
ESTIMATION

CHAPTER OVERVIEW

This chapter covers estimation, one of the two types of statistical inference. As discussed in earlier chapters, statistics, such as means and variances, can be calculated from samples drawn from populations. These statistics serve as estimates of the corresponding population parameters. We expect these estimates to differ by some amount from the parameters they estimate. This chapter introduces estimation procedures that take these differences into account, thereby providing a foundation for statistical inference procedures discussed in the remaining chapters of the book.

TOPICS

- 6.1 INTRODUCTION
- 6.2 CONFIDENCE INTERVAL FOR A POPULATION MEAN
- 6.3 THE t DISTRIBUTION
- 6.4 CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO POPULATION MEANS
- 6.5 CONFIDENCE INTERVAL FOR A POPULATION PROPORTION
- 6.6 CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS
- 6.7 DETERMINATION OF SAMPLE SIZE FOR ESTIMATING MEANS
- 6.8 DETERMINATION OF SAMPLE SIZE FOR ESTIMATING PROPORTIONS
- 6.9 CONFIDENCE INTERVAL FOR THE VARIANCE OF A NORMALLY DISTRIBUTED POPULATION
- 6.10 CONFIDENCE INTERVAL FOR THE RATIO OF THE VARIANCES OF TWO NORMALLY DISTRIBUTED POPULATIONS
- 6.11 SUMMARY

LEARNING OUTCOMES

After studying this chapter, the student will

1. understand the importance and basic principles of estimation.
2. be able to calculate interval estimates for a variety of parameters.
3. be able to interpret a confidence interval from both a practical and a probabilistic viewpoint.
4. understand the basic properties and uses of the t distribution, chi-square distribution, and F distribution.

6.1 INTRODUCTION

We come now to a consideration of *estimation*, the first of the two general areas of statistical inference. The second general area, *hypothesis testing*, is examined in the next chapter.

We learned in Chapter 1 that inferential statistics is defined as follows.

DEFINITION

Statistical inference is the procedure by which we reach a conclusion about a population on the basis of the information contained in a sample drawn from that population.

The process of estimation entails calculating, from the data of a sample, some statistic that is offered as an approximation of the corresponding parameter of the population from which the sample was drawn.

The rationale behind estimation in the health sciences field rests on the assumption that workers in this field have an interest in the parameters, such as means and proportions, of various populations. If this is the case, there is a good reason why one must rely on estimating procedures to obtain information regarding these parameters. Many populations of interest, although finite, are so large that a 100 percent examination would be prohibitive from the standpoint of cost.

Suppose the administrator of a large hospital is interested in the mean age of patients admitted to his hospital during a given year. He may consider it too expensive to go through the records of all patients admitted during that particular year and, consequently, elect to examine a sample of the records from which he can compute an estimate of the mean age of patients admitted that year.

A physician in general practice may be interested in knowing what proportion of a certain type of individual, treated with a particular drug, suffers undesirable side effects. No doubt, her concept of the population consists of all those persons who ever have been or ever will be treated with this drug. Deferring a conclusion until the entire population has been observed could have an adverse effect on her practice.

These two examples have implied an interest in estimating, respectively, a population mean and a population proportion. Other parameters, the estimation of which we will cover in this chapter, are the difference between two means, the difference between two proportions, the population variance, and the ratio of two variances.

We will find that for each of the parameters we discuss, we can compute two types of estimate: a point estimate and an interval estimate.

DEFINITION

A *point estimate* is a single numerical value used to estimate the corresponding population parameter.

DEFINITION

An *interval estimate* consists of two numerical values defining a range of values that, with a specified degree of confidence, most likely includes the parameter being estimated.

These concepts will be elaborated on in the succeeding sections.

Choosing an Appropriate Estimator Note that a single computed value has been referred to as an *estimate*. The rule that tells us how to compute this value, or estimate, is referred to as an *estimator*. Estimators are usually presented as formulas. For example,

$$\bar{x} = \frac{\sum x_i}{n}$$

is an estimator of the population mean, μ . The single numerical value that results from evaluating this formula is called an estimate of the parameter μ .

In many cases, a parameter may be estimated by more than one estimator. For example, we could use the sample median to estimate the population mean. How then do we decide which estimator to use for estimating a given parameter? The decision is based on an objective measure or set of criteria that reflect some desired property of a particular estimator. When measured against these criteria, some estimators are better than others. One of these criteria is the property of *unbiasedness*.

DEFINITION

An estimator, say, T , of the parameter θ is said to be an unbiased estimator of θ if $E(T) = \theta$.

$E(T)$ is read, “the expected value of T .” For a finite population, $E(T)$ is obtained by taking the average value of T computed from all possible samples of a given size that may be drawn from the population. That is, $E(T) = \mu_T$. For an infinite population, $E(T)$ is defined in terms of calculus.

In the previous chapter we have seen that the sample mean, the sample proportion, the difference between two sample means, and the difference between two sample proportions are each unbiased estimates of their corresponding parameters. This property was implied when the parameters were said to be the means of the respective sampling distributions. For example, since the mean of the sampling distribution of \bar{x} is equal to μ , we know that \bar{x} is an unbiased estimator of μ . The other criteria of good estimators will not

be discussed in this book. The interested reader will find them covered in detail in most mathematical statistics texts.

Sampled Populations and Target Populations The health researcher who uses statistical inference procedures must be aware of the difference between two kinds of population—the sampled population and the target population.

DEFINITION _____

The *sampled population* is the population from which one actually draws a sample.

DEFINITION _____

The *target population* is the population about which one wishes to make an inference.

These two populations may or may not be the same. Statistical inference procedures allow one to make inferences about sampled populations (provided proper sampling methods have been employed). Only when the target population and the sampled population are the same is it possible for one to use statistical inference procedures to reach conclusions about the target population. If the sampled population and the target population are different, the researcher can reach conclusions about the target population only on the basis of nonstatistical considerations.

Suppose, for example, that a researcher wishes to assess the effectiveness of some method for treating rheumatoid arthritis. The target population consists of all patients suffering from the disease. It is not practical to draw a sample from this population. The researcher may, however, select a sample from all rheumatoid arthritis patients seen in some specific clinic. These patients constitute the sampled population, and, if proper sampling methods are used, inferences about this sampled population may be drawn on the basis of the information in the sample. If the researcher wishes to make inferences about all rheumatoid arthritis sufferers, he or she must rely on nonstatistical means to do so. Perhaps the researcher knows that the sampled population is similar, with respect to all important characteristics, to the target population. That is, the researcher may know that the age, sex, severity of illness, duration of illness, and so on are similar in both populations. And on the strength of this knowledge, the researcher may be willing to extrapolate his or her findings to the target population.

In many situations the sampled population and the target population are identical; when this is the case, inferences about the target population are straightforward. The researcher, however, should be aware that this is not always the case and not fall into the trap of drawing unwarranted inferences about a population that is different from the one that is sampled.

Random and Nonrandom Samples In the examples and exercises of this book, we assume that the data available for analysis have come from random samples. The strict validity of the statistical procedures discussed depends on this assumption. In many instances in real-world applications it is impossible or impractical to use truly random samples. In animal experiments, for example, researchers usually use whatever animals are available from suppliers or their own breeding stock. If the researchers had to depend on

randomly selected material, very little research of this type would be conducted. Again, nonstatistical considerations must play a part in the generalization process. Researchers may contend that the samples actually used are equivalent to simple random samples, since there is no reason to believe that the material actually used is not representative of the population about which inferences are desired.

In many health research projects, samples of convenience, rather than random samples, are employed. Researchers may have to rely on volunteer subjects or on readily available subjects such as students in their classes. Samples obtained from such sources are examples of *convenience samples*. Again, generalizations must be made on the basis of nonstatistical considerations. The consequences of such generalizations, however, may be useful or they may range from misleading to disastrous.

In some situations it is possible to introduce randomization into an experiment even though available subjects are not randomly selected from some well-defined population. In comparing two treatments, for example, each subject may be randomly assigned to one or the other of the treatments. Inferences in such cases apply to the treatments and not the subjects, and hence the inferences are valid.

6.2 CONFIDENCE INTERVAL FOR A POPULATION MEAN

Suppose researchers wish to estimate the mean of some normally distributed population. They draw a random sample of size n from the population and compute \bar{x} , which they use as a point estimate of μ . Although this estimator of μ possesses all the qualities of a good estimator, we know that because random sampling inherently involves chance, \bar{x} cannot be expected to be equal to μ .

It would be much more meaningful, therefore, to estimate μ by an interval that somehow communicates information regarding the probable magnitude of μ .

Sampling Distributions and Estimation To obtain an interval estimate, we must draw on our knowledge of sampling distributions. In the present case, because we are concerned with the sample mean as an estimator of a population mean, we must recall what we know about the sampling distribution of the sample mean.

In the previous chapter we learned that if sampling is from a normally distributed population, the sampling distribution of the sample mean will be normally distributed with a mean $\mu_{\bar{x}}$ equal to the population mean μ , and a variance $\sigma_{\bar{x}}^2$ equal to σ^2/n . We could plot the sampling distribution if we only knew where to locate it on the \bar{x} -axis. From our knowledge of normal distributions, in general, we know even more about the distribution of \bar{x} in this case. We know, for example, that regardless of where the distribution of \bar{x} is located, approximately 95 percent of the possible values of \bar{x} constituting the distribution are within two standard deviations of the mean. The two points that are two standard deviations from the mean are $\mu - 2\sigma_{\bar{x}}$ and $\mu + 2\sigma_{\bar{x}}$, so that the interval $\mu \pm 2\sigma_{\bar{x}}$ will contain approximately 95 percent of the possible values of \bar{x} . We know that μ and, hence $\mu_{\bar{x}}$, are unknown, but we may arbitrarily place the sampling distribution of \bar{x} on the \bar{x} -axis.

Since we do not know the value of μ , not a great deal is accomplished by the expression $\mu \pm 2\sigma_{\bar{x}}$. We do, however, have a point estimate of μ , which is \bar{x} . Would it be

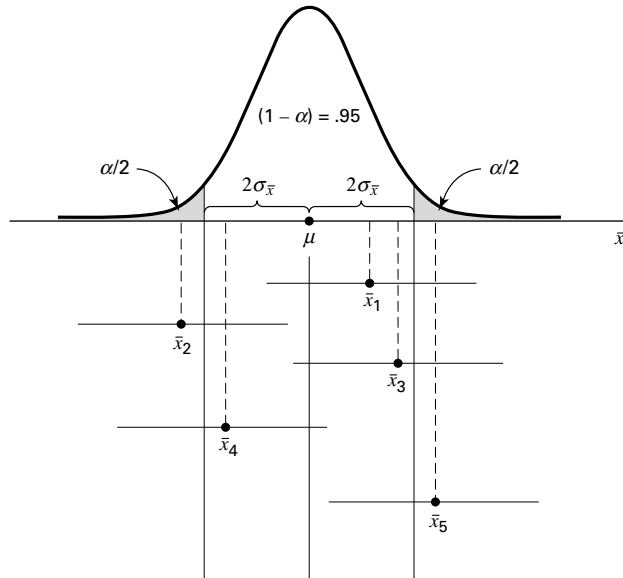


FIGURE 6.2.1 The 95 percent confidence interval for μ .

useful to construct an interval about this point estimate of μ ? The answer is yes. Suppose we constructed intervals about every possible value of \bar{x} computed from all possible samples of size n from the population of interest. We would have a large number of intervals of the form $\bar{x} \pm 2\sigma_{\bar{x}}$ with widths all equal to the width of the interval about the unknown μ . Approximately 95 percent of these intervals would have centers falling within the $\pm 2\sigma_{\bar{x}}$ interval about μ . Each of the intervals whose centers fall within $2\sigma_{\bar{x}}$ of μ would contain μ . These concepts are illustrated in Figure 6.2.1, in which we see that \bar{x}_1 , \bar{x}_3 , and \bar{x}_4 all fall within the interval about μ , and, consequently, the $2\sigma_{\bar{x}}$ intervals about these sample means include the value of μ . The sample means \bar{x}_2 and \bar{x}_5 do not fall within the $2\sigma_{\bar{x}}$ interval about μ , and the $2\sigma_{\bar{x}}$ intervals about them do not include μ .

EXAMPLE 6.2.1

Suppose a researcher, interested in obtaining an estimate of the average level of some enzyme in a certain human population, takes a sample of 10 individuals, determines the level of the enzyme in each, and computes a sample mean of $\bar{x} = 22$. Suppose further it is known that the variable of interest is approximately normally distributed with a variance of 45. We wish to estimate μ .

Solution: An approximate 95 percent confidence interval for μ is given by

$$\begin{aligned} & \bar{x} \pm 2\sigma_{\bar{x}} \\ & 22 \pm 2\sqrt{\frac{45}{10}} \\ & 22 \pm 2(2.1213) \\ & 17.76, 26.24 \end{aligned}$$

■

Interval Estimate Components Let us examine the composition of the interval estimate constructed in Example 6.2.1. It contains in its center the point estimate of μ . The 2 we recognize as a value from the standard normal distribution that tells us within how many standard errors lie approximately 95 percent of the possible values of \bar{x} . This value of z is referred to as the *reliability coefficient*. The last component, $\sigma_{\bar{x}}$, is the standard error, or standard deviation of the sampling distribution of \bar{x} . In general, then, an interval estimate may be expressed as follows:

$$\text{estimator} \pm (\text{reliability coefficient}) \times (\text{standard error}) \quad (6.2.1)$$

In particular, when sampling is from a normal distribution with known variance, an interval estimate for μ may be expressed as

$$\bar{x} \pm z_{(1-\alpha/2)}\sigma_{\bar{x}} \quad (6.2.2)$$

where $z_{(1-\alpha/2)}$ is the value of z to the left of which lies $1 - \alpha/2$ and to the right of which lies $\alpha/2$ of the area under its curve.

Interpreting Confidence Intervals How do we interpret the interval given by Expression 6.2.2? In the present example, where the reliability coefficient is equal to 2, we say that in repeated sampling approximately 95 percent of the intervals constructed by Expression 6.2.2 will include the population mean. This interpretation is based on the probability of occurrence of different values of \bar{x} . We may generalize this interpretation if we designate the total area under the curve of \bar{x} that is outside the interval $\mu \pm 2\sigma_{\bar{x}}$ as α and the area within the interval as $1 - \alpha$ and give the following *probabilistic interpretation* of Expression 6.2.2.

Probabilistic Interpretation

In repeated sampling, from a normally distributed population with a known standard deviation, 100(1 - α) percent of all intervals of the form $\bar{x} \pm z_{(1-\alpha/2)}\sigma_{\bar{x}}$ will in the long run include the population mean μ .

The quantity $1 - \alpha$, in this case .95, is called the *confidence coefficient* (or confidence level), and the interval $\bar{x} \pm z_{(1-\alpha/2)}\sigma_{\bar{x}}$ is called a *confidence interval* for μ . When $(1 - \alpha) = .95$, the interval is called the 95 percent confidence interval for μ . In the present example we say that we are 95 percent confident that the population mean is between 17.76 and 26.24. This is called the *practical interpretation* of Expression 6.2.2. In general, it may be expressed as follows.

Practical Interpretation

When sampling is from a normally distributed population with known standard deviation, we are 100(1 - α) percent confident that the single computed interval, $\bar{x} \pm z_{(1-\alpha/2)}\sigma_{\bar{x}}$, contains the population mean μ .

In the example given here we might prefer, rather than 2, the more exact value of z , 1.96, corresponding to a confidence coefficient of .95. Researchers may use any confidence coefficient they wish; the most frequently used values are .90, .95, and .99, which have associated reliability factors, respectively, of 1.645, 1.96, and 2.58.

Precision The quantity obtained by multiplying the reliability factor by the standard error of the mean is called the *precision* of the estimate. This quantity is also called the *margin of error*.

EXAMPLE 6.2.2

A physical therapist wished to estimate, with 99 percent confidence, the mean maximal strength of a particular muscle in a certain group of individuals. He is willing to assume that strength scores are approximately normally distributed with a variance of 144. A sample of 15 subjects who participated in the experiment yielded a mean of 84.3.

Solution: The z value corresponding to a confidence coefficient of .99 is found in Appendix Table D to be 2.58. This is our reliability coefficient. The standard error is $\sigma_{\bar{x}} = 12/\sqrt{15} = 3.0984$. Our 99 percent confidence interval for μ , then, is

$$84.3 \pm 2.58(3.0984)$$

$$84.3 \pm 8.0$$

$$76.3, 92.3$$

We say we are 99 percent confident that the population mean is between 76.3 and 92.3 since, in repeated sampling, 99 percent of all intervals that could be constructed in the manner just described would include the population mean. ■

Situations in which the variable of interest is approximately normally distributed with a known variance are quite rare. The purpose of the preceding examples, which assumed that these ideal conditions existed, was to establish the theoretical background for constructing confidence intervals for population means. In most practical situations either the variables are not approximately normally distributed or the population variances are not known or both. Example 6.2.3 and Section 6.3 explain the procedures that are available for use in the less than ideal, but more common, situations.

