{R1A3})$
   For the alternative hypotheses, we will generalize to the statement that not all the groups have equal means for each of the three hypotheses.
2. *Statistical test.* The F test for repeated measures is chosen because we have related trials on the dependent variable for k samples, accept the assumptions of analysis of variance, and have interval data.
3. *Significance level.* Let $\alpha = .05$ and d.f. = [airline (2, 57), ratings (1, 57), ratings by airline (2, 57)].
4. *Calculated values.* See summary in Exhibit 18-17.
5. *Critical test value.* Enter Appendix C, Exhibit C-8, with d.f. (2, 57), $\alpha = .05$ and (1, 57), $\alpha = .05$. The critical values are 3.16 (2, 57) and 4.01 (1, 57).
6. *Interpretation.* The statistical results are grounds for rejecting all three null hypotheses and concluding there are statistically significant differences between means in all three instances. We conclude the perceptions of in-flight service were significantly affected by the different airlines, the interval between the two measures had a significant effect on the ratings, and the measures' time interval and the airlines interacted to a significant degree.

The ANOVA summary table in Exhibit 18-17 records the results of the tests. A means table provides the means and standard deviations for all combinations of ratings by airline. A second table of means reports the differences between flight service ratings 1 and 2. In Exhibit 18-18, there is an interaction plot for these data. Note that the second in-flight service rating was improved in two of the three groups after one week, but for the third carrier, there was a decrease in favorable response. The intersecting lines in the interaction plot reflect this finding.

Nonparametric Tests

When the k related samples have been measured on a nominal scale, the Cochran Q test is a good choice.[14] This test extends the McNemar test, discussed earlier, for studies having more than two samples. It tests the hypothesis that the proportion of cases in a category is equal for several related categories.

When the data are at least ordinal, the Friedman two-way analysis of variance is appropriate. It tests matched samples, ranking each case and calculating the mean rank for each variable across all cases. It uses these ranks to compute a test statistic. The product is a two-way table where the rows represent subjects and the columns represent the treatment conditions.[15] See Appendix B for additional nonparametric tests.

>summary

1 In classical statistics we make inferences about a population based on evidence gathered from a sample. Although we cannot state unequivocally what is true about the entire population, representative samples allow us to make statements about what is probably true and how much error is likely to be encountered in arriving at a decision. The Bayesian approach also employs sampling statistics but has an additional element of prior information to improve the decision maker's judgment.

2 A difference between two or more sets of data is statistically significant if it actually occurs in a population. To have a statistically significant finding based on sampling evidence, we must be able to calculate the probability that some observed difference is large enough that there is little chance it could result from random sampling. Probability is the foundation for deciding on the acceptability of the null hypothesis, and sampling statistics facilitate acquiring the estimates.

3 Hypothesis testing can be viewed as a six-step procedure:

 a Establish a null hypothesis as well as the alternative hypothesis. It is a one-tailed test of significance if the alternative hypothesis states the direction of difference. If no direction of difference is given, it is a two-tailed test.

 b Choose the statistical test on the basis of the assumption about the population distribution and measurement level. The form of the data can also be a factor. In light of these considerations, one typically chooses the test that has the greatest power efficiency or ability to reduce decision errors.

 c Select the desired level of confidence. While $\alpha = .05$ is the most frequently used level, many others are also used. The α is the significance level that we desire and is typically set in advance of the study. Alpha or Type I error is the risk of rejecting a true null hypothesis and represents a decision error. The β or Type II error is the decision error that results from accepting a false null hypothesis. Usually, one determines a level of acceptable α error and then seeks to reduce the β error by increasing the sample size, shifting from a two-tailed to a one-tailed significance test, or both.

 d Compute the actual test value of the data.

 e Obtain the critical test value, usually by referring to a table for the appropriate type of distribution.

 f Interpret the result by comparing the actual test value with the critical test value.

4 Parametric and nonparametric tests are applicable under the various conditions described in the chapter. They were also summarized in Exhibit 18-6. Parametric tests operate with interval and ratio data and are preferred when their assumptions can be met. Diagnostic tools examine the data for violations of those assumptions. Nonparametric tests do not require stringent assumptions about population distributions and are useful with less powerful nominal and ordinal measures.

5 In selecting a significance test, one needs to know, at a minimum, the number of samples, their independence or relatedness, and the measurement level of the data. Statistical tests emphasized in the chapter were the Z and t-tests, analysis of variance, and chi-square. The Z and t-tests may be used to test for the difference between two means. The t-test is chosen when the sample size is small. Variations on the t-test are used for both independent and related samples.

One-way analysis of variance compares the means of several groups. It has a single grouping variable, called a factor, and a continuous dependent variable. Analysis of variance (ANOVA) partitions the total variation among scores into between-groups (treatment) and within-groups (error) variance. The *F* ratio, the test statistic, determines if the differences are large enough to reject the null hypothesis. ANOVA may be extended to two-way, *n*-way, repeated-measures, and multivariate applications.

Chi-square is a nonparametric statistic that is used frequently for cross-tabulation or contingency tables. Its applications include testing for differences between proportions in populations and testing for independence. Corrections for chi-square were discussed.

>keyterms

>discussionquestions

Terms in Review

1 Distinguish between the following:
   a Parametric tests and nonparametric tests.
   b Type I error and Type II error.
   c Null hypothesis and alternative hypothesis.
   d Acceptance region and rejection region.
   e One-tailed tests and two-tailed tests.
   f Type II error and the power of the test.

2 Summarize the steps of hypothesis testing. What is the virtue of this procedure?

3 In analysis of variance, what is the purpose of the mean square between and the mean square within? If the null hypothesis is accepted, what do these quantities look like?

4 Describe the assumptions for ANOVA, and explain how they may be diagnosed.

Making Research Decisions

5 Suggest situations where the researcher should be more concerned with Type II error than with Type I error.
   a How can the probability of a Type I error be reduced? A Type II error?
   b How does practical significance differ from statistical significance?
   c Suppose you interview all the members of the freshman and senior classes and find that 65 percent of the freshmen and 62 percent of the seniors favor a proposal to send Help Centers offshore. Is this difference significant?

6 What hypothesis testing procedure would you use in the following situations?
   a A test classifies applicants as accepted or rejected. On the basis of data on 200 applicants, we test the hypothesis that ad placement success is not related to gender.

b A company manufactures and markets automobiles in two different countries. We want to know if the gas mileage is the same for vehicles from both facilities. There are samples of 45 units from each facility.

c A company has three categories of marketing analysts: (1) with professional qualifications but without work experience, (2) with professional qualifications and with work experience, and (3) without professional qualifications but with work experience. A study exists that measures each analyst's motivation level (classified as high, normal, and low). A hypothesis of no relation between analyst category and motivation is to be tested.

d A company has 24 salespersons. The test must evaluate whether their sales performance is unchanged or has improved after a training program.

e A company has to evaluate whether it should attribute increased sales to product quality, advertising, or an interaction of product quality and advertising.

7 You conduct a survey of a sample of 25 members of this year's graduating marketing students and find that the average GPA is 3.2. The standard deviation of the sample is 0.4. Over the last 10 years, the average GPA has been 3.0. Is the GPA of this year's students significantly different from the long-run average? At what alpha level would it be significant?

8 You are curious about whether the professors and students at your school are of different political persuasions, so you take a sample of 20 professors and 20 students drawn randomly from each population. You find that 10 professors say they are conservative and 6 students say they are conservative. Is this a statistically significant difference?

9 You contact a random sample of 36 graduates of Western University and learn that their starting salaries averaged $28,000 last year. You then contact a random sample of 40 graduates from Eastern University and find that their average starting salary was $28,800. In each case, the standard deviation of the sample was $1,000.

a Test the null hypothesis that there is no difference between average salaries received by the graduates of the two schools.

b What assumptions are necessary for this test?

10 A random sample of students is interviewed to determine if there is an association between class and attitude toward corporations. With the following results, test the hypothesis that there is no difference among students on this attitude.

	Favorable	**Neutral**	**Unfavorable**
Freshmen	100	50	70
Sophomores	80	60	70
Juniors	50	50	80
Seniors	40	60	90

11 You do a survey of marketing students and liberal arts school students to find out how many times a week they read a daily newspaper. In each case, you interview 100 students. You find the following:

$\bar{X}_m = 4.5$ times per week

$S_m = 1.5$

$\bar{X}_{la} = 5.6$ times per week

$S_{la} = 2.0$

Test the hypothesis that there is no significant difference between these two samples.

12 One-Koat Paint Company has developed a new type of porch paint that it hopes will be the most durable on the market. The R&D group tests the new product against the two leading competing products by using a machine that scrubs until it wears through the coating. One-Koat runs five trials with each product and secures the following results (in thousands of scrubs):

Trial	**One-Koat**	**Competitor A**	**Competitor B**
1	37	34	24
2	30	19	25
3	34	22	23
4	28	31	20
5	29	27	20

Test the hypothesis that there are no differences between the means of these products ($\alpha = .05$).

13 A computer manufacturer is introducing a new product specifically targeted at the home market and wishes to compare the effectiveness of three sales strategies: computer stores, home electronics stores, and department stores. Numbers of sales by 15 salespeople are recorded below:

Electronics store: 5, 4, 3, 3, 3

Department store: 9, 7, 8, 6, 5

Computer store: 7, 4, 8, 4, 3

a Test the hypothesis that there is no difference between the means of the retailers ($\alpha = .05$).

b Select a multiple comparison test, if necessary, to determine which groups differ in mean sales ($\alpha = .05$).

14 Following the collapse of the dot-coms in 2000, the *Fortune* magazine annual list of the 40 richest self-made Americans under the age of 40 revealed some interesting changes.

Rank	Name	Company	Net Worth ($, millions)
1	Michael Dell	Dell Computer	16,300
2	Pierre Omidyar	eBay	4,390
3	Jeff Skoll	eBay	2,630
4	Ted Waitt	Gateway, Inc.	1,870
5	Jeff Bezos	Amazon.Com, Inc.	1,230
6	Vinny Smith	Quest Software	780
7	David Filo	Yahoo	730
8	Jerry Yang	Yahoo	721
9	Rob Glaser	RealNetworks	635
10	Dan Snyder	Washington Redskins	604
11	Greg Reyes	Brocade Communications Systems	518
12	Jen-Hsun Huang	Nvidia	507
13	Michael Jordan	Washington Wizards	398
14	Joe Liemandt	Trilogy Software	390
15	Jeanette Symons	Zhone Technologies	374
16	Pantas Sutardja	Marvell Technology	363
17	John Schnatter	Papa John's International	293
18	Sanjay Kumar	Computer Associate International, Inc.	270
19	Tom Cruise	Cruise/Wagner Productions	251
20	Percy Miller (Master P)	No Limit	249
21	James T. Demetriades	SeeBeyond Technology	239
22	Sean Combs (P. Diddy)	Bad Boy Entertainment	231
23	Jerry Greenberg	Sapient	225
24	J. Stuart Moore	Sapient	224
25	Sudhakar Ravi	SonicWALL	219
26	Sreekanth Ravi	SonicWALL	219
27	David Hitz	Network Appliance	202
28	John L. MacFarlane	Openwave Systems Inc.	198
29	Jeffrey Citron	Vonage	194
30	Raul Fernandez	Dimension Data North America	188
31	Eric Greenberg	Innovation Investments	187
32	Chris Klaus	Internet Security Systems	187
33	Anousheh Ansari	Sonus	180
34	Halsey Minor	12 Entrepreneuring	180
35	Michael Saylor	Microstrategy	180
36	Jim Carrey	Pit Bull Productions	171
37	Jonathan M. Rothberg	CuraGen	168
38	Marc Andreessen	Loudcloud	166
39	Nav Sooch	Silicon Laboratories	162
40	Tiger Woods	ETW	160

a Devise a grouping variable to classify the companies in the accompanying table (e.g., Internet, computers, celebrities).

b Using one-way analysis of variance, test the hypothesis that there is no difference in net worth among the groups.

15 A consumer testing firm is interested in testing two competing antivirus products for personal computers. It wants to know how many strains of virus will be removed. The data are:

Virus Removed by Anti-V?	Virus Removed by Q-Cure? Yes	No
Yes	45	33
No	58	20

Are Anti-V and Q-Cure equally effective ($\alpha = .05$)?

16 A researcher for an auto manufacturer is examining preferences for styling features on larger sedans. Buyers were classified as "first time" and "repeat," resulting in the following table.

	Preference: European Styling	Japanese Styling
Repeat	40	20
First time	8	32

a Test the hypothesis that buying characteristic is independent of styling preference ($\alpha = .05$).

b Should the statistic be adjusted?

Bringing Research to Life

17 If you were Sally, what test would you propose should be run on the City Center for Performing Arts project?

From Concept to Practice

18 Using the data in Exhibit 18-11 for the variables flight service rating 2 and airline (2, 3), test the hypothesis of no difference between means.

>wwwexercise

Find a study reported on the Web and identify its likely hypothesis. How would you test this hypothesis given the data that was collected. (Hint: The Henry Kaiser Family Foundation Web site is a great source of fully cited research reports: **http://www.kff.org.**)

>cases*

Inquiring Minds Want to Know—NOW!

NCRCC: Teeing Up a New Strategic Direction

Mastering Teacher Leadership

Yahoo!: ***Consumer Direct*** **Marries Purchase Metrics to Banner Ads**

* All cases appear on the text CD; you will find abstracts of these cases in the Case Abstracts section of this text.

>chapter 19

Measures of Association

"To truly understand consumers' motives and actions, you must determine relationships between what they think and feel and what they actually do."

David Singleton, vice president of insights, Zyman Marketing Group

>**learning**objectives

After reading this chapter, you should understand . . .

1 How correlation analysis may be applied to study relationships between two or more variables.

2 The uses, requirements, and interpretation of the product moment correlation coefficient.

3 How predictions are made with regression analysis using the method of least squares to minimize errors in drawing a line of best fit.

4 How to test regression models for linearity and whether the equation is effective in fitting the data.

5 The nonparametric measures of association and the alternatives they offer when key assumptions and requirements for parametric techniques cannot be met.

Sally arrives for an analysis meeting with Jason and finds a round, bald little man sitting at Jason's desk, studying the screen of a laptop computer, stroking his gray beard and smiling broadly.

"Sally," says Jason, "meet Jack Adams, rising political consultant."

Jack, who seems to be caressing his laptop, grins broadly. "Hello, Sally," says Jack. "I wanted Jason to know this little computer has made me the marketing kingpin of the Boca Beach political scene."

"Jack sold his painting business on Long Island to his three boys and moved to Boca Beach after his wife passed away," offers Jason in explanation.

"For three months I played golf in the morning and sat by the pool and played cards in the afternoon. For three months, seven days, I did this. I was going crazy. Then my next-door neighbor Marty died and his wife gave me his MindWriter."

"Jason came through Boca Beach and stopped for a visit. He downloaded a statistical program, free from the Internet. I must say, statistics in college never generated as much excitement as they have recently," grins Jack. "We had a wise guy, Sandy Plover, a former electrical contractor in Jersey, who got himself into local politics. Being a natural-born troublemaker, he waited for his chance to agitate. As it happens, the sheriff released data to the newspaper that the incidence of arrests resulting from police calls to Oceanside—the richer of the two neighborhoods where the sheriff happens to live—is higher than in Gladeside."

Jack types the following:

Research hypothesis: Gladeside residents get special treatment when it comes to solving crimes and thus live in a safer environment due to their higher incomes and greater political power.

Null hypothesis: Gladeside and Oceanside receive the same attention from the police.

	Gladeside	Oceanside
Police calls without arrest	46	40
Police calls resulting in arrest	4	10
Total calls	50	50

"I doubt that Sandy would have paid attention, except that in both neighborhoods the total number of police calls happened to be 50, which made it easy for him to see that in Oceanside the rate of arrests was twice that in Gladeside."

"Actually," says Sally, "I'm surprised there would be any police calls in such an upscale community."

"We are old," says Jack, "but not dead."