Sampling from Nonnormal Populations As noted, it will not always be possible or prudent to assume that the population of interest is normally distributed. Thanks to the central limit theorem, this will not deter us if we are able to select a large enough sample. We have learned that for large samples, the sampling distribution of \bar{x} is approximately normally distributed regardless of how the parent population is distributed.

EXAMPLE 6.2.3

Punctuality of patients in keeping appointments is of interest to a research team. In a study of patient flow through the offices of general practitioners, it was found that a sample of 35 patients was 17.2 minutes late for appointments, on the average. Previous research had shown the standard deviation to be about 8 minutes. The population distribution was felt to be nonnormal. What is the 90 percent confidence interval for μ , the true mean amount of time late for appointments?

Solution: Since the sample size is fairly large (greater than 30), and since the population standard deviation is known, we draw on the central limit theorem and assume the sampling distribution of \bar{x} to be approximately normally distributed. From Appendix Table D we find the reliability coefficient corresponding to a confidence coefficient of .90 to be about 1.645, if we interpolate. The standard error is $\sigma_{\bar{x}} = 8/\sqrt{35} = 1.3522$, so that our 90 percent confidence interval for μ is

$$17.2 \pm 1.645(1.3522)$$

$$17.2 \pm 2.2$$

$$15.0, 19.4$$

Frequently, when the sample is large enough for the application of the central limit theorem, the population variance is unknown. In that case we use the sample variance as a replacement for the unknown population variance in the formula for constructing a confidence interval for the population mean.

Computer Analysis When confidence intervals are desired, a great deal of time can be saved if one uses a computer, which can be programmed to construct intervals from raw data.

EXAMPLE 6.2.4

The following are the activity values (micromoles per minute per gram of tissue) of a certain enzyme measured in normal gastric tissue of 35 patients with gastric carcinoma.

.360	1.189	.614	.788	.273	2.464	.571
1.827	.537	.374	.449	.262	.448	.971
.372	.898	.411	.348	1.925	.550	.622
.610	.319	.406	.413	.767	.385	.674
.521	.603	.533	.662	1.177	.307	1.499

We wish to use the MINITAB computer software package to construct a 95 percent confidence interval for the population mean. Suppose we know that the population variance is .36. It is not necessary to assume that the sampled population of values is normally distributed since the sample size is sufficiently large for application of the central limit theorem.

Solution: We enter the data into Column 1 and proceed as shown in Figure 6.2.2. These instructions tell the computer that the reliability factor is z , that a 95 percent confidence interval is desired, that the population standard deviation is .6, and that the data are in Column 1. The output tells us that the sample mean is .718, the sample standard deviation is .511, and the standard error of the mean, σ/\sqrt{n} is $.6/\sqrt{35} = .101$.

We are 95 percent confident that the population mean is somewhere between .519 and .917. Confidence intervals may be obtained through the use of many other software packages. Users of SAS[®], for example, may wish to use the output from PROC MEANS or PROC UNIVARIATE to construct confidence intervals.

<p>Dialog box:</p> <p>Stat > Basic Statistics > 1-Sample z</p> <p>Type <i>C1</i> in Samples in Columns. Type <i>.6</i> in Standard deviation. Click OK.</p> <p>Output:</p> <p>One-Sample Z: C1</p> <p>The assumed standard deviation = 0.600</p> <table border="1"> <thead> <tr> <th>Variable</th> <th>N</th> <th>Mean</th> <th>StDev</th> <th>SE Mean</th> <th>95.0 % C.I.</th> </tr> </thead> <tbody> <tr> <td>MicMoles</td> <td>35</td> <td>0.718</td> <td>0.511</td> <td>0.101</td> <td>(0.519, 0.917)</td> </tr> </tbody> </table>	Variable	N	Mean	StDev	SE Mean	95.0 % C.I.	MicMoles	35	0.718	0.511	0.101	(0.519, 0.917)	<p>Session command:</p> <p>MTB > ZINTERVAL 95 .6 C1</p>
Variable	N	Mean	StDev	SE Mean	95.0 % C.I.								
MicMoles	35	0.718	0.511	0.101	(0.519, 0.917)								

FIGURE 6.2.2 MINITAB procedure for constructing 95 percent confidence interval for a population mean, Example 6.2.4.

Alternative Estimates of Central Tendency As noted previously, the mean is sensitive to extreme values—those values that deviate appreciably from most of the measurements in a data set. They are sometimes referred to as *outliers*. We also noted earlier that the median, because it is not so sensitive to extreme measurements, is sometimes preferred over the mean as a measure of central tendency when outliers are present. For the same reason, we may prefer to use the sample median as an estimator of the population median when we wish to make an inference about the central tendency of a population. Not only may we use the sample median as a point estimate of the population median, we also may construct a confidence interval for the population median. The formula is not given here but may be found in the book by Rice (1).

Trimmed Mean Estimators that are insensitive to outliers are called *robust estimators*. Another robust measure and estimator of central tendency is the *trimmed mean*. For a set of sample data containing n measurements we calculate the 100α percent trimmed mean as follows:

1. Order the measurements.
2. Discard the smallest 100α percent and the largest 100α percent of the measurements. The recommended value of α is something between .1 and .2.
3. Compute the arithmetic mean of the remaining measurements.

Note that the median may be regarded as a 50 percent trimmed mean.

EXERCISES

For each of the following exercises construct 90, 95, and 99 percent confidence intervals for the population mean, and state the practical and probabilistic interpretations of each. Indicate which interpretation you think would be more appropriate to use when discussing confidence intervals with

someone who has not had a course in statistics, and state the reason for your choice. Explain why the three intervals that you construct are not of equal width. Indicate which of the three intervals you would prefer to use as an estimate of the population mean, and state the reason for your choice.

- 6.2.1.** We wish to estimate the average number of heartbeats per minute for a certain population. The average number of heartbeats per minute for a sample of 49 subjects was found to be 90. Assume that these 49 patients constitute a random sample, and that the population is normally distributed with a standard deviation of 10.
- 6.2.2.** We wish to estimate the mean serum indirect bilirubin level of 4-day-old infants. The mean for a sample of 16 infants was found to be 5.98 mg/100 cc. Assume that bilirubin levels in 4-day-old infants are approximately normally distributed with a standard deviation of 3.5 mg/100 cc.
- 6.2.3.** In a length of hospitalization study conducted by several cooperating hospitals, a random sample of 64 peptic ulcer patients was drawn from a list of all peptic ulcer patients ever admitted to the participating hospitals and the length of hospitalization per admission was determined for each. The mean length of hospitalization was found to be 8.25 days. The population standard deviation is known to be 3 days.
- 6.2.4.** A sample of 100 apparently normal adult males, 25 years old, had a mean systolic blood pressure of 125. It is believed that the population standard deviation is 15.
- 6.2.5.** Some studies of Alzheimer's disease (AD) have shown an increase in $^{14}\text{CO}_2$ production in patients with the disease. In one such study the following $^{14}\text{CO}_2$ values were obtained from 16 neocortical biopsy samples from AD patients.

1009	1280	1180	1255	1547	2352	1956	1080
1776	1767	1680	2050	1452	2857	3100	1621

Assume that the population of such values is normally distributed with a standard deviation of 350.

6.3 THE t DISTRIBUTION

In Section 6.2, a procedure was outlined for constructing a confidence interval for a population mean. The procedure requires knowledge of the variance of the population from which the sample is drawn. It may seem somewhat strange that one can have knowledge of the population variance and not know the value of the population mean. Indeed, it is the usual case, in situations such as have been presented, that the population variance, as well as the population mean, is unknown. This condition presents a problem with respect to constructing confidence intervals. Although, for example, the statistic

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

is normally distributed when the population is normally distributed and is at least approximately normally distributed when n is large, regardless of the functional form of the population, we cannot make use of this fact because σ is unknown. However, all is not lost, and the most logical solution to the problem is the one followed. We use the sample standard deviation

$$s = \sqrt{\sum (x_i - \bar{x})^2 / (n - 1)}$$

to replace σ . When the sample size is large, say, greater than 30, our faith in s as an approximation of σ is usually substantial, and we may be appropriately justified in using normal distribution theory to construct a confidence interval for the population mean. In that event, we proceed as instructed in Section 6.2.

It is when we have small samples that it becomes mandatory for us to find an alternative procedure for constructing confidence intervals.

As a result of the work of Gosset (2), writing under the pseudonym of “Student,” an alternative, known as *Student’s t distribution*, usually shortened to *t distribution*, is available to us.

The quantity

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad (6.3.1)$$

follows this distribution.

Properties of the t Distribution The t distribution has the following properties.

1. It has a mean of 0.
2. It is symmetrical about the mean.
3. In general, it has a variance greater than 1, but the variance approaches 1 as the sample size becomes large. For $df > 2$, the variance of the t distribution is $df/(df - 2)$, where df is the degrees of freedom. Alternatively, since here $df = n - 1$ for $n > 3$, we may write the variance of the t distribution as $(n - 1)/(n - 3)$.
4. The variable t ranges from $-\infty$ to $+\infty$.
5. The t distribution is really a family of distributions, since there is a different distribution for each sample value of $n - 1$, the divisor used in computing s^2 . We recall that $n - 1$ is referred to as degrees of freedom. Figure 6.3.1 shows t distributions corresponding to several degrees-of-freedom values.

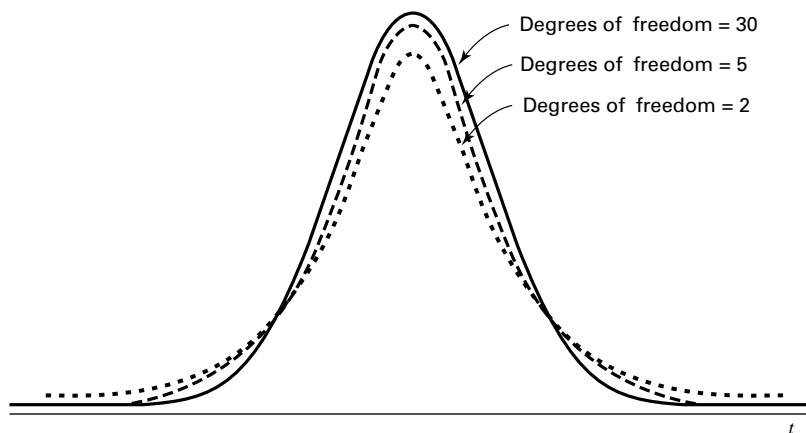


FIGURE 6.3.1 The t distribution for different degrees-of-freedom values.

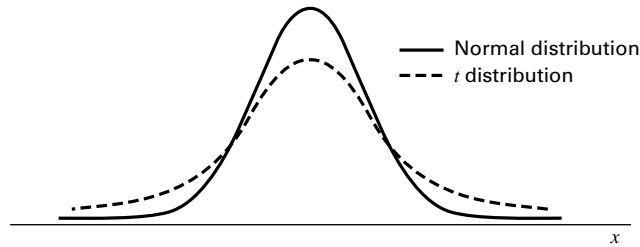


FIGURE 6.3.2 Comparison of normal distribution and t distribution.

6. Compared to the normal distribution, the t distribution is less peaked in the center and has thicker tails. Figure 6.3.2 compares the t distribution with the normal.
7. The t distribution approaches the normal distribution as $n - 1$ approaches infinity.

The t distribution, like the standard normal, has been extensively tabulated. One such table is given as Table E in the Appendix. As we will see, we must take both the confidence coefficient and degrees of freedom into account when using the table of the t distribution.

You may use MINITAB to graph the t distribution (for specified degrees-of-freedom values) and other distributions. After designating the horizontal axis by following directions in the Set Patterned Data box, choose menu path Calc and then Probability Distributions. Finally, click on the distribution desired and follow the instructions. Use the Plot dialog box to plot the graph.

Confidence Intervals Using t The general procedure for constructing confidence intervals is not affected by our having to use the t distribution rather than the standard normal distribution. We still make use of the relationship expressed by

$$\text{estimator} \pm (\text{reliability coefficient}) \times (\text{standard error of the estimator})$$

What is different is the source of the reliability coefficient. It is now obtained from the table of the t distribution rather than from the table of the standard normal distribution. To be more specific, *when sampling is from a normal distribution whose standard deviation, σ , is unknown, the $100(1 - \alpha)$ percent confidence interval for the population mean, μ , is given by*

$$\bar{x} \pm t_{(1-\alpha/2)} \frac{s}{\sqrt{n}} \quad (6.3.2)$$

We emphasize that a requirement for the strictly valid use of the t distribution is that the sample must be drawn from a normal distribution. Experience has shown, however, that moderate departures from this requirement can be tolerated. As a consequence, the t distribution is used even when it is known that the parent population deviates somewhat from normality. Most researchers require that an assumption of, at least, a mound-shaped population distribution be tenable.

EXAMPLE 6.3.1

Maffulli et al. (A-1) studied the effectiveness of early weightbearing and ankle mobilization therapies following acute repair of a ruptured Achilles tendon. One of the variables

they measured following treatment was the isometric gastrocsoleus muscle strength. In 19 subjects, the mean isometric strength for the operated limb (in newtons) was 250.8 with a standard deviation of 130.9. We assume that these 19 patients constitute a random sample from a population of similar subjects. We wish to use these sample data to estimate for the population the mean isometric strength after surgery.

Solution: We may use the sample mean, 250.8, as a point estimate of the population mean but, because the population standard deviation is unknown, we must assume the population of values to be at least approximately normally distributed before constructing a confidence interval for μ . Let us assume that such an assumption is reasonable and that a 95 percent confidence interval is desired. We have our estimator, \bar{x} , and our standard error is $s/\sqrt{n} = 130.9/\sqrt{19} = 30.0305$. We need now to find the reliability coefficient, the value of t associated with a confidence coefficient of .95 and $n - 1 = 18$ degrees of freedom. Since a 95 percent confidence interval leaves .05 of the area under the curve of t to be equally divided between the two tails, we need the value of t to the right of which lies .025 of the area. We locate in Appendix Table E the column headed $t_{.975}$. This is the value of t to the left of which lies .975 of the area under the curve. The area to the right of this value is equal to the desired .025. We now locate the number 18 in the degrees-of-freedom column. The value at the intersection of the row labeled 18 and the column labeled $t_{.975}$ is the t we seek. This value of t , which is our reliability coefficient, is found to be 2.1009. We now construct our 95 percent confidence interval as follows:

$$250.8 \pm 2.1009(30.0305)$$

$$250.8 \pm 63.1$$

$$187.7, 313.9$$

This interval may be interpreted from both the probabilistic and practical points of view. We are 95 percent confident that the true population mean, μ , is somewhere between 187.7 and 313.9 because, in repeated sampling, 95 percent of intervals constructed in like manner will include μ .

Deciding Between z and t When we construct a confidence interval for a population mean, we must decide whether to use a value of z or a value of t as the reliability factor. To make an appropriate choice we must consider sample size, whether the sampled population is normally distributed, and whether the population variance is known. Figure 6.3.3 provides a flowchart that one can use to decide quickly whether the reliability factor should be z or t .

Computer Analysis If you wish to have MINITAB construct a confidence interval for a population mean when the t statistic is the appropriate reliability factor, the command is TINTERVAL. In Windows choose 1-Sample t from the Basic Statistics menu.

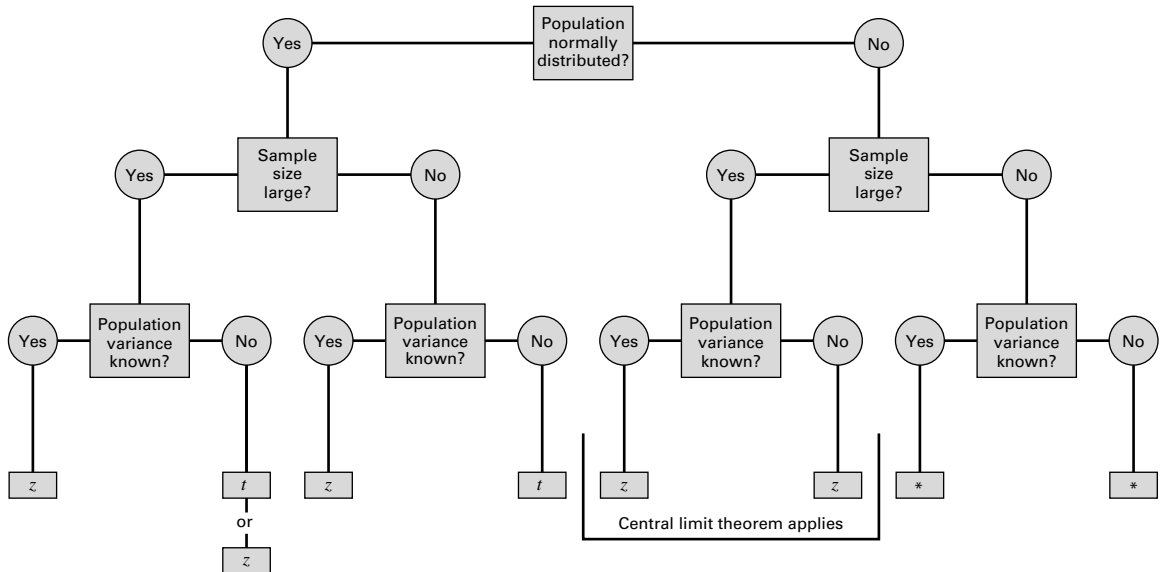


FIGURE 6.3.3 Flowchart for use in deciding between z and t when making inferences about population means. (*Use a nonparametric procedure. See Chapter 13.)