"In any case, Sandy's finely honed political instincts told him he was going nowhere by trying to turn the community against the sheriff. It would be much, much better to turn voters of Oceanside against those in Gladeside. So he complained about the disparate impact of arrests. While both the communities are roughly the same size, in Oceanside folks are mostly from Brooklyn, and in Gladeside folks come from the Bronx."

"But the ethics . . ."

". . . meant nothing to Sandy. He told me, 'I think I'm gonna kick some butt and make a name for myself down here.'"

"The trouble with the police calls as an issue is that sheriffs' offices nowadays are well staffed with

>cont'd

statistically educated analysts, who know very well how to rebut oddball claims," interrupts Jason.

"While I personally miss the old days, I punched the numbers into this MindWriter here to double-check the stats. I did the obvious first, just what I supposed a police analyst would have done. I ran a cross-tabulation and a chi-square test of the hypothesis that the arrests in Oceanside were disproportionate to those in Gladeside."

"To an untrained observer it would appear that they are," contributes Sally as she peers over Jack's shoulder. "But then I'm not so easily hoodwinked."

"Good for you, Sally!" exclaims Jack. "What you have here is the 'eyeball' fallacy, as my dear old professor called it almost 50 years ago. As I explained to Sandy, 100 police encounters resulting in a few arrests is nothing, nada, not a large enough sample to trust a quick peek and a leap to a conclusion. You run it through the computer, and, sure enough, although the ratios seem to be out of whack, the difference is not statistically significant. You cannot support 'disparate impact.' No way."

"Granting that 10 arrests per 50 is bigger than 4 per 50," observes Jason, "Jack saw that a statistician would say that it is not disproportionate enough to convince a scientist that the police were acting differently in the two neighborhoods. A statistician would say, wait and see, let the story unfold, collect a bigger sample."

"How did Sandy accept your explanation, Jack?"

"He was ready to shoot the messenger, very much bothered, at first, that I would not support his political strategy. But I was pretty sure the sheriff would come roaring back with a statistical analysis to throw cold water over Sandy."

"Did you bring him around?"

"That jerk, come around? Never. He ran to the papers, and spilled his numbers and accusations in a letter to the editor that was printed on a Monday, and Tuesday the sheriff came back with his experts and made Sandy look like a fool—on page one, if you can believe it. So Sandy was washed up, but he mentioned to the reporter that I had provided the same interpretation before he went to the paper, so I now have a new career: resident political genius. What I do is look at the opponents' polling results and deny their validity for the newspapers and TV. If the opposing party is ahead by a few poll points, I scoff at the thinness of the margin. If their lead is wide, I belittle the size of the sample and intimate that any statistician would see through them."

"Jack is colorful, amusing, and good-natured in debunking his opponents' polls, and the newspaper writers have never challenged him to substantiate his claims or interpretations," contributes Jason. "What he learned from me is that statistics is so complicated, and scares so many people, that you can claim or deny anything. And he is usually right to debunk the polls, since for a preelection political poll to be taken seriously there has to be a large enough sample to produce significant results. And there has to be a big enough spread between winners and losers to protect against a last-minute shift in voter sentiment. In the small, closely contested voting precincts of condominium politics, hardly any poll can meet two such stringent criteria."

"So now I sit in my condo's clubhouse and people come over and want to know what I think about the Middle East, campaign reform, everything." Jack, rising, extends his hand to Sally. "I can tell you have things to do, so I'll move along. It was nice meeting you, Sally."

Sally watches Jack Adams give Jason a bear hug and then walk out the door.

"So, Sally, what did you think of Jack's knowledge of statistics?"

> Introduction

In the previous chapter, we emphasized testing hypotheses of difference. However, management questions frequently revolve around the study of relationships between two or more variables. Then, a *relational hypothesis* is necessary. In the research question "Are U.S. kitchen appliances perceived by American consumers to be of better quality than foreign kitchen appliances?" the nature of the relationship between the two variables ("country of origin" and "perceived quality") is not specified. The implication, nonetheless, is that one variable is responsible for the other. A correct relational hypothesis for this question would state that the variables occur together in some specified manner without implying that one causes the other.

< You might want to review our discussion of relational hypotheses in Chapter 2.

Various objectives are served with correlation analysis. The strength, direction, shape, and other features of the relationship may be discovered. Or tactical and strategic questions may be answered by predicting the values of one variable from those of another. Let's look at some typical management questions:

- In the mail-order business, excessive catalog costs quickly squeeze margins. Many mailings fail to reach receptive or active buyers. What is the relationship between mailings that delete inactive customers and the improvement in profit margins?
- Medium-size companies often have difficulty attracting the cream of the MBA crop, and when they are successful, they have trouble retaining them. What is the relationship between the candidate's rank based on an executive interview and the rank obtained from testing or managerial assessment?
- Cigarette company marketing allocations shifted a few years ago as a result of multistate settlements eliminating outdoor and transit advertising. More recently, advertising in magazines with large youth readerships came under scrutiny. During a given period, what is the relationship between point-of-sale expenditures and net profits?
- Aggressive U.S. high-tech companies have advertised heavily in the European chip market, and their sales have grown 20 percent over sales of the three largest European firms. Can we predict next year's sales based on present advertising?

All these questions may be evaluated by means of measures of association. And all call for different techniques based on the level at which the variables were measured or the intent of the question. The first three use nominal, ordinal, and interval data, respectively. The last one is answered through simple linear regression.

With correlation, one calculates an index to measure the nature of the relationship between variables. With regression, an equation is developed to predict the values of a dependent variable. Both are affected by the assumptions of measurement level and the distributions that underlie the data.

Exhibit 19-1 lists some common measures and their uses. The chapter follows the progression of the exhibit, first covering bivariate linear correlation, then examining simple regression, and concluding with nonparametric measures of association. Exploration of data through visual inspection and diagnostic evaluation of assumptions continues to be emphasized.

> Bivariate Correlation Analysis

Bivariate correlation analysis differs from nonparametric measures of association and regression analysis in two important ways. First, parametric correlation requires two continuous variables measured on an interval or ratio scale. Second, the coefficient does not distinguish between independent and dependent variables. It treats the variables symmetrically since the coefficient r_{xy} has the same interpretation as r_{yx}.

> **Exhibit 19-1** Commonly Used Measures of Association

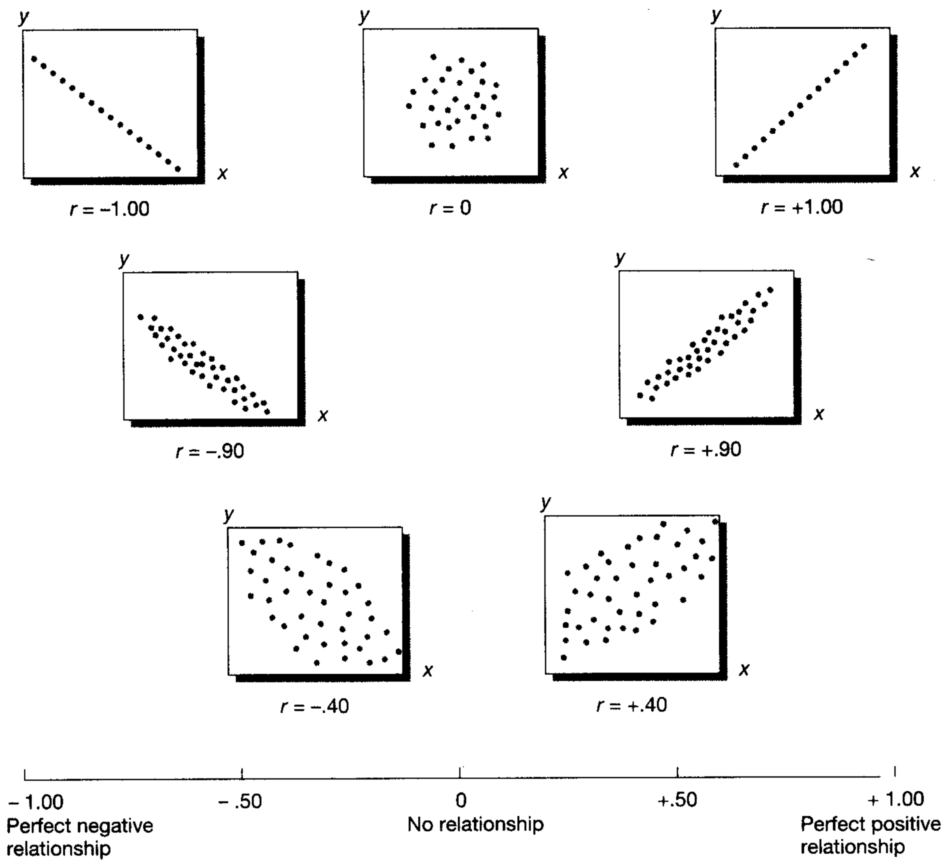
Measurement	Coefficient	Comments and Uses
	Pearson (product moment) correlation coefficient	For continuous linearly related variables.
	Correlation ratio (eta)	For nonlinear data or relating a main effect to a continuous dependent variable.
	Biserial	One continuous and one dichotomous variable with an underlying normal distribution.
	Partial correlation	Three variables; relating two with the third's effect taken out.
	Multiple correlation	Three variables; relating one variable with two others.
	Bivariate linear regression	Predicting one variable from another's scores.
	Gamma	Based on concordant-discordant pairs: (*P* – *Q*); proportional reduction in error (PRE) interpretation.
	Kendall's tau *b*	*P* – *Q* based; adjustment for tied ranks.
	Kendall's tau *c*	*P* – *Q* based; adjustment for table dimensions.
	Somers's *d*	*P* – *Q* based; asymmetrical extension of gamma.
	Spearman's rho	Product moment correlation for ranked data.
	Phi	Chi-square (CS) based for 2 × 2 tables.
	Cramer's *V*	CS based; adjustment when one table dimension > 2.
	Contingency coefficient *C*	CS based; flexible data and distribution assumptions.
	Lambda	PRE-based interpretation.
	Goodman & Kruskal's tau	PRE based with table marginals emphasis.
	Uncertainty coefficient	Useful for multidimensional tables.
	Kappa	Agreement measure.

Pearson's Product Moment Coefficient *r*

The **Pearson** (product moment) **correlation coefficient** varies over a range of +1 through 0 to −1. The designation *r* symbolizes the coefficient's estimate of linear association based on sampling data. The coefficient ρ represents the population correlation.

Correlation coefficients reveal the magnitude and direction of relationships. The ***magnitude*** is the degree to which variables move in unison or opposition. The size of a correlation of +.40 is the same as one of −.40. The sign says nothing about size. The degree of correlation is modest. The coefficient's sign signifies the *direction* of the relationship. Direction tells us whether large values on one variable are associated with large values on the other (and small values with small values). When the values correspond in this way, the two variables have a positive relationship: As one increases, the other also increases. Family income, for exam-

> **Exhibit 19-2** Scatterplots of Correlations between Two Variables



ple, is positively related to household food expenditures. As income increases, food expenditures increase. Other variables are inversely related. Large values on the first variable are associated with small values on the second (and vice versa). The prices of products and services are inversely related to their scarcity. In general, as products decrease in available quantity, their prices rise. The absence of a relationship is expressed by a coefficient of approximately zero.

Scatterplots for Exploring Relationships

Scatterplots are essential for understanding the relationships between variables. They provide a means for visual inspection of data that a list of values for two variables cannot. Both the direction and the shape of a relationship are conveyed in a plot. With a little practice, the magnitude of the relationship can be seen.

Exhibit 19-2 contains a series of scatterplots that depict some relationships across the range *r*. The three plots on the left side of the figure have their points sloping from the upper left to the lower right of each *x*-*y* plot.[1] They represent different magnitudes of negative relationships. On the right side of the figure, the three plots have opposite patterns and show positive relationships.

When stronger relationships are apparent (for example, the ±.90 correlations), the points cluster close to an imaginary straight line passing through the data. The weaker relationships (±.40) depict a more diffuse data cloud with points spread farther from the line.

> **Exhibit 19-3** Four Data Sets with the Same Summary Statistics

S	X_1	Y_1	X_2	Y_2	X_3	Y_3	X_4	Y_4
1	10	8.04	10	9.14	10	7.46	8	6.58
2	8	6.95	8	8.14	8	6.77	8	5.76
3	13	7.58	13	8.74	13	12.74	8	7.71
4	9	8.81	9	8.77	9	7.11	8	8.84
5	11	8.33	11	9.26	11	7.81	8	8.47
6	14	9.96	14	8.10	14	8.84	8	7.04
7	6	7.24	6	6.13	6	6.08	8	5.25
8	4	4.26	4	3.10	4	5.39	19	12.50
9	12	10.84	12	9.13	12	8.15	8	5.56
10	7	4.82	7	7.26	7	6.42	8	7.91
11	5	5.68	5	4.74	5	5.73	8	6.89

The shape of linear relationships is characterized by a straight line, whereas nonlinear relationships have curvilinear, parabolic, and compound curves representing their shapes. Pearson's r measures relationships in variables that are linearly related. It cannot distinguish linear from nonlinear data. Summary statistics alone do not reveal the appropriateness of the data for the model, which is why inspecting the data is important.

The need for data visualization is illustrated with four small data sets possessing identical summary statistics but displaying strikingly different patterns.[2] Exhibit 19-3 contains these data, and Exhibit 19-4 plots them. In Plot 1 of the figure, the variables are positively related. Their points follow a superimposed straight line through the data. This example is well suited to correlation analysis. In Plot 2, the data are curvilinear in relation to the line, and r is an inappropriate measure of their relationship. Plot 3 shows the presence of an influential point that changed a coefficient that would have otherwise been a perfect $+1.0$. The last plot displays constant values of x (similar to what you might find in an animal or quality control experiment). One leverage point establishes the fitted line for these data.

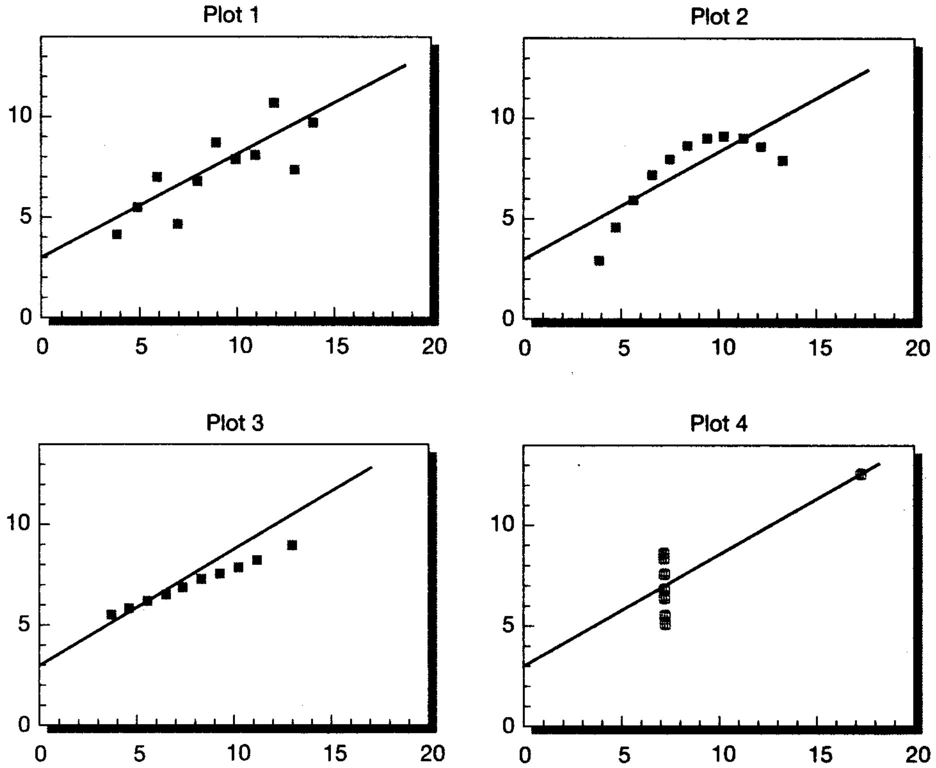
We will return to these concepts and the process of drawing the line when we discuss regression. For now, comparing Plots 2 through 4 with Plot 1 suggests the importance of visually inspecting correlation data for underlying patterns to ensure linearity. Careful analysts make scatterplots an integral part of the inspection and exploration of their data. Although small samples may be plotted by hand, statistical software packages save time and offer a variety of plotting procedures.