EXERCISES

- 6.3.1.** Use the t distribution to find the reliability factor for a confidence interval based on the following confidence coefficients and sample sizes:

	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
Confidence coefficient	.95	.99	.90	.95
Sample size	15	24	8	30

- 6.3.2.** In a study of the effects of early Alzheimer’s disease on nondeclarative memory, Reber et al. (A-2) used the Category Fluency Test to establish baseline persistence and semantic memory and language abilities. The eight subjects in the sample had Category Fluency Test scores of 11, 10, 6, 3, 11, 10, 9, 11. Assume that the eight subjects constitute a simple random sample from a normally distributed population of similar subjects with early Alzheimer’s disease.
- What is the point estimate of the population mean?
 - What is the standard deviation of the sample?
 - What is the estimated standard error of the sample mean?
 - Construct a 95 percent confidence interval for the population mean category fluency test score.
 - What is the precision of the estimate?
 - State the probabilistic interpretation of the confidence interval you constructed.
 - State the practical interpretation of the confidence interval you constructed.
- 6.3.3.** Pedroletti et al. (A-3) reported the maximal nitric oxide diffusion rate in a sample of 15 asthmatic schoolchildren and 15 controls as mean \pm standard error of the mean. For asthmatic children, they

reported 3.5 ± 0.4 nL/s (nanoliters per second) and for control subjects they reported $0.7 \pm .1$ nL/s. For each group, determine the following:

- (a) What was the sample standard deviation?
 - (b) What is the 95 percent confidence interval for the mean maximal nitric oxide diffusion rate of the population?
 - (c) What assumptions are necessary for the validity of the confidence interval you constructed?
 - (d) What are the practical and probabilistic interpretations of the interval you constructed?
 - (e) Which interpretation would be more appropriate to use when discussing confidence intervals with someone who has not had a course in statistics? State the reasons for your choice.
 - (f) If you were to construct a 90 percent confidence interval for the population mean from the information given here, would the interval be wider or narrower than the 95 percent confidence interval? Explain your answer without actually constructing the interval.
 - (g) If you were to construct a 99 percent confidence interval for the population mean from the information given here, would the interval be wider or narrower than the 95 percent confidence interval? Explain your answer without actually constructing the interval.
- 6.3.4.** The concern of a study by Beynnon et al. (A-4) were nine subjects with chronic anterior cruciate ligament (ACL) tears. One of the variables of interest was the laxity of the anteroposterior, where higher values indicate more knee instability. The researchers found that among subjects with ACL-deficient knees, the mean laxity value was 17.4 mm with a standard deviation of 4.3 mm.
- (a) What is the estimated standard error of the mean?
 - (b) Construct the 99 percent confidence interval for the mean of the population from which the nine subjects may be presumed to be a random sample.
 - (c) What is the precision of the estimate?
 - (d) What assumptions are necessary for the validity of the confidence interval you constructed?
- 6.3.5.** A sample of 16 ten-year-old girls had a mean weight of 71.5 and a standard deviation of 12 pounds, respectively. Assuming normality, find the 90, 95, and 99 percent confidence intervals for μ .
- 6.3.6.** The subjects of a study by Dugoff et al. (A-5) were 10 obstetrics and gynecology interns at the University of Colorado Health Sciences Center. The researchers wanted to assess competence in performing clinical breast examinations. One of the baseline measurements was the number of such examinations performed. The following data give the number of breast examinations performed for this sample of 10 interns.

Intern Number	No. of Breast Exams Performed
1	30
2	40
3	8
4	20
5	26
6	35
7	35
8	20
9	25
10	20

Source: Lorraine Dugoff, Mauritha R. Everett, Louis Vontver, and Gwyn E. Barley, "Evaluation of Pelvic and Breast Examination Skills of Interns in Obstetrics and Gynecology and Internal Medicine," *American Journal of Obstetrics and Gynecology*, 189 (2003), 655–658.

Construct a 95 percent confidence interval for the mean of the population from which the study subjects may be presumed to have been drawn.

6.4 CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO POPULATION MEANS

Sometimes there arise cases in which we are interested in estimating the difference between two population means. From each of the populations an independent random sample is drawn and, from the data of each, the sample means \bar{x}_1 and \bar{x}_2 , respectively, are computed. We learned in the previous chapter that the estimator $\bar{x}_1 - \bar{x}_2$ yields an unbiased estimate of $\mu_1 - \mu_2$, the difference between the population means. The variance of the estimator is $(\sigma_1^2/n_1) + (\sigma_2^2/n_2)$. We also know from Chapter 5 that, depending on the conditions, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ may be, at least, approximately normally distributed, so that in many cases we make use of the theory relevant to normal distributions to compute a confidence interval for $\mu_1 - \mu_2$. When the population variances are known, the $100(1 - \alpha)$ percent confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad (6.4.1)$$

An examination of a confidence interval for the difference between population means provides information that is helpful in deciding whether or not it is likely that the two population means are equal. When the constructed interval does not include zero, we say that the interval provides evidence that the two population means are not equal. When the interval includes zero, we say that the population means may be equal.

Let us illustrate a case where sampling is from the normal distributions.

EXAMPLE 6.4.1

A research team is interested in the difference between serum uric acid levels in patients with and without Down's syndrome. In a large hospital for the treatment of the mentally challenged, a sample of 12 individuals with Down's syndrome yielded a mean of $\bar{x}_1 = 4.5$ mg/100 ml. In a general hospital a sample of 15 normal individuals of the same age and sex were found to have a mean value of $\bar{x}_2 = 3.4$. If it is reasonable to assume that the two populations of values are normally distributed with variances equal to 1 and 1.5, find the 95 percent confidence interval for $\mu_1 - \mu_2$.

Solution: For a point estimate of $\mu_1 - \mu_2$, we use $\bar{x}_1 - \bar{x}_2 = 4.5 - 3.4 = 1.1$. The reliability coefficient corresponding to .95 is found in Appendix Table D to be 1.96. The standard error is

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{1}{12} + \frac{1.5}{15}} = .4282$$

The 95 percent confidence interval, then, is

$$\begin{aligned} 1.1 \pm 1.96(.4282) \\ 1.1 \pm .84 \\ (.26, 1.94) \end{aligned}$$

We say that we are 95 percent confident that the true difference, $\mu_1 - \mu_2$, is somewhere between .26 and 1.94 because, in repeated sampling, 95 percent of the intervals constructed in this manner would include the difference between the true means.

Since the interval does not include zero, we conclude that the two population means are not equal. ■

Sampling from Non-normal Populations The construction of a confidence interval for the difference between two population means when sampling is from non-normal populations proceeds in the same manner as in Example 6.4.1 if the sample sizes n_1 and n_2 are large. Again, this is a result of the central limit theorem. If the population variances are unknown, we use the sample variances to estimate them.

EXAMPLE 6.4.2

Despite common knowledge of the adverse effects of doing so, many women continue to smoke while pregnant. Mayhew et al. (A-6) examined the effectiveness of a smoking cessation program for pregnant women. The mean number of cigarettes smoked daily at the close of the program by the 328 women who completed the program was 4.3 with a standard deviation of 5.22. Among 64 women who did not complete the program, the mean number of cigarettes smoked per day at the close of the program was 13 with a standard deviation of 8.97. We wish to construct a 99 percent confidence interval for the difference between the means of the populations from which the samples may be presumed to have been selected.

Solution: No information is given regarding the shape of the distribution of cigarettes smoked per day. Since our sample sizes are large, however, the central limit theorem assures us that the sampling distribution of the difference between sample means will be approximately normally distributed even if the distribution of the variable in the populations is not normally distributed. We may use this fact as justification for using the z statistic as the reliability factor in the construction of our confidence interval. Also, since the population standard deviations are not given, we will use the sample standard deviations to estimate them. The point estimate for the difference between population means is the difference between sample means, $4.3 - 13.0 = -8.7$. In Appendix Table D we find the reliability factor to be 2.58. The estimated standard error is

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{5.22^2}{328} + \frac{8.97^2}{64}} = 1.1577$$

By Equation 6.4.1, our 99 percent confidence interval for the difference between population means is

$$\begin{aligned} & -8.7 \pm 2.58(1.1577) \\ & (-11.7, -5.7) \end{aligned}$$

We are 99 percent confident that the mean number of cigarettes smoked per day for women who complete the program is between 5.7 and 11.7 lower than the mean for women who do not complete the program. ■

The t Distribution and the Difference Between Means When population variances are unknown, and we wish to estimate the difference between two population means with a confidence interval, we can use the t distribution as a source of the reliability factor if certain assumptions are met. We must know, or be willing to assume, that the two sampled populations are normally distributed. With regard to the population variances, we distinguish between two situations: (1) the situation in which the population variances are equal, and (2) the situation in which they are not equal. Let us consider each situation separately.

Population Variances Equal If the assumption of equal population variances is justified, the two sample variances that we compute from our two independent samples may be considered as estimates of the same quantity, the common variance. It seems logical, then, that we should somehow capitalize on this in our analysis. We do just that and obtain a *pooled estimate* of the common variance. This pooled estimate is obtained by computing the weighted average of the two sample variances. Each sample variance is weighted by its degrees of freedom. If the sample sizes are equal, this weighted average is the arithmetic mean of the two sample variances. If the two sample sizes are unequal, the weighted average takes advantage of the additional information provided by the larger sample. The pooled estimate is given by the formula

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (6.4.2)$$

The standard error of the estimate, then, is given by

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \quad (6.4.3)$$

and the $100(1 - \alpha)$ percent confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) \pm t_{(1-\alpha/2)} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} \quad (6.4.4)$$

The number of degrees of freedom used in determining the value of t to use in constructing the interval is $n_1 + n_2 - 2$, the denominator of Equation 6.4.2. We interpret this interval in the usual manner.

Methods that may be used in reaching a decision about the equality of population variances are discussed in Sections 6.10 and 7.8.

EXAMPLE 6.4.3

The purpose of a study by Granholm et al. (A-7) was to determine the effectiveness of an integrated outpatient dual-diagnosis treatment program for mentally ill subjects. The authors were addressing the problem of substance abuse issues among people with severe mental disorders. A retrospective chart review was performed on 50 consecutive patient referrals to the Substance Abuse/Mental Illness program at the VA San Diego Healthcare System. One of the outcome variables examined was the number of inpatient treatment days for psychiatric disorder during the year following the end of the program. Among 18 subjects with schizophrenia, the mean number of treatment days was 4.7 with a standard deviation of 9.3. For 10 subjects with bipolar disorder, the mean number of psychiatric disorder treatment days was 8.8 with a standard deviation of 11.5. We wish to construct a 95 percent confidence interval for the difference between the means of the populations represented by these two samples.

Solution: First we use Equation 6.4.2 to compute the pooled estimate of the common population variance.

$$s_p^2 = \frac{(18 - 1)(9.3^2) + (10 - 1)(11.5)^2}{18 + 10 - 2} = 102.33$$

When we enter Appendix Table E with $18 + 10 - 2 = 26$ degrees of freedom and a desired confidence level of .95, we find that the reliability factor is 2.0555. By Expression 6.4.4 we compute the 95 percent confidence interval for the difference between population means as follows:

$$\begin{aligned} (4.7 - 8.8) \pm 2.0555 \sqrt{\frac{102.33}{18} + \frac{102.33}{10}} \\ -4.1 \pm 8.20 \\ (-12.3, 4.1) \end{aligned}$$

We are 95 percent confident that the difference between population means is somewhere between -12.3 and 4.10 . We can say this because we know that if we were to repeat the study many, many times, and compute confidence intervals in the same way, about 95 percent of the intervals would include the difference between the population means.

Since the interval includes zero, we conclude that the population means may be equal. ■

Population Variances Not Equal When one is unable to conclude that the variances of two populations of interest are equal, even though the two populations may be assumed to be normally distributed, it is not proper to use the t distribution as just outlined in constructing confidence intervals.

As a practical rule in applied problems, one may wish to assume the inequality of variances if the ratio of the larger to the smaller variance exceeds 2; however, a more formal test is described in Section 6.10.

A solution to the problem of unequal variances was proposed by Behrens (3) and later was verified and generalized by Fisher (4,5). Solutions have also been proposed by Neyman (6), Scheffé (7,8), and Welch (9,10). The problem is discussed in detail by Cochran (11).

The problem revolves around the fact that the quantity

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

does not follow a t distribution with $n_1 + n_2 - 2$ degrees of freedom when the population variances are not equal. Consequently, the t distribution cannot be used in the usual way to obtain the reliability factor for the confidence interval for the difference between the means of two populations that have unequal variances. The solution proposed by Cochran consists of computing the reliability factor, $t'_{1-\alpha/2}$, by the following formula:

$$t'_{1-\alpha/2} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2} \quad (6.4.5)$$

where $w_1 = s_1^2/n_1$, $w_2 = s_2^2/n_2$, $t_1 = t_{1-\alpha/2}$ for $n_1 - 1$ degrees of freedom, and $t_2 = t_{1-\alpha/2}$ for $n_2 - 1$ degrees of freedom. An approximate 100(1 - α) percent confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) \pm t'_{(1-\alpha/2)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \quad (6.4.6)$$

Adjustments to the reliability coefficient may also be made by reducing the number of degrees of freedom instead of modifying t in the manner just demonstrated. Many computer programs calculate an adjusted reliability coefficient in this way.

EXAMPLE 6.4.4

Let us reexamine the data presented in Example 6.4.3 from the study by Granholm et al. (A-7). Recall that among the 18 subjects with schizophrenia, the mean number of treatment days was 4.7 with a standard deviation of 9.3. In the bipolar disorder treatment group of 10 subjects, the mean number of psychiatric disorder treatment days was 8.8 with a standard deviation of 11.5. We assume that the two populations of number of psychiatric disorder days are approximately normally distributed. Now let us assume, however, that the two population variances are not equal. We wish to construct a 95 percent confidence interval for the difference between the means of the two populations represented by the samples.

Solution: We will use t' as found in Equation 6.4.5 for the reliability factor. Reference to Appendix Table E shows that with 17 degrees of freedom and $1 - .05/2 = .975$, $t_1 = 2.1098$. Similarly, with 9 degrees of freedom and


```

R Code:
> tsum.test(mean.x=4.7, s.x=9.3, n.x=18, mean.y=8.8, s.y=11.5, n.y=10, alternative =
"two.sided", mu=0, var.equal=FALSE, conf.level=0.95)

ROutput:
Welch Modified Two-Sample t-Test

data: Summarized x and y

t = -0.9656, df = 15.635, p-value = 0.349

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:
-13.118585  4.918585

sample estimates:
mean of x mean of y
  4.7      8.8

```

FIGURE 6.4.2 Program R example calculation for the confidence interval between two means assuming unequal variances using the data in Example 6.4.4.

EXERCISES

For each of the following exercises construct 90, 95, and 99 percent confidence intervals for the difference between population means. Where appropriate, state the assumptions that make your method valid. State the practical and probabilistic interpretations of each interval that you construct. Consider the variables under consideration in each exercise, and state what use you think researchers might make of your results.

- 6.4.1. Iannelo et al. (A-8) performed a study that examined free fatty acid concentrations in 18 lean subjects and 11 obese subjects. The lean subjects had a mean level of $299 \mu\text{Eq/L}$ with a standard error of the mean of 30, while the obese subjects had a mean of $744 \mu\text{Eq/L}$ with a standard error of the mean of 62.
- 6.4.2. Chan et al. (A-9) developed a questionnaire to assess knowledge of prostate cancer. There was a total of 36 questions to which respondents could answer “agree,” “disagree,” or “don’t know.” Scores could range from 0 to 36. The mean scores for Caucasian study participants was 20.6 with a standard deviation of 5.8, while the mean scores for African-American men was 17.4 with a standard deviation of 5.8. The number of Caucasian study participants was 185, and the number of African-Americans was 86.
- 6.4.3. The objectives of a study by van Vollenhoven et al. (A-10) were to examine the effectiveness of etanercept alone and etanercept in combination with methotrexate in the treatment of rheumatoid arthritis. The researchers conducted a retrospective study using data from the STURE database, which collects efficacy and safety data for all patients starting biological treatments at the major hospitals in Stockholm, Sweden. The researchers identified 40 subjects who were prescribed etanercept only and 57 subjects who were given etanercept with methotrexate. Using a 100-mm visual analogue scale (the higher the value, the greater the pain), researchers found that after 3 months of treatment, the mean pain score was 36.4 with a standard error of the mean of 5.5 for subjects taking etanercept only. In the sample receiving etanercept plus methotrexate, the mean score was 30.5 with a standard error of the mean of 4.6.
- 6.4.4. The purpose of a study by Nozawa et al. (A-11) was to determine the effectiveness of segmental wire fixation in athletes with spondylolysis. Between 1993 and 2000, 20 athletes (6 women and 14 men)

with lumbar spondylolysis were treated surgically with the technique. The following table gives the Japanese Orthopaedic Association (JOA) evaluation score for lower back pain syndrome for men and women prior to the surgery. The lower score indicates less pain.

Gender	JOA scores
Female	14, 13, 24, 21, 20, 21
Male	21, 26, 24, 24, 22, 23, 18, 24, 13, 22, 25, 23, 21, 25

Source: Satoshi Nozawa, Katsuji Shimizu, Kei Miyamoto, and Mizuo Tanaka, "Repair of Pars Interarticularis Defect by Segmental Wire Fixation in Young Athletes with Spondylolysis," *American Journal of Sports Medicine*, 31 (2003), 359–364.

- 6.4.5.** Krantz et al. (A-12) investigated dose-related effects of methadone in subjects with torsade de pointes, a polymorphic ventricular tachycardia. In the study of 17 subjects, nine were being treated with methadone for opiate dependency and eight for chronic pain. The mean daily dose of methadone in the opiate dependency group was 541 mg/day with a standard deviation of 156, while the chronic pain group received a mean dose of 269 mg/day with a standard deviation of 316.
- 6.4.6.** Transverse diameter measurements on the hearts of adult males and females gave the following results:

Group	Sample Size	\bar{x} (cm)	s (cm)
Males	12	13.21	1.05
Females	9	11.00	1.01

Assume normally distributed populations with equal variances.

- 6.4.7.** Twenty-four experimental animals with vitamin D deficiency were divided equally into two groups. Group 1 received treatment consisting of a diet that provided vitamin D. The second group was not treated. At the end of the experimental period, serum calcium determinations were made with the following results:

$$\begin{aligned} \text{Treated group: } & \bar{x} = 11.1 \text{ mg}/100 \text{ ml}, s = 1.5 \\ \text{Untreated group: } & \bar{x} = 7.8 \text{ mg}/100 \text{ ml}, s = 2.0 \end{aligned}$$

Assume normally distributed populations with equal variances.

- 6.4.8.** Two groups of children were given visual acuity tests. Group 1 was composed of 11 children who receive their health care from private physicians. The mean score for this group was 26 with a standard deviation of 5. Group 2 was composed of 14 children who receive their health care from the health department, and had an average score of 21 with a standard deviation of 6. Assume normally distributed populations with equal variances.
- 6.4.9.** The average length of stay of a sample of 20 patients discharged from a general hospital was 7 days with a standard deviation of 2 days. A sample of 24 patients discharged from a chronic disease hospital had an average length of stay of 36 days with a standard deviation of 10 days. Assume normally distributed populations with unequal variances.
- 6.4.10.** In a study of factors thought to be responsible for the adverse effects of smoking on human reproduction, cadmium level determinations (nanograms per gram) were made on placenta tissue of a

sample of 14 mothers who were smokers and an independent random sample of 18 nonsmoking mothers. The results were as follows:

Nonsmokers: 10.0, 8.4, 12.8, 25.0, 11.8, 9.8, 12.5, 15.4, 23.5,
9.4, 25.1, 19.5, 25.5, 9.8, 7.5, 11.8, 12.2, 15.0

Smokers: 30.0, 30.1, 15.0, 24.1, 30.5, 17.8, 16.8, 14.8,
13.4, 28.5, 17.5, 14.4, 12.5, 20.4

Does it appear likely that the mean cadmium level is higher among smokers than nonsmokers? Why do you reach this conclusion?

6.5 CONFIDENCE INTERVAL FOR A POPULATION PROPORTION

Many questions of interest to the health worker relate to population proportions. What proportion of patients who receive a particular type of treatment recover? What proportion of some population has a certain disease? What proportion of a population is immune to a certain disease?

To estimate a population proportion we proceed in the same manner as when estimating a population mean. A sample is drawn from the population of interest, and the sample proportion, \hat{p} , is computed. This sample proportion is used as the point estimator of the population proportion. A confidence interval is obtained by the general formula

$$\text{estimator} \pm (\text{reliability coefficient}) \times (\text{standard error of the estimator})$$

In the previous chapter we saw that when both np and $n(1 - p)$ are greater than 5, we may consider the sampling distribution of \hat{p} to be quite close to the normal distribution. When this condition is met, our reliability coefficient is some value of z from the standard normal distribution. The standard error, we have seen, is equal to $\sigma_{\hat{p}} = \sqrt{p(1 - p)/n}$. Since p , the parameter we are trying to estimate, is unknown, we must use \hat{p} as an estimate. Thus, we estimate $\sigma_{\hat{p}}$ by $\sqrt{\hat{p}(1 - \hat{p})/n}$, and our $100(1 - \alpha)$ percent confidence interval for p is given by

$$\hat{p} \pm z_{1-\alpha/2} \sqrt{\hat{p}(1 - \hat{p})/n} \quad (6.5.1)$$

We give this interval both the probabilistic and practical interpretations.

EXAMPLE 6.5.1

The Pew Internet and American Life Project (A-13) reported in 2003 that 18 percent of Internet users have used it to search for information regarding experimental treatments or medicines. The sample consisted of 1220 adult Internet users, and information was collected from telephone interviews. We wish to construct a 95 percent confidence interval for the proportion of Internet users in the sampled population who have searched for information on experimental treatments or medicines.

Solution: We shall assume that the 1220 subjects were sampled in random fashion. The best point estimate of the population proportion is $\hat{p} = .18$. The size of the sample and our estimate of p are of sufficient magnitude to justify use of the standard normal distribution in constructing a confidence interval. The reliability coefficient corresponding to a confidence level of .95 is 1.96, and our estimate of the standard error $\sigma_{\hat{p}}$ is $\sqrt{\hat{p}(1 - \hat{p})/n} = \sqrt{(.18)(.82)/1220} = .0110$. The 95 percent confidence interval for p , based on these data, is

$$.18 \pm 1.96(.0110)$$

$$.18 \pm .022$$

$$.158, .202$$

We are 95 percent confident that the population proportion p is between .158 and .202 because, in repeated sampling, about 95 percent of the intervals constructed in the manner of the present single interval would include the true p . On the basis of these results we would expect, with 95 percent confidence, to find somewhere between 15.8 percent and 20.2 percent of adult Internet users to have used it for information on medicine or experimental treatments. ■

EXERCISES

For each of the following exercises state the practical and probabilistic interpretations of the interval that you construct. Identify each component of the interval: point estimate, reliability coefficient, and standard error. Explain why the reliability coefficients are not the same for all exercises.

- 6.5.1. Luna et al. (A-14) studied patients who were mechanically ventilated in the intensive care unit of six hospitals in Buenos Aires, Argentina. The researchers found that of 472 mechanically ventilated patients, 63 had clinical evidence of ventilator-associated pneumonia (VAP). Construct a 95 percent confidence interval for the proportion of all mechanically ventilated patients at these hospitals who may be expected to develop VAP.
- 6.5.2. Q waves on the electrocardiogram, according to Schinkel et al. (A-15), are often considered to be reflective of irreversibly scarred myocardium. These researchers assert, however, that there are some indications that residual viable tissue may be present in Q-wave-infarcted regions. Their study of 150 patients with chronic electrocardiographic Q-wave infarction found 202 dysfunctional Q-wave regions. With dobutamine stress echocardiography (DSE), they noted that 118 of these 202 regions were viable with information from the DSE testing. Construct a 90 percent confidence interval for the proportion of viable regions that one might expect to find a population of dysfunctional Q-wave regions.
- 6.5.3. In a study by von zur Muhlen et al. (A-16), 136 subjects with syncope or near syncope were studied. Syncope is the temporary loss of consciousness due to a sudden decline in blood flow to the brain. Of these subjects, 75 also reported having cardiovascular disease. Construct a 99 percent confidence interval for the population proportion of subjects with syncope or near syncope who also have cardiovascular disease.
- 6.5.4. In a simple random sample of 125 unemployed male high-school dropouts between the ages of 16 and 21, inclusive, 88 stated that they were regular consumers of alcoholic beverages. Construct a 95 percent confidence interval for the population proportion.

6.6 CONFIDENCE INTERVAL FOR THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS

The magnitude of the difference between two population proportions is often of interest. We may want to compare, for example, men and women, two age groups, two socioeconomic groups, or two diagnostic groups with respect to the proportion possessing some characteristic of interest. An unbiased point estimator of the difference between two population proportions is provided by the difference between sample proportions, $\hat{p}_1 - \hat{p}_2$. As we have seen, when n_1 and n_2 are large and the population proportions are not too close to 0 or 1, the central limit theorem applies and normal distribution theory may be employed to obtain confidence intervals. The standard error of the estimate usually must be estimated by

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}$$

because, as a rule, the population proportions are unknown. A $100(1 - \alpha)$ percent confidence interval for $p_1 - p_2$ is given by

$$(\hat{p}_1 - \hat{p}_2) \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} \quad (6.6.1)$$

We may interpret this interval from both the probabilistic and practical points of view.

EXAMPLE 6.6.1

Connor et al. (A-17) investigated gender differences in proactive and reactive aggression in a sample of 323 children and adolescents (68 females and 255 males). The subjects were from unsolicited consecutive referrals to a residential treatment center and a pediatric psychopharmacology clinic serving a tertiary hospital and medical school. In the sample, 31 of the females and 53 of the males reported sexual abuse. We wish to construct a 99 percent confidence interval for the difference between the proportions of sexual abuse in the two sampled populations.

Solution: The sample proportions for the females and males are, respectively, $\hat{p}_F = 31/68 = .4559$ and $\hat{p}_M = 53/255 = .2078$. The difference between sample proportions is $\hat{p}_F - \hat{p}_M = .4559 - .2078 = .2481$. The estimated standard error of the difference between sample proportions is

$$\begin{aligned} \hat{\sigma}_{\hat{p}_F - \hat{p}_M} &= \sqrt{\frac{(.4559)(.5441)}{68} + \frac{(.2078)(.7922)}{255}} \\ &= .0655 \end{aligned}$$

The reliability factor from Appendix Table D is 2.58, so that our confidence interval, by Expression 6.6.1, is

$$\begin{aligned} &.2481 \pm 2.58(.0655) \\ &.0791, .4171 \end{aligned}$$

We are 99 percent confident that for the sampled populations, the proportion of cases of reported sexual abuse among females exceeds the proportion of cases of reported sexual abuse among males by somewhere between .0791 and .4171.

Since the interval does not include zero, we conclude that the two population proportions are not equal. ■

EXERCISES

For each of the following exercises state the practical and probabilistic interpretations of the interval that you construct. Identify each component of the interval: point estimate, reliability coefficient, and standard error. Explain why the reliability coefficients are not the same for all exercises.

- 6.6.1.** Horwitz et al. (A-18) studied 637 persons who were identified by court records from 1967 to 1971 as having experienced abuse or neglect. For a control group, they located 510 subjects who as children attended the same elementary school and lived within a five-block radius of those in the abused/neglected group. In the abused/neglected group, and control group, 114 and 57 subjects, respectively, had developed antisocial personality disorders over their lifetimes. Construct a 95 percent confidence interval for the difference between the proportions of subjects developing antisocial personality disorders one might expect to find in the populations of subjects from which the subjects of this study may be presumed to have been drawn.
- 6.6.2.** The objective of a randomized controlled trial by Adab et al. (A-19) was to determine whether providing women with additional information on the pros and cons of screening for cervical cancer would increase the willingness to be screened. A treatment group of 138 women received a leaflet on screening that contained more information (average individual risk for cervical cancer, likelihood of positive finding, the possibility of false positive/negative results, etc.) than the standard leaflet developed by the British National Health Service that 136 women in a control group received. In the treatment group, 109 women indicated they wanted to have the screening test for cervical cancer while in the control group, 120 indicated they wanted the screening test. Construct a 95 percent confidence interval for the difference in proportions for the two populations represented by these samples.
- 6.6.3.** Spertus et al. (A-20) performed a randomized single blind study for subjects with stable coronary artery disease. They randomized subjects into two treatment groups. The first group had current angina medications optimized, and the second group was tapered off existing medications and then started on long-acting diltiazem at 180 mg/day. The researchers performed several tests to determine if there were significant differences in the two treatment groups at baseline. One of the characteristics of interest was the difference in the percentages of subjects who had reported a history of congestive heart failure. In the group where current medications were optimized, 16 of 49 subjects reported a history of congestive heart failure. In the subjects placed on the diltiazem, 12 of the 51 subjects reported a history of congestive heart failure. State the assumptions that you think are necessary and construct a 95 percent confidence interval for the difference between the proportions of those reporting congestive heart failure within the two populations from which we presume these treatment groups to have been selected.
- 6.6.4.** To study the difference in drug therapy adherence among subjects with depression who received usual care and those who received care in a collaborative care model was the goal of a study conducted by Finley et al. (A-21). The collaborative care model emphasized the role of clinical pharmacists in providing drug therapy management and treatment follow-up. Of the 50 subjects receiving usual care, 24 adhered to the prescribed drug regimen, while 50 out of 75 subjects in the collaborative care model

adhered to the drug regimen. Construct a 90 percent confidence interval for the difference in adherence proportions for the populations of subjects represented by these two samples.

6.7 DETERMINATION OF SAMPLE SIZE FOR ESTIMATING MEANS

The question of how large a sample to take arises early in the planning of any survey or experiment. This is an important question that should not be treated lightly. To take a larger sample than is needed to achieve the desired results is wasteful of resources, whereas very small samples often lead to results that are of no practical use. Let us consider, then, how one may go about determining the sample size that is needed in a given situation. In this section, we present a method for determining the sample size required for estimating a population mean, and in the next section we apply this method to the case of sample size determination when the parameter to be estimated is a population proportion. By straightforward extensions of these methods, sample sizes required for more complicated situations can be determined.

Objectives The objectives in interval estimation are to obtain narrow intervals with high reliability. If we look at the components of a confidence interval, we see that the width of the interval is determined by the magnitude of the quantity

$$(\text{reliability coefficient}) \times (\text{standard error of the estimator})$$

since the total width of the interval is twice this amount. We have learned that this quantity is usually called the precision of the estimate or the margin of error. For a given standard error, increasing reliability means a larger reliability coefficient. But a larger reliability coefficient for a fixed standard error makes for a wider interval.

On the other hand, if we fix the reliability coefficient, the only way to reduce the width of the interval is to reduce the standard error. Since the standard error is equal to σ/\sqrt{n} , and since σ is a constant, the only way to obtain a small standard error is to take a large sample. How large a sample? That depends on the size of σ , the population standard deviation, the desired degree of reliability, and the desired interval width.

Let us suppose we want an interval that extends d units on either side of the estimator. We can write

$$d = (\text{reliability coefficient}) \times (\text{standard error of the estimator}) \quad (6.7.1)$$

If sampling is to be with replacement, from an infinite population, or from a population that is sufficiently large to warrant our ignoring the finite population correction, Equation 6.7.1 becomes

$$d = z \frac{\sigma}{\sqrt{n}} \quad (6.7.2)$$

which, when solved for n , gives

$$n = \frac{z^2 \sigma^2}{d^2} \quad (6.7.3)$$

When sampling is without replacement from a small finite population, the finite population correction is required and Equation 6.7.1 becomes

$$d = z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \quad (6.7.4)$$

which, when solved for n , gives

$$n = \frac{Nz^2\sigma^2}{d^2(N-1) + z^2\sigma^2} \quad (6.7.5)$$

If the finite population correction can be ignored, Equation 6.7.5 reduces to Equation 6.7.3.