The Assumptions of *r*

Like other parametric techniques, correlation analysis makes certain assumptions about the data. Many of these assumptions are necessary to test hypotheses about the coefficient.

The first requirement for r is **linearity.** All of the examples in Exhibit 19-2 with the exception of $r = 0$ illustrate a relationship between variables that can be described by a straight line passing through the data cloud.

> Exhibit 19-4 Different Scatterplots for the Same Summary Statistics



When $r = 0$, no pattern is evident that could be described with a single line. Parenthetically, it is also possible to find coefficients of 0 where the variables are highly related but in a nonlinear form. As we have seen, plots make such findings evident.

The second assumption for correlation is a **bivariate normal distribution**—that is, the data are from a random sample of a population where the two variables are normally distributed in a joint manner.

Often these assumptions or the required measurement level cannot be met. Then the analyst should select a nonlinear or nonparametric measure of association, many of which are described later in this chapter.

Computation and Testing of *r*

The formula for calculating Pearson's r is

$$r = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{(n - 1)s_x s_y} \quad (1)$$

where

n = the number of pairs of cases

s_x, s_y = the standard deviations for X and Y

Alternatively,

$$r = \frac{\Sigma xy}{\sqrt{(\Sigma x^2)(\Sigma y^2)}} \quad (2)$$

> **Exhibit 19-5** Computation of Pearson's Product Moment Correlation

(1) x^2	(2) Net Profits (S, millions) X y^2	(3) Cash Flow (S, millions) Y	(4) Deviations from Means Corporation	(5)	(6)	(7) $(X\ \ \bar{X})x$	(8) $(Y\ \ \bar{Y})y$
1	82.6	126.5	−93.84	−178.64	16763.58	8805.95	31912.25
2	89.0	191.2	−87.44	−113.94	9962.91	7645.75	12982.32
3	176.0	267.0	−0.44	−38.14	16.78	0.19	1454.66
4	82.3	137.1	−94.14	−168.04	15819.29	8862.34	28237.44
5	413.5	806.8	237.06	501.66	118923.52	56197.44	251602.56
6	18.1	35.2	158.34	−269.94	42742.3	25071.56	72867.60
7	337.3	425.5	160.86	120.36	19361.11	25875.94	14486.53
8	145.8	380.0	−30.64	74.86	−2293.71	938.81	5604.02
9	172.6	326.6	−3.84	21.36	82.02	14.75	456.25
10	247.2	355.5	70.76	50.36	3563.47	5006.98	2536.13
	$\bar{X} = 176.44$	$\bar{Y} = 305.14$			$\Sigma xy = 224777.23$		
	$s_x = 216.59$	$s_y = 124.01$				$\Sigma x^2 = 138419.71$	
							$\Sigma y^2 = 422139.76$

since

$$s_x = \sqrt{\frac{\Sigma x^2}{N}} \qquad s_y = \sqrt{\frac{\Sigma y^2}{N}}$$

If the numerator of equation (2) is divided by n, we have the *covariance*, the amount of deviation that the X and Y distributions have in common. With a positive covariance, the variables move in unison; with a negative one, they move in opposition. When the covariance is 0, there is no relationship. The denominator for equation (2) represents the maximum potential variation that the two distributions share. Thus, correlation may be thought of as a ratio.

Exhibit 19-5 contains a random subsample of 10 firms of the Forbes 500 sample. The variables chosen to illustrate the computation of r are cash flow and net profits. Beneath each variable is its mean and standard deviation. In columns 4 and 5 we obtain the deviations of the X and Y values from their means, and in column 6 we find the product. Columns 7 and 8 are the squared deviation scores.

Substituting into the formula, we get

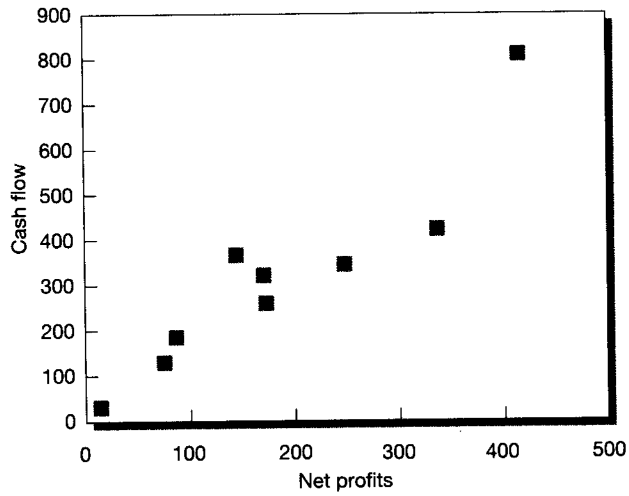
$$r = \frac{224777.23}{\sqrt{138419.71} * \sqrt{422139.76}} = .9298$$

In this subsample, net profits and cash flow are positively related and have a very high coefficient. As net profits increase, cash flow increases; the opposite is also true. Linearity of the variables may be examined with a scatterplot such as the one shown in Exhibit 19-6. The data points fall along a straight line.

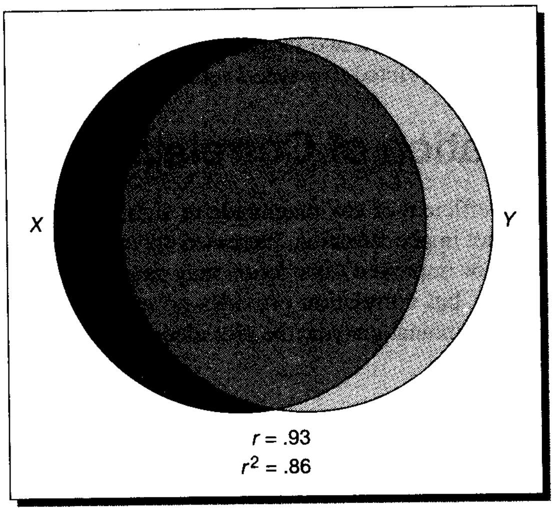
Common Variance as an Explanation

The amount of common variance in X (net profits) and Y (cash flow) may be summarized by r^2, the **coefficient of determination.** As Exhibit 19-7 shows, the overlap between the two variables is the proportion of their common or shared variance.

> Exhibit 19-6 Plot of Forbes 500 Net Profits with Cash Flow



> Exhibit 19-7 Diagram of Common Variance



The area of overlap represents the percentage of the total relationship accounted for by one variable or the other. So 86 percent of the variance in X is explained by Y, and vice versa.

Testing the Significance of *r*

Is the coefficient representing the relationship between net profits and cash flow real, or does it occur by chance? This question tries to discover whether our r is a chance deviation from a population p of zero. In other situations, the researcher may wish to know if significant differences exist between two or more rs. In either case, r's significance should be checked before r is used in other calculations or comparisons. For this test, we must have independent random samples from a bivariate normal distribution. Then the Z or t-test may be used for the null hypothesis, $p = 0$.

The formula for small samples is

$$t = \frac{r}{\sqrt{\frac{1 - r^2}{n - 2}}}$$

where

$r = .93$

$n = 10$

Substituting into the equation, we calculate t:

$$t = \frac{.93}{\sqrt{\frac{1 - .86}{8}}} = 7.03$$

With $n - 2$ degrees of freedom, the statistical program calculates the value of t (7.03) at a probability less than .005 for the one-tailed alternative, H_A: $p > 0$. We reject the hypothesis that there is no linear relationship between net profits and cash flow in the population. The above statistic is appropriate when the null hypothesis states a correlation of 0. It should be used only for a one-tailed test.[3] However, it is often difficult to know in advance whether the variables are positively or negatively related, particularly when a computer removes our contact with the raw data. Software programs produce two-tailed tests for this eventuality. The observed significance level for a one-tailed test is half of the printed two-tailed version in most programs.