Estimating σ^2 The formulas for sample size require knowledge of σ^2 but, as has been pointed out, the population variance is, as a rule, unknown. As a result, σ^2 has to be estimated. The most frequently used sources of estimates for σ^2 are the following:

1. A *pilot* or preliminary sample may be drawn from the population, and the variance computed from this sample may be used as an estimate of σ^2 . Observations used in the pilot sample may be counted as part of the final sample, so that n (the computed sample size) $- n_1$ (the pilot sample size) $= n_2$ (the number of observations needed to satisfy the total sample size requirement).
2. Estimates of σ^2 may be available from previous or similar studies.
3. If it is thought that the population from which the sample is to be drawn is approximately normally distributed, one may use the fact that the range is approximately equal to six standard deviations and compute $\sigma \approx R/6$. This method requires some knowledge of the smallest and largest value of the variable in the population.

EXAMPLE 6.7.1

A health department nutritionist, wishing to conduct a survey among a population of teenage girls to determine their average daily protein intake (measured in grams), is seeking the advice of a biostatistician relative to the sample size that should be taken.

What procedure does the biostatistician follow in providing assistance to the nutritionist? Before the statistician can be of help to the nutritionist, the latter must provide three items of information: (1) the desired width of the confidence interval, (2) the level of confidence desired, and (3) the magnitude of the population variance.

Solution: Let us assume that the nutritionist would like an interval about 10 grams wide; that is, the estimate should be within about 5 grams of the population mean in either direction. In other words, a margin of error of 5 grams is desired. Let us also assume that a confidence coefficient of .95 is decided on and that, from past experience, the nutritionist feels that the population standard deviation is probably about 20 grams. The statistician now has

the necessary information to compute the sample size: $z = 1.96$, $\sigma = 20$ and $d = 5$. Let us assume that the population of interest is large so that the statistician may ignore the finite population correction and use Equation 6.7.3. On making proper substitutions, the value of n is found to be

$$\begin{aligned} n &= \frac{(1.96)^2(20)^2}{(5)^2} \\ &= 61.47 \end{aligned}$$

The nutritionist is advised to take a sample of size 62. When calculating a sample size by Equation 6.7.3 or Equation 6.7.5, we round up to the next-largest whole number if the calculations yield a number that is not itself an integer. ■

EXERCISES

- 6.7.1. A hospital administrator wishes to estimate the mean weight of babies born in her hospital. How large a sample of birth records should be taken if she wants a 99 percent confidence interval that is 1 pound wide? Assume that a reasonable estimate of σ is 1 pound. What sample size is required if the confidence coefficient is lowered to .95?
- 6.7.2. The director of the rabies control section in a city health department wishes to draw a sample from the department's records of dog bites reported during the past year in order to estimate the mean age of persons bitten. He wants a 95 percent confidence interval, he will be satisfied to let $d = 2.5$, and from previous studies he estimates the population standard deviation to be about 15 years. How large a sample should be drawn?
- 6.7.3. A physician would like to know the mean fasting blood glucose value (milligrams per 100 ml) of patients seen in a diabetes clinic over the past 10 years. Determine the number of records the physician should examine in order to obtain a 90 percent confidence interval for μ if the desired width of the interval is 6 units and a pilot sample yields a variance of 60.
- 6.7.4. For multiple sclerosis patients we wish to estimate the mean age at which the disease was first diagnosed. We want a 95 percent confidence interval that is 10 years wide. If the population variance is 90, how large should our sample be?

6.8 DETERMINATION OF SAMPLE SIZE FOR ESTIMATING PROPORTIONS

The method of sample size determination when a population proportion is to be estimated is essentially the same as that described for estimating a population mean. We make use of the fact that one-half the desired interval, d , may be set equal to the product of the reliability coefficient and the standard error.

Assuming that random sampling and conditions warranting approximate normality of the distribution of \hat{p} leads to the following formula for n when sampling is with

replacement, when sampling is from an infinite population, or when the sampled population is large enough to make use of the finite population correction unnecessary,

$$n = \frac{z^2 pq}{d^2} \quad (6.8.1)$$

where $q = 1 - p$.

If the finite population correction cannot be disregarded, the proper formula for n is

$$n = \frac{Nz^2 pq}{d^2(N - 1) + z^2 pq} \quad (6.8.2)$$

When N is large in comparison to n (that is, $n/N \leq .05$ the finite population correction may be ignored, and Equation 6.8.2 reduces to Equation 6.8.1.

Estimating p As we see, both formulas require knowledge of p , the proportion in the population possessing the characteristic of interest. Since this is the parameter we are trying to estimate, it, obviously, will be unknown. One solution to this problem is to take a pilot sample and compute an estimate to be used in place of p in the formula for n . Sometimes an investigator will have some notion of an upper bound for p that can be used in the formula. For example, if it is desired to estimate the proportion of some population who have a certain disability, we may feel that the true proportion cannot be greater than, say, .30. We then substitute .30 for p in the formula for n . If it is impossible to come up with a better estimate, one may set p equal to .5 and solve for n . Since $p = .5$ in the formula yields the maximum value of n , this procedure will give a large enough sample for the desired reliability and interval width. It may, however, be larger than needed and result in a more expensive sample than if a better estimate of p had been available. This procedure should be used only if one is unable to arrive at a better estimate of p .

EXAMPLE 6.8.1

A survey is being planned to determine what proportion of families in a certain area are medically indigent. It is believed that the proportion cannot be greater than .35. A 95 percent confidence interval is desired with $d = .05$. What size sample of families should be selected?

Solution: If the finite population correction can be ignored, we have

$$n = \frac{(1.96)^2 (.35)(.65)}{(.05)^2} = 349.59$$

The necessary sample size, then, is 350. ■

EXERCISES

- 6.8.1.** An epidemiologist wishes to know what proportion of adults living in a large metropolitan area have subtype A hepatitis B virus. Determine the sample size that would be required to estimate the true proportion to within .03 with 95 percent confidence. In a similar metropolitan area the proportion of adults with the characteristic is reported to be .20. If data from another metropolitan area were not available and a pilot sample could not be drawn, what sample size would be required?
- 6.8.2.** A survey is planned to determine what proportion of the high-school students in a metropolitan school system have regularly smoked marijuana. If no estimate of p is available from previous studies, a pilot sample cannot be drawn, a confidence coefficient of .95 is desired, and $d = .04$ is to be used, determine the appropriate sample size. What sample size would be required if 99 percent confidence were desired?
- 6.8.3.** A hospital administrator wishes to know what proportion of discharged patients is unhappy with the care received during hospitalization. How large a sample should be drawn if we let $d = .05$, the confidence coefficient is .95, and no other information is available? How large should the sample be if p is approximated by .25?
- 6.8.4.** A health planning agency wishes to know, for a certain geographic region, what proportion of patients admitted to hospitals for the treatment of trauma die in the hospital. A 95 percent confidence interval is desired, the width of the interval must be .06, and the population proportion, from other evidence, is estimated to be .20. How large a sample is needed?

6.9 CONFIDENCE INTERVAL FOR THE VARIANCE OF A NORMALLY DISTRIBUTED POPULATION

Point Estimation of the Population Variance In previous sections it has been suggested that when a population variance is unknown, the sample variance may be used as an estimator. You may have wondered about the quality of this estimator. We have discussed only one criterion of quality—unbiasedness—so let us see if the sample variance is an unbiased estimator of the population variance. To be unbiased, the average value of the sample variance over all possible samples must be equal to the population variance. That is, the expression $E(s^2) = \sigma^2$ must hold. To see if this condition holds for a particular situation, let us refer to the example of constructing a sampling distribution given in Section 5.3. In Table 5.3.1 we have all possible samples of size 2 from the population consisting of the values 6, 8, 10, 12, and 14. It will be recalled that two measures of dispersion for this population were computed as follows:

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = 8 \quad \text{and} \quad S^2 = \frac{\sum (x_i - \mu)^2}{N - 1} = 10$$

If we compute the sample variance $s^2 = \sum (x_i - \bar{x})^2 / (n - 1)$ for each of the possible samples shown in Table 5.3.1, we obtain the sample variances shown in Table 6.9.1.

TABLE 6.9.1 Variances Computed from Samples Shown in Table 5.3.1

		Second Draw				
		6	8	10	12	14
First Draw	6	0	2	8	18	32
	8	2	0	2	8	18
	10	8	2	0	2	8
	12	18	8	2	0	2
	14	32	18	8	2	0

Sampling with Replacement If sampling is with replacement, the expected value of s^2 is obtained by taking the mean of all sample variances in Table 6.9.1. When we do this, we have

$$E(s^2) = \frac{\sum s_i^2}{N^n} = \frac{0 + 2 + \cdots + 2 + 0}{25} = \frac{200}{25} = 8$$

and we see, for example, that when sampling is with replacement $E(s^2) = \sigma^2$, where $s^2 = \sum (x_i - \bar{x})^2 / (n - 1)$ and $\sigma^2 = \sum (x_i - \mu)^2 / N$.

Sampling Without Replacement If we consider the case where sampling is without replacement, the expected value of s^2 is obtained by taking the mean of all variances above (or below) the principal diagonal. That is,

$$E(s^2) = \frac{\sum s_i^2}{{}_N C_n} = \frac{2 + 8 + \cdots + 2}{10} = \frac{100}{10} = 10$$

which, we see, is not equal to σ^2 , but is equal to $S^2 = \sum (x_i - \mu)^2 / (N - 1)$.

These results are examples of general principles, as it can be shown that, in general,

$$\begin{aligned} E(s^2) &= \sigma^2 && \text{when sampling is with replacement} \\ E(s^2) &= S^2 && \text{when sampling is without replacement} \end{aligned}$$

When N is large, $N - 1$ and N will be approximately equal and, consequently, σ^2 and S^2 will be approximately equal.

These results justify our use of $s^2 = \sum (x_i - \bar{x})^2 / (n - 1)$ when computing the sample variance. In passing, let us note that although s^2 is an unbiased estimator of σ^2 , s is not an unbiased estimator of σ . The bias, however, diminishes rapidly as n increases.

Interval Estimation of a Population Variance With a point estimate available, it is logical to inquire about the construction of a confidence interval for a population variance. Whether we are successful in constructing a confidence interval for σ^2 will depend on our ability to find an appropriate sampling distribution.

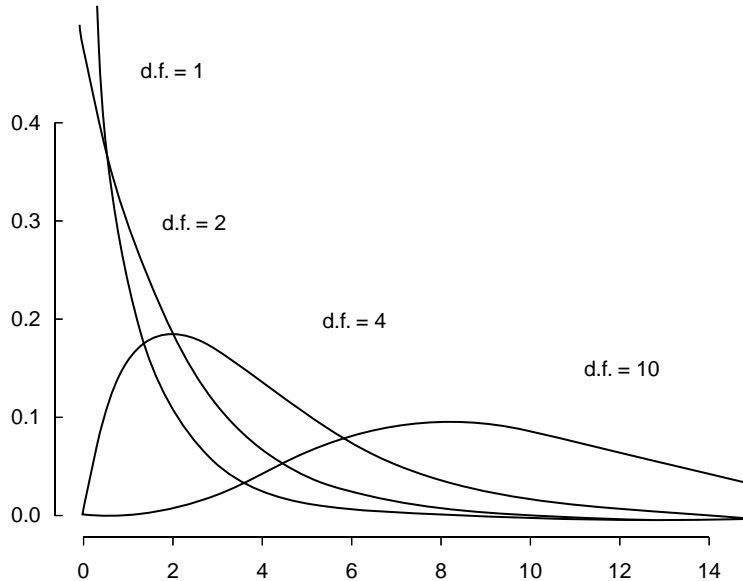


FIGURE 6.9.1 Chi-square distributions.

(Source: Gerald van Belle, Lloyd D. Fisher, Patrick J. Heagerty, and Thomas Lumley, *Biostatistics: A Methodology for the Health Sciences*, 2nd Ed., © 2004 John Wiley & Sons, Inc. This material is reproduced with permission of John Wiley & Sons, Inc.)

The Chi-Square Distribution Confidence intervals for σ^2 are usually based on the sampling distribution of $(n - 1)s^2/\sigma^2$. If samples of size n are drawn from a normally distributed population, this quantity has a distribution known as the *chi-square* (χ^2) *distribution* with $n - 1$ degrees of freedom. As we will say more about this distribution in chapter 12, we only say here that it is the distribution that the quantity $(n - 1)s^2/\sigma^2$ follows and that it is useful in finding confidence intervals for σ^2 when the assumption that the population is normally distributed holds true.

Figure 6.9.1 shows chi-square distributions for several values of degrees of freedom. Percentiles of the chi-square distribution are given in Appendix Table F. The column headings give the values of χ^2 to the left of which lies a proportion of the total area under the curve equal to the subscript of χ^2 . The row labels are the degrees of freedom.

To obtain a $100(1 - \alpha)$ percent confidence interval for σ^2 , we first obtain the $100(1 - \alpha)$ percent confidence interval for $(n - 1)s^2/\sigma^2$. To do this, we select the values of χ^2 from Appendix Table F in such a way that $\alpha/2$ is to the left of the smaller value and $\alpha/2$ is to the right of the larger value. In other words, the two values of χ^2 are selected in such a way that α is divided equally between the two tails of the distribution. We may designate these two values of χ^2 as $\chi^2_{\alpha/2}$ and $\chi^2_{1-(\alpha/2)}$, respectively. The $100(1 - \alpha)$ percent confidence interval for $(n - 1)s^2/\sigma^2$, then, is given by

$$\chi^2_{\alpha/2} < \frac{(n - 1)s^2}{\sigma^2} < \chi^2_{1-(\alpha/2)}$$

We now manipulate this expression in such a way that we obtain an expression with σ^2 alone as the middle term. First, let us divide each term by $(n - 1)s^2$ to get

$$\frac{\chi_{\alpha/2}^2}{(n - 1)s^2} < \frac{1}{\sigma^2} < \frac{\chi_{1-(\alpha/2)}^2}{(n - 1)s^2}$$

If we take the reciprocal of this expression, we have

$$\frac{(n - 1)s^2}{\chi_{\alpha/2}^2} > \sigma^2 > \frac{(n - 1)s^2}{\chi_{1-(\alpha/2)}^2}$$

Note that the direction of the inequalities changed when we took the reciprocals. If we reverse the order of the terms, we have

$$\frac{(n - 1)s^2}{\chi_{1-(\alpha/2)}^2} < \sigma^2 < \frac{(n - 1)s^2}{\chi_{\alpha/2}^2} \quad (6.9.1)$$

which is the $100(1 - \alpha)$ percent confidence interval for σ^2 . If we take the square root of each term in Expression 6.9.1, we have the following $100(1 - \alpha)$ percent confidence interval for σ , the population standard deviation:

$$\sqrt{\frac{(n - 1)s^2}{\chi_{1-(\alpha/2)}^2}} < \sigma < \sqrt{\frac{(n - 1)s^2}{\chi_{\alpha/2}^2}} \quad (6.9.2)$$

EXAMPLE 6.9.1

In a study of the effectiveness of a gluten-free diet in first-degree relatives of patients with type I diabetics, Hummel et al. (A-22) placed seven subjects on a gluten-free diet for 12 months. Prior to the diet, they took baseline measurements of several antibodies and autoantibodies, one of which was the diabetes related insulin autoantibody (IAA). The IAA levels were measured by radiobinding assay. The seven subjects had IAA units of

9.7, 12.3, 11.2, 5.1, 24.8, 14.8, 17.7

We wish to estimate from the data in this sample the variance of the IAA units in the population from which the sample was drawn and construct a 95 percent confidence interval for this estimate.

Solution: The sample yielded a value of $s^2 = 39.763$. The degrees of freedom are $n - 1 = 6$. The appropriate values of χ^2 from Appendix Table F are

$\chi_{1-(\alpha/2)}^2 = 14.449$ and $\chi_{\alpha/2}^2 = 1.237$. Our 95 percent confidence interval for σ^2 is

$$\frac{6(39.763)}{14.449} < \sigma^2 < \frac{6(39.763)}{1.237}$$

$$16.512 < \sigma^2 < 192.868$$

The 95 percent confidence interval for σ is

$$4.063 < \sigma < 13.888$$

We are 95 percent confident that the parameters being estimated are within the specified limits, because we know that in the long run, in repeated sampling, 95 percent of intervals constructed as illustrated would include the respective parameters. ■

Some Precautions Although this method of constructing confidence intervals for σ^2 is widely used, it is not without its drawbacks. First, the assumption of the normality of the population from which the sample is drawn is crucial, and results may be misleading if the assumption is ignored.

Another difficulty with these intervals results from the fact that the estimator is not in the center of the confidence interval, as is the case with the confidence interval for μ . This is because the chi-square distribution, unlike the normal, is not symmetric. The practical implication of this is that the method for the construction of confidence intervals for σ^2 , which has just been described, does not yield the shortest possible confidence intervals. Tate and Klett (12) give tables that may be used to overcome this difficulty.

EXERCISES

- 6.9.1.** A study by Aizenberg et al. (A-23) examined the efficacy of sildenafil, a potent phosphodiesterase inhibitor, in the treatment of elderly men with erectile dysfunction induced by antidepressant treatment for major depressive disorder. The ages of the 10 enrollees in the study were

74, 81, 70, 70, 74, 77, 76, 70, 71, 72

Assume that the subjects in this sample constitute a simple random sample drawn from a population of similar subjects. Construct a 95 percent confidence interval for the variance of the ages of subjects in the population.

- 6.9.2.** Borden et al. (A-24) performed experiments on cadaveric knees to test the effectiveness of several meniscal repair techniques. Specimens were loaded into a servohydraulic device and tension-loaded to failure. The biomechanical testing was performed by using a slow loading rate to simulate the stresses that the medial meniscus might be subjected to during early rehabilitation exercises and activities of daily living. One of the measures is the amount of displacement that occurs. Of the 12 specimens receiving the vertical mattress suture and the Fast-FIX method, the displacement values

measured in millimeters are 16.9, 20.2, 20.1, 15.7, 13.9, 14.9, 18.0, 18.5, 9.2, 18.8, 22.8, 17.5. Construct a 90 percent confidence interval for the variance of the displacement in millimeters for a population of subjects receiving these repair techniques.

- 6.9.3.** Forced vital capacity determinations were made on 20 healthy adult males. The sample variance was 1,000,000. Construct 90 percent confidence intervals for σ^2 and σ .
- 6.9.4.** In a study of myocardial transit times, appearance transit times were obtained on a sample of 30 patients with coronary artery disease. The sample variance was found to be 1.03. Construct 99 percent confidence intervals for σ^2 and σ .
- 6.9.5.** A sample of 25 physically and mentally healthy males participated in a sleep experiment in which the percentage of each participant's total sleeping time spent in a certain stage of sleep was recorded. The variance computed from the sample data was 2.25. Construct 95 percent confidence intervals for σ^2 and σ .
- 6.9.6.** Hemoglobin determinations were made on 16 animals exposed to a harmful chemical. The following observations were recorded: 15.6, 14.8, 14.4, 16.6, 13.8, 14.0, 17.3, 17.4, 18.6, 16.2, 14.7, 15.7, 16.4, 13.9, 14.8, 17.5. Construct 95 percent confidence intervals for σ^2 and σ .
- 6.9.7.** Twenty air samples taken at the same site over a period of 6 months showed the following amounts of suspended particulate matter (micrograms per cubic meter of air):

68	22	36	32
42	24	28	38
30	44	28	27
28	43	45	50
79	74	57	21

Consider these measurements to be a random sample from a population of normally distributed measurements, and construct a 95 percent confidence interval for the population variance.

6.10 CONFIDENCE INTERVAL FOR THE RATIO OF THE VARIANCES OF TWO NORMALLY DISTRIBUTED POPULATIONS

It is frequently of interest to compare two variances, and one way to do this is to form their ratio, σ_1^2/σ_2^2 . If two variances are equal, their ratio will be equal to 1. We usually will not know the variances of populations of interest, and, consequently, any comparisons we make will be based on sample variances. In other words, we may wish to estimate the ratio of two population variances. We learned in Section 6.4 that the valid use of the t distribution to construct a confidence interval for the difference between two population means requires that the population variances be equal. The use of the ratio of two population variances for determining equality of variances has been formalized into a statistical test. The distribution of this test provides test values for determining if the ratio exceeds the value 1 to a large enough extent that we may conclude that the variances are not equal. The test is referred to as the F-max Test by Hartley (13) or the Variance Ratio Test by Zar (14). Many computer programs provide some formalized test of the equality of variances so that the assumption of equality of variances associated with many of the tests in the following chapters can be

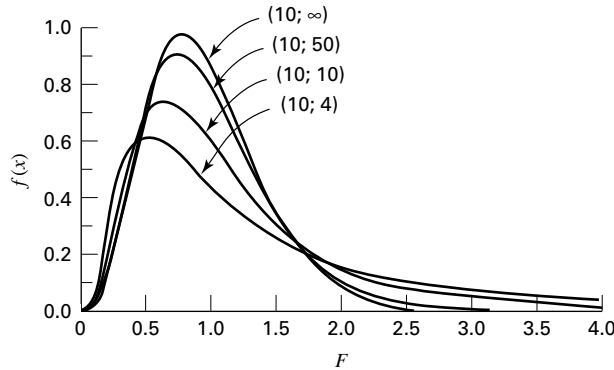


FIGURE 6.10.1 The F distribution for various degrees of freedom.
 (From *Documenta Geigy, Scientific Tables*, Seventh Edition, 1970. Courtesy of Ciba-Geigy Limited, Basel, Switzerland.)

examined. If the confidence interval for the ratio of two population variances includes 1, we conclude that the two population variances may, in fact, be equal. Again, since this is a form of inference, we must rely on some sampling distribution, and this time the distribution of $(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$ is utilized provided certain assumptions are met. The assumptions are that s_1^2 and s_2^2 are computed from independent samples of size n_1 and n_2 respectively, drawn from two normally distributed populations. We use s_1^2 to designate the larger of the two sample variances.

The F Distribution If the assumptions are met, $(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$ follows a distribution known as the F distribution. We defer a more complete discussion of this distribution until chapter 8, but note that this distribution depends on two-degrees-of-freedom values, one corresponding to the value of $n_1 - 1$ used in computing s_1^2 and the other corresponding to the value of $n_2 - 1$ used in computing s_2^2 . These are usually referred to as the *numerator degrees of freedom* and the *denominator degrees of freedom*. Figure 6.10.1 shows some F distributions for several numerator and denominator degrees-of-freedom combinations. Appendix Table G contains, for specified combinations of degrees of freedom and values of α , F values to the right of which lies $\alpha/2$ of the area under the curve of F .

A Confidence Interval for σ_1^2/σ_2^2 To find the $100(1 - \alpha)$ percent confidence interval for σ_1^2/σ_2^2 , we begin with the expression

$$F_{\alpha/2} < \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} < F_{1-(\alpha/2)}$$

where $F_{\alpha/2}$ and $F_{1-(\alpha/2)}$ are the values from the F table to the left and right of which, respectively, lies $\alpha/2$ of the area under the curve. The middle term of this expression may be rewritten so that the entire expression is

$$F_{\alpha/2} < \frac{s_1^2}{s_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} < F_{1-(\alpha/2)}$$

If we divide through by s_1^2/s_2^2 , we have

$$\frac{F_{\alpha/2}}{s_1^2/s_2^2} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{F_{1-(\alpha/2)}}{s_1^2/s_2^2}$$

Taking the reciprocals of the three terms gives

$$\frac{s_1^2/s_2^2}{F_{\alpha/2}} > \frac{\sigma_1^2}{\sigma_2^2} > \frac{s_1^2/s_2^2}{F_{1-(\alpha/2)}}$$

and if we reverse the order, we have the following $100(1 - \alpha)$ percent confidence interval for σ_1^2/σ_2^2 :

$$\frac{s_1^2/s_2^2}{F_{1-(\alpha/2)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2/s_2^2}{F_{\alpha/2}} \quad (6.10.1)$$

EXAMPLE 6.10.1

Allen and Gross (A-25) examine toe flexors strength in subjects with plantar fasciitis (pain from heel spurs, or general heel pain), a common condition in patients with musculo-skeletal problems. Inflammation of the plantar fascia is often costly to treat and frustrating for both the patient and the clinician. One of the baseline measurements was the body mass index (BMI). For the 16 women in the study, the standard deviation for BMI was 8.1 and for four men in the study, the standard deviation was 5.9. We wish to construct a 95 percent confidence interval for the ratio of the variances of the two populations from which we presume these samples were drawn.

Solution: We have the following information:

$$\begin{aligned} n_1 &= 16 & n_2 &= 4 \\ s_1^2 &= (8.1)^2 = 65.61 & s_2^2 &= (5.9)^2 = 34.81 \\ df_1 &= \text{numerator degrees of freedom} = n_1 - 1 = 15 \\ df_2 &= \text{denominator degrees of freedom} = n_2 - 1 = 3 \\ \alpha &= .05 \\ F_{.025} &= .24096 & F_{.975} &= 14.25 \end{aligned}$$

We are now ready to obtain our 95 percent confidence interval for σ_1^2/σ_2^2 by substituting appropriate values into Expression 6.10.1:

$$\begin{aligned} \frac{65.61/34.81}{14.25} &< \frac{\sigma_1^2}{\sigma_2^2} < \frac{65.61/34.81}{.24096} \\ .1323 &< \frac{\sigma_1^2}{\sigma_2^2} < 7.8221 \end{aligned}$$

We give this interval the appropriate probabilistic and practical interpretations.

Since the interval .1323 to 7.8221 includes 1, we are able to conclude that the two population variances may be equal. ■

Finding $F_{1-(\alpha/2)}$ and $F_{\alpha/2}$ At this point we must make a cumbersome, but unavoidable, digression and explain how the values $F_{.975} = 14.25$ and $F_{.025} = .24096$ were obtained. The value of $F_{.975}$ at the intersection of the column headed $df_1 = 15$ and the row labeled $df_2 = 3$ is 14.25. If we had a more extensive table of the F distribution, finding $F_{.025}$ would be no trouble; we would simply find $F_{.025}$ as we found $F_{.975}$. We would take the value at the intersection of the column headed 15 and the row headed 3. To include every possible percentile of F would make for a very lengthy table. Fortunately, however, there exists a relationship that enables us to compute the lower percentile values from our limited table. The relationship is as follows:

$$F_{\alpha,df_1,df_2} = \frac{1}{F_{1-\alpha,df_2,df_1}} \quad (6.10.2)$$

We proceed as follows.

Interchange the numerator and denominator degrees of freedom and locate the appropriate value of F . For the problem at hand we locate 4.15, which is at the intersection of the column headed 3 and the row labeled 15. We now take the reciprocal of this value, $1/4.15 = .24096$. In summary, the lower confidence limit (LCL) and upper confidence limit (UCL) σ_1^2/σ_2^2 are as follows:

$$LCL = \frac{s_1^2}{s_2^2} \frac{1}{F_{(1-\alpha/2),df_1,df_2}}$$

$$UCL = \frac{s_1^2}{s_2^2} F_{1-(\alpha/2),df_2,df_1}$$

Alternative procedures for making inferences about the equality of two variances when the sampled populations are not normally distributed may be found in the book by Daniel (15).

Some Precautions Similar to the discussion in the previous section of constructing confidence intervals for σ^2 , the assumption of normality of the populations from which the samples are drawn is crucial to obtaining correct intervals for the ratio of variances discussed in this section. Fortunately, most statistical computer programs provide alternatives to the F -ratio, such as Levene's test, when the underlying distributions cannot be assumed to be normally distributed. Computationally, Levene's test uses a measure of distance from a sample median instead of a sample mean, hence removing the assumption of normality.

EXERCISES

- 6.10.1.** The purpose of a study by Moneim et al. (A-26) was to examine thumb amputations from team roping at rodeos. The researchers reviewed 16 cases of thumb amputations. Of these, 11 were complete amputations while five were incomplete. The ischemia time is the length of time that insufficient oxygen is supplied to the amputated thumb. The ischemia times (hours) for 11 subjects experiencing complete amputations were

4.67, 10.5, 2.0, 3.18, 4.00, 3.5, 3.33, 5.32, 2.0, 4.25, 6.0

For five victims of incomplete thumb amputation, the ischemia times were

3.0, 10.25, 1.5, 5.22, 5.0

Treat the two reported sets of data as sample data from the two populations as described. Construct a 95 percent confidence interval for the ratio of the two unknown population variances.

- 6.10.2.** The objective of a study by Horesh et al. (A-27) was to explore the hypothesis that some forms of suicidal behavior among adolescents are related to anger and impulsivity. The sample consisted of 65 adolescents admitted to a university-affiliated adolescent psychiatric unit. The researchers used the Impulsiveness-Control Scale (ICS, A-28) where higher numbers indicate higher degrees of impulsiveness and scores can range from 0 to 45. The 33 subjects classified as suicidal had an ICS score standard deviation of 8.4 while the 32 nonsuicidal subjects had a standard deviation of 6.0. Assume that these two groups constitute independent simple random samples from two populations of similar subjects. Assume also that the ICS scores in these two populations are normally distributed. Find the 99 percent confidence interval for the ratio of the two population variances of scores on the ICS.
- 6.10.3.** Stroke index values were statistically analyzed for two samples of patients suffering from myocardial infarction. The sample variances were 12 and 10. There were 21 patients in each sample. Construct the 95 percent confidence interval for the ratio of the two population variances.
- 6.10.4.** Thirty-two adult asphasics seeking speech therapy were divided equally into two groups. Group 1 received treatment 1, and group 2 received treatment 2. Statistical analysis of the treatment effectiveness scores yielded the following variances: $s_1^2 = 8, s_2^2 = 15$. Construct the 90 percent confidence interval for σ_2^2/σ_1^2 .
- 6.10.5.** Sample variances were computed for the tidal volumes (milliliters) of two groups of patients suffering from atrial septal defect. The results and sample sizes were as follows:

$$\begin{aligned}n_1 &= 31, s_1^2 = 35,000 \\n_2 &= 41, s_2^2 = 20,000\end{aligned}$$

Construct the 95 percent confidence interval for the ratio of the two population variances.

- 6.10.6.** Glucose responses to oral glucose were recorded for 11 patients with Huntington's disease (group 1) and 13 control subjects (group 2). Statistical analysis of the results yielded the following sample variances: $s_1^2 = 105, s_2^2 = 148$. Construct the 95 percent confidence interval for the ratio of the two population variances.

6.10.7. Measurements of gastric secretion of hydrochloric acid (milliequivalents per hour) in 16 normal subjects and 10 subjects with duodenal ulcer yielded the following results:

Normal subjects: 6.3, 2.0, 2.3, 0.5, 1.9, 3.2, 4.1, 4.0, 6.2, 6.1, 3.5, 1.3, 1.7, 4.5, 6.3, 6.2
 Ulcer subjects: 13.7, 20.6, 15.9, 28.4, 29.4, 18.4, 21.1, 3.0, 26.2, 13.0

Construct a 95 percent confidence interval for the ratio of the two population variances. What assumptions must be met for this procedure to be valid?

6.11 SUMMARY

This chapter is concerned with one of the major areas of statistical inference—estimation. Both point estimation and interval estimation are covered. The concepts and methods involved in the construction of confidence intervals are illustrated for the following parameters: means, the difference between two means, proportions, the difference between two proportions, variances, and the ratio of two variances. In addition, we learned in this chapter how to determine the sample size needed to estimate a population mean and a population proportion at specified levels of precision.

We learned, also, in this chapter that interval estimates of population parameters are more desirable than point estimates because statements of confidence can be attached to interval estimates.