< You might want to review the nature of causation in Chapter 6.

Interpretation of Correlations

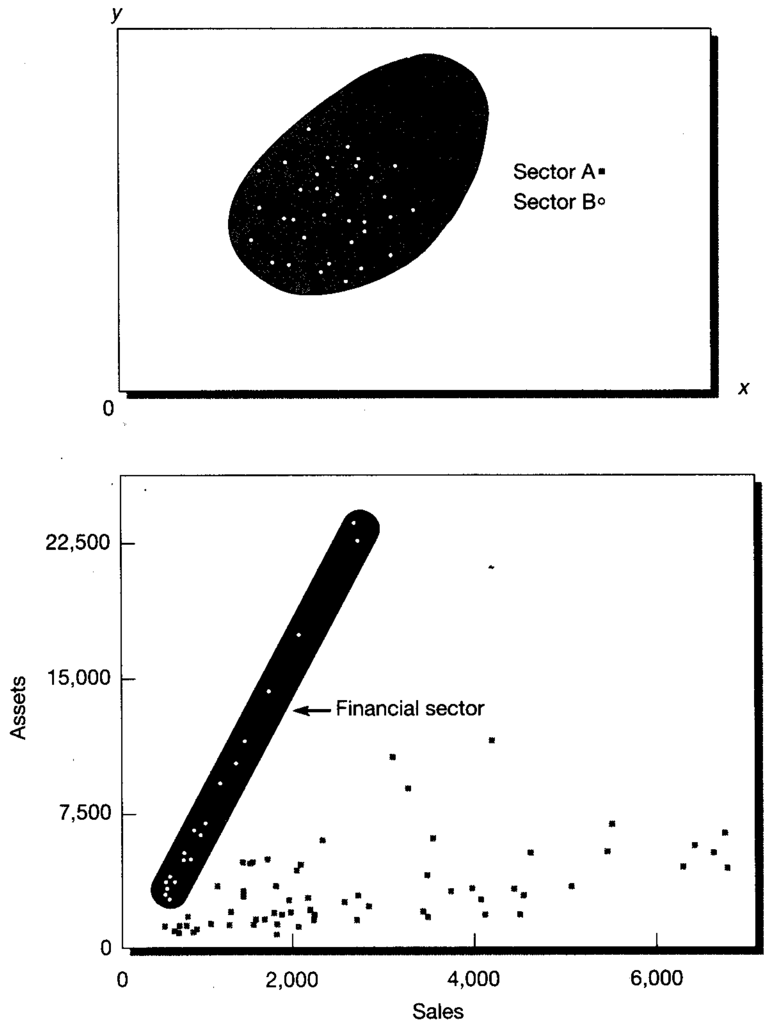
A correlation coefficient of any magnitude or sign, whatever its statistical significance, does not imply causation. Increased net profits may cause an increase in market value, or improved satisfaction may cause improved performance in certain situations, but correlation provides no evidence of cause and effect. Several alternate explanations may be provided for correlation results:

- *X* causes *Y*.
- *Y* causes *X*.
- *X* and *Y* are activated by one or more other variables.
- *X* and *Y* influence each other reciprocally.

Ex post facto studies seldom possess sufficiently powerful designs to demonstrate which of these conditions could be true. By controlling variables under an experimental design, we may obtain more rigorous evidence of causality.

Take care to avoid so-called **artifact correlations,** where distinct groups combine to give the impression of one. The upper panel of Exhibit 19-8 shows data from two business sectors. If all the data points for the *X* and *Y* variables are aggregated and a correlation is computed for a single group, a positive correlation results. Separate calculations for each sector (note that points for sector A form a circle, as do points for sector B) reveal *no* relationship between the *X* and *Y* variables. A second example, shown in the lower panel, contains a plot of data on assets and sales. We have enclosed and highlighted the data for the financial sector. This is shown as a narrow band enclosed by an ellipse. The companies in this sector score high on assets and low in sales—all are banks. When the data for banks are removed and treated separately, the correlation is nearly perfect (.99). When banks are returned to the sample and the correlation is recalcu-

> Exhibit 19-8 Artifact Correlations



lated, the overall relationship drops to the middle .80s. In short, data hidden or nested within an aggregated set may present a radically different picture.

Another issue affecting interpretation of coefficients concerns practical significance. Even when a coefficient is statistically significant, it must be practically meaningful. In many relationships, other factors combine to make the coefficient's meaning misleading. For example, in nature we expect rainfall and the height of reservoirs to be positively correlated. But in states where water management and flood control mechanisms are complex, an apparently simple relationship may not hold. Techniques like partial and multiple correlation or multiple regression are helpful in sorting out confounding effects.

With large samples, even exceedingly low coefficients can be statistically significant. This "significance" only reflects the likelihood of a linear relationship in the population. Should magnitudes less than .30 be reported when they are significant? It depends. We might consider the correlations between variables such as cash flow, sales, market value, or net profits to be interesting revelations of a particular phenomenon whether they were high, moderate, or low. The nature of the study, the characteristics of the sample, or other reasons will be determining factors. *A coefficient is not remarkable simply because it is statistically significant.*

>snapshot

Envirosell: Studies Reveal Left-Hand Retail

World retailers collect and subscribe to numerous data sources, but they need knowledge from the data to craft their merchandising, staffing, and promotion strategies, as well as their store designs. Retail giants (e.g., The Gap, Limited, Starbucks, Radio Shack, McDonald's) turn to consultant Paco Underhill when they want to know how consumers buy what they do and what barriers prevent or discourage buying. Underhill describes himself as a "commercial researcher, which means I am part scientist, part artist, and part entrepreneur." His company, Envirosell, has offices in the United States, Milan, Sidney, and São Paulo. Envirosell concentrates on the third segment of retail information, drawn from observation (segment 1 is register data, and segment 2 is communication studies). In an ABC News live e-chat, Underhill said, "The principal differences in 1st world shopping patterns are governed more by education and income than by ethnicity . . . but the Brits and Aussies [do] tend to walk as they drive. This sets up some very peculiar retail [shopping] patterns, because their walking patterns set up a left-hand dominance, whereas in the U.S. and much of the rest of the world, our walking patterns set up a right-hand dominance." How would you set up an observation study to verify Underhill's conclusions?

www.envirosell.com

If you were Gap and about to design a store to open in London, how would you design a study to verify Paco Underhill's conclusions about left-hand dominance?

By probing the evidence of direction, magnitude, statistical significance, and common variance together with the study's objectives and limitations, we reduce the chances of reporting trivial findings. Simultaneously, the communication of practical implications to the reader will be improved.

> Simple Linear Regression[4]

In the previous section, we focused on relationships between variables. The product moment correlation was found to represent an index of the magnitude of the relationship, the sign governed the direction, and r^2 explained the common variance. Relationships also serve as a basis for estimation and prediction.

When we take the observed values of X to estimate or predict corresponding Y values, the process is called **simple prediction.**[5] When more than one X variable is used, the outcome is a function of multiple predictors. Simple and multiple predictions are made with a technique called **regression analysis.**

The similarities and differences of correlation and regression are summarized in Exhibit 19-9. Their relatedness would suggest that beneath many correlation problems is a regression analysis that could provide further insight about the relationship of Y with X.

> Exhibit 19-9 Comparison of Bivariate Linear Correlation and Regression

	[illegible]	Regression
Measurement level	Interval or ratio scale	Interval or ratio scale
Nature of variables	Both continuous, linearly related	Both continuous, linearly related
X – Y relationship	X and Y are symmetric; $r_{xy} = r_{yx}$	Y is dependent, X is independent; regression of X on Y differs from Y on X
Correlation	The correlation of x and y produces an estimate of linear association based on sampling data	Correlation of Y – X is the same as the correlation between the predicted values of Y and observed values of Y
Coefficient of determination	Explains common variance of X and Y	Proportion of variability of Y explained by its least-squares regression on X

The Basic Model

A straight line is fundamentally the best way to model the relationship between two continuous variables. The bivariate linear regression may be expressed as

$$Y = \beta_0 + \beta_1 X_i$$

where the value of the dependent variable Y is a linear function of the corresponding value of the independent variable X_i in the ith observation. The slope and the Y intercept are known as **regression coefficients.** The **slope,** β_1, is the change in Y for a 1-unit change in X. It is sometimes called the "rise over run." This is defined by the formula

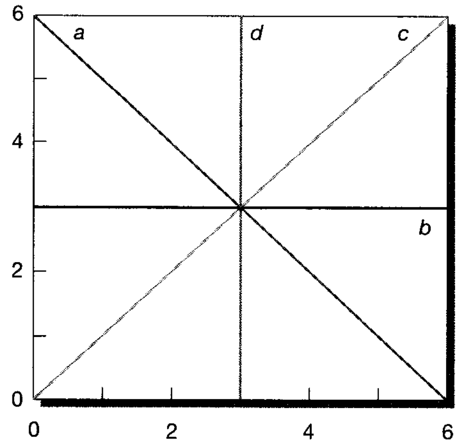
$$\beta_1 = \frac{\Delta Y}{\Delta X}$$

This is the ratio of change (Δ) in the rise of the line relative to the run or travel along the X axis. Exhibit 19-10 shows a few of the many possible slopes you may encounter.

The **intercept,** β_0, is the value for the linear function when it crosses the Y axis; it is the estimate of Y when $X = 0$. A formula for the intercept based on the mean scores of the X and Y variables is

$$\beta_0 = \bar{Y} - \beta_1 \bar{X}$$

> Exhibit 19-10 Examples of Different Slopes



Line	Slope
a	−1
b	0
c	+1
d	∞

>The test for a 0 slope is described later, in "Testing the Goodness of Fit."

Concept Application

What makes Generation X-ers all over the world select a glass of wine rather than a beer, Jack Daniels and Coke, or Bacardi Breezer? A research report from Australia highlights Generation X attitudes toward wine. The results suggest the top influencers are friends and family, wine reviews, and visits to wineries.[6] From the winery's perspective, tasting from the barrel is not only a widespread sales tool but also a major determinant of market *en primeur* or futures contracts, which represent about 60 percent of the harvest.

Weather is widely regarded as responsible for pronouncements about a wine's taste and potential quality. A Princeton economist has elaborated on that notion. He suggested that just a few facts about local weather conditions may be better predictors of vintage French red wines than the most refined palates and noses.[7] The regression model developed predicts an auction price index for about 80 wines from winter and harvest rainfall amounts and average growing-season temperatures. Interestingly, the calculations suggested that the 1989 Bordeaux would be one of the best since 1893. French traditionalists reacted hysterically to these methods yet agreed with the conclusion.

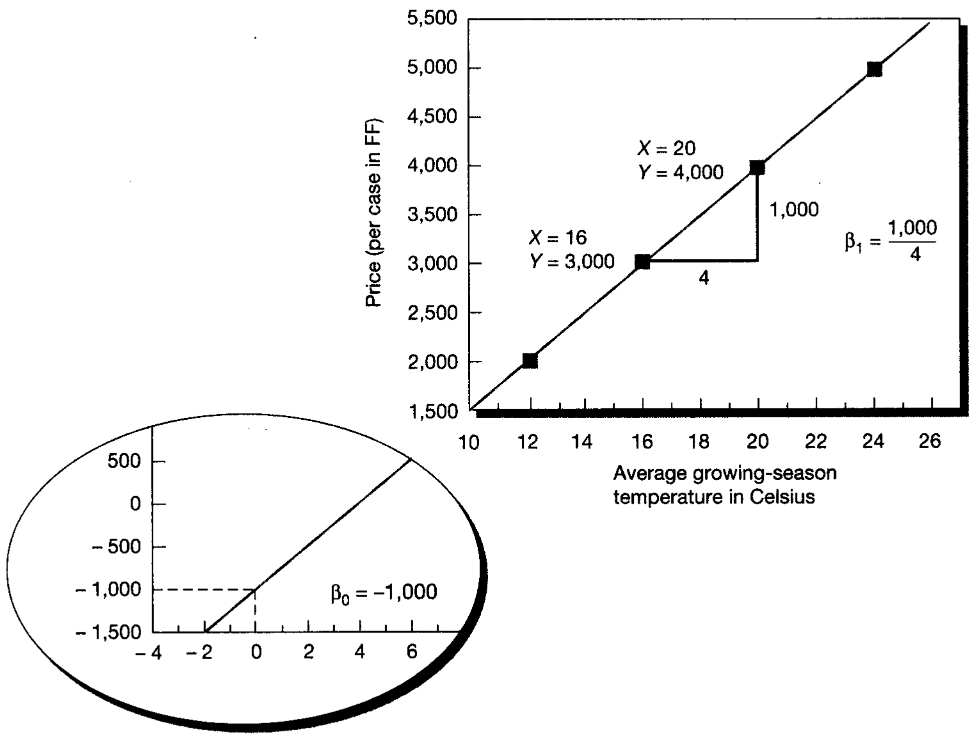
Our first example will use one predictor with highly simplified data. Let X represent the average growing-season temperature in degrees Celsius and Y the price of a 12-bottle case in French francs (6.8 French francs = 1 euro). The data appear below.

X Average Temperature (Celsius)	Y Price per Case (FF)
12	2,000
16	3,000
20	4,000
24	5,000
$\bar{X} = 18$	$\bar{Y} = 3{,}500$

The plotted data in Exhibit 19-11 show a linear relationship between the pairs of points and a perfect positive correlation, $r_{yx} = 1.0$. The slope of the line is calculated:

$$\beta_1 = \frac{Y_i - Y_j}{X_i - X_j} = \frac{4{,}000 - 3{,}000}{20 - 16} = \frac{1{,}000}{4} = 250$$

> Exhibit 19-11 Plot of Wine Price by Average Growing Temperature



where the X_iY_i values are the data points (20, 4,000) and X_jY_j are points (16, 3,000). The intercept β_0 is $-1{,}000$, the point at which $X = 0$ in this plot. This area is off the graph and appears in an insert on the figure.

$$\beta_0 = \bar{Y} - \beta_1\bar{X} = 3{,}500 - 250(18) = -1{,}000$$

Substituting into the formula, we have the simple regression equation

$$Y = -1{,}000 + 250X_i$$

We could now predict that a warm growing season with 25.5°C temperature would bring a case price of 5,375 French francs. $\hat{Y}$ (called *Y-hat*) is the predicted value of *Y*:

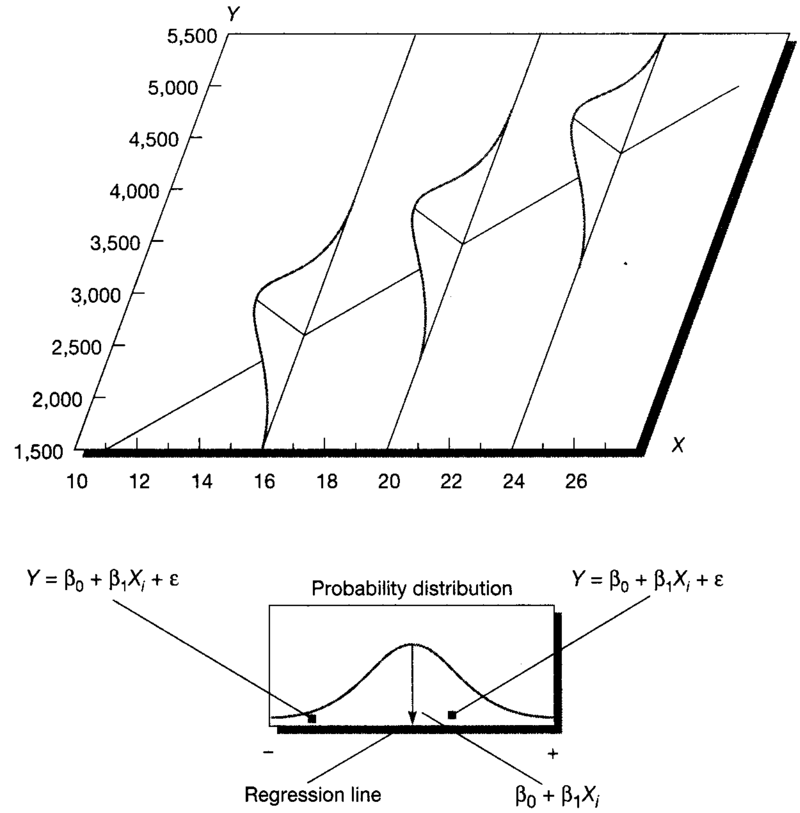
$$\hat{Y} = -1{,}000 + 250(25.5) = 5{,}375$$