SUMMARY OF FORMULAS FOR CHAPTER 6

Formula Number	Name	Formula
6.2.1	Expression of an interval estimate	estimator \pm (reliability coefficient) \times (standard error of the estimator)
6.2.2	Interval estimate for μ when σ is known	$\bar{x} \pm z_{(1-\alpha/2)}\sigma_{\bar{x}}$
6.3.1	t -transformation	$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$
6.3.2	Interval estimate for μ when σ is unknown	$\bar{x} \pm t_{(1-\alpha/2)} = \frac{s}{\sqrt{n}}$
6.4.1	Interval estimate for the difference between two population means when σ_1 and σ_2 are known	$(\bar{x}_1 - \bar{x}_2) \pm z_{(1-\alpha/2)}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
6.4.2	Pooled variance estimate	$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
6.4.3	Standard error of estimate	$s_{(\bar{x}_1 - \bar{x}_2)} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$

(Continued)

6.4.4	Interval estimate for the difference between two population means when σ_1 is unknown	$(\bar{x}_1 - \bar{x}_2) \pm t_{(1-\alpha/2)} \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$
6.4.5	Cochran's correction for reliability coefficient when variances are not equal	$t'_{(1-\alpha/2)} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$
6.4.6	Interval estimate using Cochran's correction for t	$(\bar{x}_1 - \bar{x}_2) \pm t'_{(1-\alpha/2)} \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
6.5.1	Interval estimate for a population proportion	$\hat{p} \pm z_{(1-\alpha/2)} \sqrt{\hat{p}(1-\hat{p})/n}$
6.6.1	Interval estimate for the difference between two population proportions	$(\hat{p}_1 - \hat{p}_2) \pm z_{(1-\alpha/2)} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
6.7.1–6.7.3	Sample size determination when sampling with replacement	$d = (\text{reliability coefficient}) \times (\text{standard error})$ $d = z \frac{\sigma}{\sqrt{n}}$ \therefore $n = \frac{z^2 \sigma^2}{d^2}$
6.7.4–6.7.5	Sample size determination when sampling without replacement	$d = z \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ \therefore $n = \frac{Nz^2 \sigma^2}{d^2(N-1) + z^2 \sigma^2}$
6.8.1	Sample size determination for proportions when sampling with replacement	$n = \frac{z^2 pq}{d^2}$
6.8.2	Sample size determination for proportions when sampling without replacement	$n = \frac{Nz^2 \sigma^2}{d^2(N-1) + z^2 \sigma^2}$
6.9.1	Interval estimate for σ^2	$\frac{(n-1)s^2}{\chi_{1-(\alpha/2)}^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_{(\alpha/2)}^2}$
6.9.2	Interval estimate for σ	$\sqrt{\frac{(n-1)s^2}{\chi_{1-(\alpha/2)}^2}} < \sigma < \sqrt{\frac{(n-1)s^2}{\chi_{(\alpha/2)}^2}}$
6.10.1	Interval estimate for the ratio of two variances	$\frac{s_1^2/s_2^2}{F_{1-(\alpha/2)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2/s_2^2}{F_{\alpha/2}}$

6.10.2	Relationship among F ratios	$F_{\alpha,df_1,df_2} = \frac{1}{F_{1-\alpha,df_2,df_1}}$
Symbol Key	<ul style="list-style-type: none"> • α = Type 1 error rate • χ^2 = Chi-square distribution • d = error component of interval estimate • df = degrees of freedom • F = F-distribution • μ = mean of population • n = sample size • p = proportion for population • $q = (1 - p)$ • \hat{p} = estimated proportion for sample • σ^2 = population variance • σ = population standard deviation • $\sigma_{\bar{x}}$ = standard error • s = standard deviation of sample • s_p = pooled standard deviation • t = Student's t-transformation • t' = Cochran's correction to t • \bar{x} = mean of sample • z = standard normal distribution 	

REVIEW QUESTIONS AND EXERCISES

1. What is statistical inference?
2. Why is estimation an important type of inference?
3. What is a point estimate?
4. Explain the meaning of unbiasedness.
5. Define the following:

(a) Reliability coefficient	(b) Confidence coefficient	(c) Precision
(d) Standard error	(e) Estimator	(f) Margin of error
6. Give the general formula for a confidence interval.
7. State the probabilistic and practical interpretations of a confidence interval.
8. Of what use is the central limit theorem in estimation?
9. Describe the t distribution.
10. What are the assumptions underlying the use of the t distribution in estimating a single population mean?
11. What is the finite population correction? When can it be ignored?
12. What are the assumptions underlying the use of the t distribution in estimating the difference between two population means?

13. Arterial blood gas analyses performed on a sample of 15 physically active adult males yielded the following resting PaO₂ values:

75, 80, 80, 74, 84, 78, 89, 72, 83, 76, 75, 87, 78, 79, 88

Compute the 95 percent confidence interval for the mean of the population.

14. What proportion of asthma patients are allergic to house dust? In a sample of 140, 35 percent had positive skin reactions. Construct the 95 percent confidence interval for the population proportion.
15. An industrial hygiene survey was conducted in a large metropolitan area. Of 70 manufacturing plants of a certain type visited, 21 received a “poor” rating with respect to absence of safety hazards. Construct a 95 percent confidence interval for the population proportion deserving a “poor” rating.
16. Refer to the previous problem. How large a sample would be required to estimate the population proportion to within .05 with 95 percent confidence (.30 is the best available estimate of p):
- (a) If the finite population correction can be ignored?
- (b) If the finite population correction is not ignored and $N = 1500$?
17. In a dental survey conducted by a county dental health team, 500 adults were asked to give the reason for their last visit to a dentist. Of the 220 who had less than a high-school education, 44 said they went for preventative reasons. Of the remaining 280, who had a high-school education or better, 150 stated that they went for preventative reasons. Construct a 95 percent confidence interval for the difference between the two population proportions.
18. A breast cancer research team collected the following data on tumor size:

Type of Tumor	n	\bar{x}	s
A	21	3.85 cm	1.95 cm
B	16	2.80 cm	1.70 cm

Construct a 95 percent confidence interval for the difference between population means.

19. A certain drug was found to be effective in the treatment of pulmonary disease in 180 of 200 cases treated. Construct the 90 percent confidence interval for the population proportion.
20. Seventy patients with stasis ulcers of the leg were randomly divided into two equal groups. Each group received a different treatment for edema. At the end of the experiment, treatment effectiveness was measured in terms of reduction in leg volume as determined by water displacement. The means and standard deviations for the two groups were as follows:

Group (Treatment)	\bar{x}	s
A	95 cc	25
B	125 cc	30

Construct a 95 percent confidence interval for the difference in population means.

21. What is the average serum bilirubin level of patients admitted to a hospital for treatment of hepatitis? A sample of 10 patients yielded the following results:

20.5, 14.8, 21.3, 12.7, 15.2, 26.6, 23.4, 22.9, 15.7, 19.2

Construct a 95 percent confidence interval for the population mean.

22. Determinations of saliva pH levels were made in two independent random samples of seventh-grade schoolchildren. Sample A children were caries-free while sample B children had a high incidence of caries. The results were as follows:

A: 7.14, 7.11, 7.61, 7.98, 7.21, 7.16, 7.89 B: 7.36, 7.04, 7.19, 7.41, 7.10, 7.15, 7.36,
7.24, 7.86, 7.47, 7.82, 7.37, 7.66, 7.62, 7.65 7.57, 7.64, 7.00, 7.25, 7.19

Construct a 90 percent confidence interval for the difference between the population means. Assume that the population variances are equal.

23. Drug A was prescribed for a random sample of 12 patients complaining of insomnia. An independent random sample of 16 patients with the same complaint received drug B. The number of hours of sleep experienced during the second night after treatment began were as follows:

A: 3.5, 5.7, 3.4, 6.9, 17.8, 3.8, 3.0, 6.4, 6.8, 3.6, 6.9, 5.7
B: 4.5, 11.7, 10.8, 4.5, 6.3, 3.8, 6.2, 6.6, 7.1, 6.4, 4.5, 5.1,
3.2, 4.7, 4.5, 3.0

Construct a 95 percent confidence interval for the difference between the population means. Assume that the population variances are equal.

24. The objective of a study by Crane et al. (A-29) was to examine the efficacy, safety, and maternal satisfaction of (a) oral misoprostol and (b) intravenous oxytocin for labor induction in women with premature rupture of membranes at term. Researchers randomly assigned women to the two treatments. For the 52 women who received oral misoprostol, the mean time in minutes to active labor was 358 minutes with a standard deviation of 308 minutes. For the 53 women taking oxytocin, the mean time was 483 minutes with a standard deviation of 144 minutes. Construct a 99 percent confidence interval for the difference in mean time to active labor for these two different medications. What assumptions must be made about the reported data? Describe the population about which an inference can be made.
25. Over a 2-year period, 34 European women with previous gestational diabetes were retrospectively recruited from West London antenatal databases for a study conducted by Kousta et al. (A-30). One of the measurements for these women was the fasting nonesterified fatty acids concentration (NEFA) measured in $\mu\text{mol/L}$. In the sample of 34 women, the mean NEFA level was 435 with a sample standard deviation of 215.0. Construct a 95 percent confidence interval for the mean fasting NEFA level for a population of women with gestational diabetes. State all necessary assumptions about the reported data and subjects.
26. Scheid et al. (A-31) questioned 387 women receiving free bone mineral density screening. The questions focused on past smoking history. Subjects undergoing hormone replacement therapy (HRT), and subjects not undergoing HRT, were asked if they had ever been a regular smoker. In the HRT group, 29.3 percent of 220 women stated that they were at some point in their life a regular smoker. In the non-HRT group, 17.3 percent of 106 women responded positively to being at some point in their life a regular smoker. (Sixty-one women chose not to answer the question.) Construct a 95 percent confidence interval for the difference in smoking percentages for the two populations of women represented by the subjects in the study. What assumptions about the data are necessary?
27. The purpose of a study by Elliott et al. (A-32) was to assess the prevalence of vitamin D deficiency in women living in nursing homes. The sample consisted of 39 women in a 120-bed skilled nursing facility. Women older than 65 years of age who were long-term residents were invited to participate if they had no diagnosis of terminal cancer or metastatic disease. In the sample, 23 women had 25-hydroxyvitamin D levels of 20 ng/ml or less. Construct a 95 percent confidence interval for the percent of women with vitamin D deficiency in the population presumed to be represented by this sample.

28. In a study of the role of dietary fats in the etiology of ischemic heart disease the subjects were 60 males between 40 and 60 years of age who had recently had a myocardial infarction and 50 apparently healthy males from the same age group and social class. One variable of interest in the study was the proportion of linoleic acid (L.A.) in the subjects' plasma triglyceride fatty acids. The data on this variable were as follows:

Subjects with Myocardial Infarction

Subject	L.A.	Subject	L.A.	Subject	L.A.	Subject	L.A.
1	18.0	2	17.6	3	9.6	4	5.5
5	16.8	6	12.9	7	14.0	8	8.0
9	8.9	10	15.0	11	9.3	12	5.8
13	8.3	14	4.8	15	6.9	16	18.3
17	24.0	18	16.8	19	12.1	20	12.9
21	16.9	22	15.1	23	6.1	24	16.6
25	8.7	26	15.6	27	12.3	28	14.9
29	16.9	30	5.7	31	14.3	32	14.1
33	14.1	34	15.1	35	10.6	36	13.6
37	16.4	38	10.7	39	18.1	40	14.3
41	6.9	42	6.5	43	17.7	44	13.4
45	15.6	46	10.9	47	13.0	48	10.6
49	7.9	50	2.8	51	15.2	52	22.3
53	9.7	54	15.2	55	10.1	56	11.5
57	15.4	58	17.8	59	12.6	60	7.2

Healthy Subjects

Subject	L.A.	Subject	L.A.	Subject	L.A.	Subject	L.A.
1	17.1	2	22.9	3	10.4	4	30.9
5	32.7	6	9.1	7	20.1	8	19.2
9	18.9	10	20.3	11	35.6	12	17.2
13	5.8	14	15.2	15	22.2	16	21.2
17	19.3	18	25.6	19	42.4	20	5.9
21	29.6	22	18.2	23	21.7	24	29.7
25	12.4	26	15.4	27	21.7	28	19.3
29	16.4	30	23.1	31	19.0	32	12.9
33	18.5	34	27.6	35	25.0	36	20.0
37	51.7	38	20.5	39	25.9	40	24.6
41	22.4	42	27.1	43	11.1	44	32.7
45	13.2	46	22.1	47	13.5	48	5.3
49	29.0	50	20.2				

Construct the 95 percent confidence interval for the difference between population means. What do these data suggest about the levels of linoleic acid in the two sampled populations?

29. The purpose of a study by Tahmassebi and Curzon (A-33) was to compare the mean salivary flow rate among subjects with cerebral palsy and among subjects in a control group. Each group had 10 subjects. The following table gives the mean flow rate in ml/minute as well as the standard error.

Group	Sample Size	Mean ml/minute	Standard Error
Cerebral palsy	10	0.220	0.0582
Control	10	0.334	0.1641

source: J. F. Tahmassebi and M. E. J. Curzon, "The Cause of Drooling in Children with Cerebral Palsy—Hypersalivation or Swallowing Defect?" *International Journal of Paediatric Dentistry*, 13 (2003), 106–111.

Construct the 90 percent confidence interval for the difference in mean salivary flow rate for the two populations of subjects represented by the sample data. State the assumptions necessary for this to be a valid confidence interval.

30. Culligan et al. (A-34) compared the long-term results of two treatments: (a) a modified Burch procedure, and (b) a sling procedure for stress incontinence with a low-pressure urethra. Thirty-six women took part in the study with 19 in the Burch treatment group and 17 in the sling procedure treatment group. One of the outcome measures at three months post-surgery was maximum urethral closure pressure (cm H₂O). In the Burch group the mean and standard deviation were 16.4 and 8.2 cm, respectively. In the sling group, the mean and standard deviation were 39.8 and 23.0, respectively. Construct the 99 percent confidence interval for the difference in mean maximum urethral closure pressure for the two populations represented by these subjects. State all necessary assumptions.
31. In general, narrow confidence intervals are preferred over wide ones. We can make an interval narrow by using a small confidence coefficient. For a given set of other conditions, what happens to the level of confidence when we use a small confidence coefficient? What would happen to the interval width and the level of confidence if we were to use a confidence coefficient of zero?
32. In general, a high level of confidence is preferred over a low level of confidence. For a given set of other conditions, suppose we set our level of confidence at 100 percent. What would be the effect of such a choice on the width of the interval?
33. The subjects of a study by Borland et al. (A-35) were children in acute pain. Thirty-two children who presented at an emergency room were enrolled in the study. Each child used the visual analogue scale to rate pain on a scale from 0 to 100 mm. The mean pain score was 61.3 mm with a 95 percent confidence interval of 53.2 mm–69.4 mm. Which would be the appropriate reliability factor for the interval, z or t ? Justify your choice. What is the precision of the estimate? The margin of error?
34. Does delirium increase hospital stay? That was the research question investigated by McCusker et al. (A-36). The researchers sampled 204 patients with prevalent delirium and 118 without delirium. The conclusion of the study was that patients with prevalent delirium did not have a higher mean length of stay compared to those without delirium. What was the target population? The sampled population?
35. Assessing driving self-restriction in relation to vision performance was the objective of a study by West et al. (A-37). The researchers studied 629 current drivers ages 55 and older for 2 years. The variables of interest were driving behavior, health, physical function, and vision function. The subjects were part of a larger vision study at the Smith-Kettlewell Eye Research Institute. A conclusion of the study was that older adults with early changes in spatial vision function and depth perception appear to recognize their limitations and restrict their driving. What was the target population? The sampled population?
36. In a pilot study conducted by Ayoubia et al. (A-38), researchers studied 123 children born of HIV-1-infected mothers in Yaoundé, Cameroon. Counseled and consenting pregnant women were given a single dose of nevirapine at the onset of labor. Babies were given a syrup containing nevirapine within the first 72 hours of life. The researchers found that 87 percent of the children were considered not infected at 6–8 weeks of age. What is the target population? What is the sampled population?

37. Refer to Exercise 2.3.11. Construct a 95 percent confidence interval for the population mean S/R ratio. Should you use t or z as the reliability coefficient? Why? Describe the population about which inferences based on this study may be made.
38. Refer to Exercise 2.3.12. Construct a 90 percent confidence interval for the population mean height. Should you use t or z as the reliability coefficient? Why? Describe the population about which inferences based on this study may be made.

Exercises for Use with Large Data Sets Available on the Following Website:

www.wiley.com/college/daniel

1. Refer to North Carolina Birth Registry Data NCBIRTH800 with 800 observations (see Large Data Exercise 1 in Chapter 2). Calculate 95 percent confidence intervals for the following:
 - (a) the percentage of male children
 - (b) the mean age of a mother giving birth
 - (c) the mean weight gained during pregnancy
 - (d) the percentage of mothers admitting to smoking during pregnancy
 - (e) the difference in the average weight gained between smoking and nonsmoking mothers
 - (f) the difference in the average birth weight in grams between married and nonmarried mothers
 - (g) the difference in the percentage of low birth weight babies between married and nonmarried mothers
2. Refer to the serum cholesterol levels for 1000 subjects (CHOLEST). Select a simple random sample of size 15 from this population and construct a 95 percent confidence interval for the population mean. Compare your results with those of your classmates. What assumptions are necessary for your estimation procedure to be valid?
3. Refer to the serum cholesterol levels for 1000 subjects (CHOLEST). Select a simple random sample of size 50 from the population and construct a 95 percent confidence interval for the proportion of subjects in the population who have readings greater than 225. Compare your results with those of your classmates.
4. Refer to the weights of 1200 babies born in a community hospital (BABY WGTS). Draw a simple random sample of size 20 from this population and construct a 95 percent confidence interval for the population mean. Compare your results with those of your classmates. What assumptions are necessary for your estimation procedure to be valid?
5. Refer to the weights of 1200 babies born in a community hospital (BABY WGTS). Draw a simple random sample of size 35 from the population and construct a 95 percent confidence interval for the population mean. Compare this interval with the one constructed in Exercise 4.
6. Refer to the heights of 1000 twelve-year-old boys (BOY HGTS). Select a simple random sample of size 15 from this population and construct a 99 percent confidence interval for the population mean. What assumptions are necessary for this procedure to be valid?
7. Refer to the heights of 1000 twelve-year-old boys (BOY HGTS). Select a simple random sample of size 35 from the population and construct a 99 percent confidence interval for the population mean. Compare this interval with the one constructed in Exercise 5.

REFERENCES

Methodology References

1. JOHN A. RICE, *Mathematical Statistics and Data Analysis*, 2nd ed., Duxbury, Belmont, CA, 1988.
2. W. S. GOSSET ("Student"), "The Probable Error of a Mean," *Biometrika*, 6 (1908), 1–25.