Unfortunately, one rarely comes across a data set composed of four paired values, a perfect correlation, and an easily drawn line. A model based on such data is ***deterministic*** in that for any value of *X*, there is only one possible corresponding value of *Y*. It is more likely that we will collect data where the values of *Y* vary for each *X* value. Considering Exhibit 19-12, we should expect a distribution of price values for the temperature $X = 16$, another for $X = 20$, and another for each value of *X*. The means of these *Y* distributions will also vary in some systematic way with *X*. These variabilities lead us to construct a ***probabilistic*** model that also uses a linear function.[8] This function is written

$$Y_i = \beta_0 + \beta_1X_i + \varepsilon_1$$

where ε symbolizes the deviation of the *i*th observation from the mean, $\beta_0 + \beta_1X_i$.

> **Exhibit 19-12** Distribution of Y for Observations of X



As shown in Exhibit 19-12, the actual values of Y may be found above or below the regression line represented by the mean value of Y ($\beta_0 + \beta_1 X_i$) for a particular value of X. These deviations are the error in fitting the line and are often called the **error term.**

Method of Least Squares

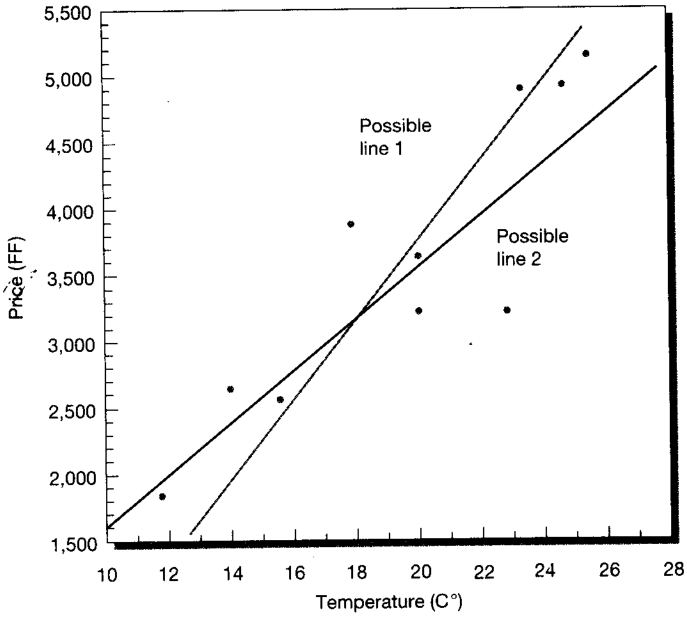
Exhibit 19-13 contains a new data set for the wine price example. Our prediction of Y from X must now account for the fact that the X and Y pairs do not fall neatly along the line. Actually, the relationship could be summarized by several lines. Exhibit 19-14 suggests two alternatives based on visual inspection—both of which produce errors, or vertical distances from the observed values to the line. The **method of least squares** allows us to find a regression line, or line of best fit, that will keep these errors to a minimum. It uses the

France's Bordeaux Business School offers a master of business administration in the wine sector. Surprised? Business schools throughout Europe are increasingly tailoring their programs with innovative degrees that respond to the changing environment of business. In addition to wine, MBA specializations focus on the music industry, luxury brands, sports management, agribusiness, e-business, consulting, and public sector specialities. **www.iht.com**

> Exhibit 19-13 Data for Wine Price Study

	Price (FF) Y	Temperature (C) X	XY	Y^2	X^2
1	1,813	11.80	21,393.40	3,286,969.00	139.24
2	2,558	15.70	40,160.60	6,543,364.00	246.49
3	2,628	14.00	36,792.00	6,906,384.00	196.00
4	3,217	22.90	73,669.30	10,349,089.00	524.41
5	3,228	20.00	64,560.00	10,419,984.00	400.00
6	3,629	20.10	72,942.90	13,169,641.00	404.01
7	3,886	17.90	69,559.40	15,100,996.00	320.41
8	4,897	23.40	114,589.80	23,980,609.00	547.56
9	4,933	24.60	121,351.80	24,334,489.00	605.16
10	5,199	25.70	133,614.30	27,029,601.00	660.49
Σ	35,988	196.10	748,633.50	141,121,126.00	4,043.77
Mean	3,598.80	19.61			
s	1,135.66	4.69			
Sum of squares (SS)	11,607,511.59	198.25	42,908.82		

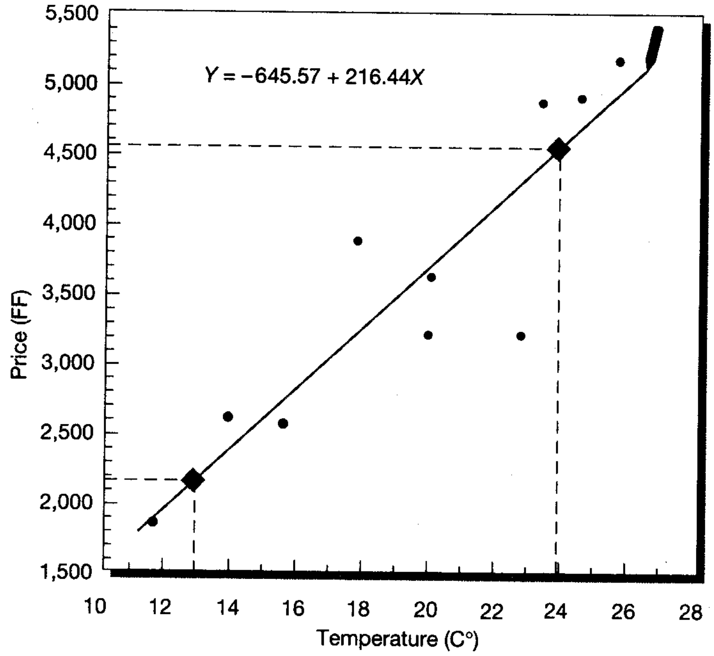
> Exhibit 19-14 Scatterplot and Possible Regression Lines Based on Visual Inspection: Wine Price Study



criterion of minimizing the total squared errors of estimate. When we predict values of Y for each X_i, the difference between the actual Y_i and the predicted $\hat{Y}$ is the error. This error is squared and then summed. The line of best fit is the one that minimizes the total squared errors of prediction.[9]

$$\sum_{i=1}^{n} e_i^2 \text{ minimized}$$

> **Exhibit 19-15** Drawing the Least-Squares Line: Wine Price Study



Regression coefficients β_0 and β_1 are used to find the least-squares solution. They are computed as follows:

$$\beta_1 = \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}}{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$$

Substituting data from Exhibit 19-13 into both formulas, we get

$$\beta_1 = \frac{748{,}633.5 - \frac{(196.1)(35{,}988)}{10}}{4{,}043.77 - \frac{(196.1)^2}{10}} = 216.439$$

$$\hat{\beta}_0 = 3{,}598.8 - (216.439)(19.61) = -645.569$$

The predictive equation is now $\hat{Y} = -645.57 + 216.44\,X_i$.

Drawing the Regression Line

Before drawing the regression line, we select two values of X to compute. Using values 13 and 24 for X_i, the points are

$$\hat{Y} = -645.57 + 216.44(13) = 2{,}168.15$$
$$\hat{Y} = -645.57 + 216.44(24) = 4{,}548.99$$