3. W. V. BEHRENS, "Ein Beitrag zu Fehlerberechnung bei wenige Beobachtungen," *Landwirtschaftliche Jahrbücher*, 68 (1929), 807–837.
4. R. A. FISHER, "The Comparison of Samples with Possibly Unequal Variances," *Annals of Eugenics*, 9 (1939), 174–180.
5. R. A. FISHER, "The Asymptotic Approach to Behrens' Integral with Further Tables for the d Test of Significance," *Annals of Eugenics*, 11 (1941), 141–172.
6. J. NEYMAN, "Fiducial Argument and the Theory of Confidence Intervals," *Biometrika*, 32 (1941), 128–150.
7. H. SCHEFFÉ, "On Solutions of the Behrens-Fisher Problem Based on the t -Distribution," *Annals of Mathematical Statistics*, 14 (1943), 35–44.
8. H. SCHEFFÉ, "A Note on the Behrens-Fisher Problem," *Annals of Mathematical Statistics*, 15 (1944), 430–432.
9. B. L. WELCH, "The Significance of the Difference Between Two Means When the Population Variances Are Unequal," *Biometrika*, 29 (1937), 350–361.
10. B. L. WELCH, "The Generalization of 'Student's Problem When Several Different Population Variances Are Involved,'" *Biometrika*, 34 (1947), 28–35.
11. WILLIAM G. COCHRAN, "Approximate Significance Levels of the Behrens-Fisher Test," *Biometrics*, 20 (1964), 191–195.
12. R. F. TATE and G. W. KLETT, "Optimal Confidence Intervals for the Variance of a Normal Distribution," *Journal of the American Statistical Association*, 54 (1959), 674–682.
13. H. O. HARTLEY, "The Maximum F-Ratio as a Short Cut Test for Heterogeneity of Variances," *Biometrika*, 37 (1950), 308–312.
14. J. H. ZAR, *Biostatistical Analysis*, 4th ed., Prentice-Hall, Upper Saddle River, NJ, 1999.
15. WAYNE W. DANIEL, *Applied Nonparametric Statistics*, 2nd ed., PWS-KENT, Boston, 1989.

Applications References

- A-1. NICOLA MAFFULLI, CHERLY TALLON, JASON WONG, KIM PENG, and ROBERT BLEAKNEY, "Early Weightbearing and Ankle Mobilization after Open Repair of Acute Midsubstance Tears of the Achilles Tendon," *American Journal of Sports Medicine*, 31 (2003), 692–700.
- A-2. PAUL J. REBER, LUCY A. MARTINEZ, and SANDRA WEINTRAUB, "Artificial Grammar Learning in Alzheimer's Disease," *Cognitive, Affective, and Behavioral Neuroscience*, 3 (2003), 145–153.
- A-3. CHRISTOPHE PEDROLETTI, MARIEANN HÖGMAN, PEKKA MERILÄINEN, LENNART S. NORDVALL, GUNILLA HEDLIN, and KJELL ALVING, "Nitric Oxide Airway Diffusing Capacity and Mucosal Concentration in Asthmatic Schoolchildren," *Pediatric Research*, 54 (2003), 496–501.
- A-4. BRUCE D. BEYNNON, BRADEN C. FLEMING, DAVID L. CHURCHILL, and DANIEL BROWN, "The Effect of Anterior Cruciate Ligament Deficiency and Functional Bracing on Translation of the Tibia Relative to the Femur during Nonweightbearing and Weightbearing," *American Journal of Sports Medicine*, 31 (2003), 99–105.
- A-5. LORRAINE DUGOFF, MAURITHA R. EVERETT, LOUIS VONTVER, and GWYN E. BARLEY, "Evaluation of Pelvic and Breast Examination Skills of Interns in Obstetrics and Gynecology and Internal Medicine," *American Journal of Obstetrics and Gynecology*, 189 (2003), 655–658.
- A-6. DIONNE MAYHEW, KAREN M. PERRIN, and WENDY STRUCHEN, "An Analysis of a Healthy Start Smoking Cessation Program," *American Journal of Health Studies*, 17 (2002), 186–190.
- A-7. ERIC GRANHOLM, ROBERT ANTHENELLI, RITA MONTEIRO, JOHN SEVCIK, and MARILYN STOLER, "Brief Integrated Outpatient Dual-Diagnosis Treatment Reduces Psychiatric Hospitalizations," *American Journal on Addictions*, 12 (2003), 306–313.
- A-8. SILVIA IANNELO, ANTONINA CAVALERI, PAOLINA MILAZZO, SANTI CANTARELLA, and FRANCESCO BELFIORE, "Low Fasting Serum Triglyceride Level as a Precocious Marker of Autoimmune Disorders," *Medscape General Medicine*, 5 (3) (2003).
- A-9. EVELYN C. Y. CHAN, SALLY W. VERNON, FREDERICK T. O'DONNELL, CHUL AHN, ANTHONY GRESINGER, and DONNIE W. AGA, "Informed Consent for Cancer Screening with Prostate-Specific Antigen: How Well Are Men Getting the Message?" *American Journal of Public Health*, 93 (2003), 779–785.
- A-10. RONALD F. VAN VOLLENHOVEN, SOFIA ERNESTAM, ANDERS HARJU, JOHAN BRATT, and LARS KLARESKOG, "Etanercept Versus Etanercept Plus Methotrexate: A Registry-Based Study Suggesting that the Combination Is Clinically More Efficacious," *Arthritis Research and Therapy*, 5 (2003), R347–R351.

- A-11. SATOSHI NOZAWA, KATSUJI SHIMIZU, KEI MIYAMOTO, and MIZUO TANAKA, "Repair of Pars Interarticularis Defect by Segmental Wire Fixation in Young Athletes with Spondylolysis," *American Journal of Sports Medicine*, 31 (2003), 359–364.
- A-12. MORI J. KRANTZ, ILANA B. KUTINSKY, ALASTAIR D. ROBERTSON, and PHILIP S. MEHLER, "Dose-Related Effects of Methadone on QT Prolongation in a Series of Patients with Torsade de Pointes," *Pharmacotherapy*, 23 (2003), 802–805.
- A-13. SUSANNAH FOX and DEBORAH FELLOWS, "Health Searches and Email Have Become More Commonplace, But There Is Room for Improvement in Searches and Overall Internet Access," *Pew Internet and American Life Project*, www.pewinternet.org/PPF/r/95/report_display.asp.
- A-14. CARLOS M. LUNA, DANIEL BLANZACO, MICHAEL S. NIDERMAN, WALTER MATARUCCO, NATALIO C. BAREDES, PABLO DESMERY, FERNANDO PALIZAS, GUILLERMO MENGA, FERNANDO RIOS, and CARLOS APEZTEGUIA, "Resolution of Ventilator-Associated Pneumonia: Prospective Evaluation of the Clinical Pulmonary Infection Score as an Early Clinical Predictor of Outcome," *Critical Care Medicine*, 31 (3) (2003), 676–682.
- A-15. AREND F. L. SCHINKEL, JEROEN J. BAX, ERIC BOERSMA, ABDOL ELHENDY, ELENI C. VOURVOURI, JOS R. T. C. ROELANDT, and DON POLDERMANS, "Assessment of Residual Myocardial Viability in Regions with Chronic Electrocardiographic Q-Wave Infarction," *American Heart Journal*, 144 (2002), 865–869.
- A-16. FRIEDERIKE VON ZUR MUHLEN, WEILUM QUAN, DAVID J. D'AGATE, and TODD J. COHEN, "A Study of Carotid Sinus Massage and Head-Up Tilt Table Testing in Patients with Syncope and Near-Syncope," *Journal of Invasive Cardiology*, 14 (2002), 477–482.
- A-17. DANIEL F. CONNOR, RONALD J. STEINGARD, JENNIFER J. ANDERSON, and RICHARD H. MELLONI, Jr., "Gender Differences in Reactive and Proactive Aggression," *Child Psychiatry and Human Development*, 33 (2003), 279–294.
- A-18. ALLAN V. HORWITZ, CATHY SPATZ WIDOM, JULIE McLAUGHLIN, and HELENE RASKIN WHITE, "The Impact of Childhood Abuse and Neglect on Adult Mental Health: A Prospective Study," *Journal of Health and Social Behavior*, 42 (2001), 184–201.
- A-19. P. ADAB, T. MARSHALL, A. ROUSE, B. RANDHAWA, H. SANGHA, and N. BHANGOO, "Randomised Controlled Trial of the Effect of Evidence Based Information on Women's Willingness to Participate in Cervical Cancer Screening," *Journal of Epidemiology and Community Health*, 57 (2003), 589–593.
- A-20. JOHN A. SPERTUS, TIM KEWHURST, CYNTHIA M. DOUGHERTY, and PAUL NICHOL, "Testing the Effectiveness of Converting Patients to Long-Acting Antianginal Medications: The Quality of Life in Angina Research," *American Heart Journal*, 141 (2001), 550–558.
- A-21. PATRICK R. FINLEY, HEIDI R. RENS, JOAN T. PONT, SUSAN L. GESS, CLIFTON LOUIE, SCOTT A. BULL, JANELLE Y. LEE, and LISA A. BERO, "Impact of a Collaborative Care Model on Depression in a Primary Care Setting: A Randomized, Controlled Trial," *Pharmacotherapy*, 23 (9) (2003), 1175–1185.
- A-22. MICHAEL HUMMEL, EZIO BONIFACIO, HEIKE E. NASERKE, and ANETTE G. ZIEGLER, "Elimination of Dietary Gluten Does Not Reduce Titers of Type 1 Diabetes-Associated Autoantibodies in High-Risk Subjects," *Diabetes Care*, 25 (2002), 1111–1116.
- A-23. DOV AIZENBERG, ABRAHAM WEIZMAN, and YORAM BARAK, "Sildenafil for Selective Serotonin Reuptake Inhibitor-Induced Erectile Dysfunction in Elderly Male Depressed Patients," *Journal of Sex and Marital Therapy*, 29 (2003), 297–303.
- A-24. PETER BORDEN, JOHN NYLAND, DAVID N. M. CABORN, and DAVID PIENKOWSKI, "Biomechanical Comparison of the FasT-Fix Meniscal Repair Suture System with Vertical Mattress Sutures and Meniscus Arrows," *American Journal of Sports Medicine*, 31 (2003), 374–378.
- A-25. RACHEL H. ALLEN and MICHAEL T. GROSS, "Toe Flexors Strength and Passive Extension Range of Motion of the First Metatarsophalangeal Joint in Individuals with Plantar Fasciitis," *Journal of Orthopaedic and Sports Physical Therapy*, 33 (2003), 468–477.
- A-26. MOHEB S. MONEIM, KEIKHOSROW FIROOZBAKHSH, DOMINIC GROSS, STEVEN D. YOUNG, and GEORGE OMER, "Thumb Amputations from Team Roping," *American Journal of Sports Medicine*, 31 (2003), 728–735.
- A-27. NETTA HORESH, ISRAEL ORBACH, DORON GOTHELF, MEIR EFRATI, and ALAN APTER, "Comparison of the Suicidal Behavior of Adolescent Inpatients with Borderline Personality Disorder and Major Depression," *Journal of Nervous and Mental Disease*, 191 (2003), 582–588.
- A-28. R. PLUTCHIK, and H. VAN PRAAG, "The Measurement of Suicidality, Aggressivity and Impulsivity," *Clinical Neuropharmacology*, 9(suppl) (1989), 380–382.
- A-29. JOAN M. G. CRANE, TINA DELANEY, and DONNA HUTCHENS, "Oral Misoprostol for Premature Rupture of Membranes at Term," *American Journal of Obstetrics and Gynecology*, 189 (2003), 720–724.

- A-30. ELENI KOUSTA, NATASHA J. LAWRENCE, IAN F. GODSLAND, ANNA PENNY, VICTOR ANYAOKU, BARBARA A. MILLAUER, ESTER CELA, DESMOND G. JOHNSTON, STEPHEN ROBINSON, and MARK I. MCCARTHY, "Insulin Resistance and Beta-Cell Dysfunction in Normoglycaemic European Women with a History of Gestational Diabetes," *Clinical Endocrinology*, 59 (2003), 289–297.
- A-31. DEWEY C. SCHEID, MARIO T. COLEMAN, and ROBERT M. HAMM, "Do Perceptions of Risk and Quality of Life Affect Use of Hormone Replacement Therapy by Postmenopausal Women?" *Journal of the American Board of Family Practice*, 16 (2003), 270–277.
- A-32. MARY E. ELLIOTT, NEIL C. BINKLEY, MOLLY CARNES, DAVID R. ZIMMERMAN, KIM PETERSEN, KATHY KNAPP, JESSICA J. BEHLKE, NANCY AHMANN, and MARY A. KIESER, "Fracture Risks for Women in Long-Term Care: High Prevalence of Calcaneal Osteoporosis and Hypovitaminosis D," *Pharmacotherapy*, 23 (2003), 702–710.
- A-33. J. F. TAHMASSEBI and M. E. J. CURZON, "The Cause of Drooling in Children with Cerebral Palsy—Hypersalivation or Swallowing Defect?" *International Journal of Paediatric Dentistry*, 13 (2003), 106–111.
- A-34. PATRICK J. CULLIGAN, ROGER P. GOLDBERG, and PETER K. SAND, "A Randomized Controlled Trial Comparing a Modified Burch Procedure and a Suburethral Sling: Long-Term Follow-Up," *International Urogynecology Journal*, 14 (2003), 229–233.
- A-35. MEREDITH L. BORLAND, IAN JACOBS, and GARY GEELHOE, "Intranasal Fentanyl Reduces Acute Pain in Children in the Emergency Department: A Safety and Efficacy Study," *Emergency Medicine*, 14 (2002), 275–280.
- A-36. JANE McCUSKER, MARTIN G. COLE, NANDINI DENDUKURI, and ERIC BELZILE, "Does Delirium Increase Hospital Stay?" *Journal of the American Geriatrics Society*, 51 (2003), 1539–1546.
- A-37. CATHERINE G. WEST, GINNY GILDENGORIN, GUNILLA HAEGERSTROM PORTNOY, LORI A. LOTT, MARILYN E. SCHNECK, and JOHN A. BRABYN, "Vision and Driving Self-Restriction in Older Adults," *Journal of the American Geriatrics Society*, 51 (2003), 1348–1355.
- A-38. AHIDJO AYOUBA, GILBERT TENE, PATRICK CUNIN, YACOUBA FOUPOUAPOUOGNIGNI, ELISABETH MENU, ANFUMBOM KFUTWAH, JOCELYN THONNON, GABRIELLA SCARLATTI, MARCEL MONNY-LOBÉ, NICOLE ETEKI, CHARLES KOUANFACK, MICHÈLE TARDY, ROBERT LEKE, MAURICE NKAM, ANNE E. NLEND, FRANÇOISE BARRÉ-SINOSSI, PAUL M. V. MARTIN, and ERIC NERRIENET, "Low Rate of Mother-to-Child Transmission of HIV-1 After Nevirapine Intervention in a Pilot Public Health Program in Yaoundé, Cameroon," *Journal of Acquired Immune Deficiency Syndrome*, 34 (2003), 274–280.

HYPOTHESIS TESTING

CHAPTER OVERVIEW

This chapter covers hypothesis testing, the second of two general areas of statistical inference. Hypothesis testing is a topic with which you as a student are likely to have some familiarity. Interval estimation, discussed in the preceding chapter, and hypothesis testing are based on similar concepts. In fact, confidence intervals may be used to arrive at the same conclusions that are reached through the use of hypothesis tests. This chapter provides a format, followed throughout the remainder of this book, for conducting a hypothesis test.

TOPICS

- 7.1 INTRODUCTION
- 7.2 HYPOTHESIS TESTING: A SINGLE POPULATION MEAN
- 7.3 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION MEANS
- 7.4 PAIRED COMPARISONS
- 7.5 HYPOTHESIS TESTING: A SINGLE POPULATION PROPORTION
- 7.6 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS
- 7.7 HYPOTHESIS TESTING: A SINGLE POPULATION VARIANCE
- 7.8 HYPOTHESIS TESTING: THE RATIO OF TWO POPULATION VARIANCES
- 7.9 THE TYPE II ERROR AND THE POWER OF A TEST
- 7.10 DETERMINING SAMPLE SIZE TO CONTROL TYPE II ERRORS
- 7.11 SUMMARY

LEARNING OUTCOMES

After studying this chapter, the student will

1. understand how to correctly state a null and alternative hypothesis and carry out a structured hypothesis test.
2. understand the concepts of type I error, type II error, and the power of a test.
3. be able to calculate and interpret z , t , F , and chi-square test statistics for making statistical inferences.
4. understand how to calculate and interpret p values.

7.1 INTRODUCTION

One type of statistical inference, estimation, is discussed in the preceding chapter. The other type, hypothesis testing, is the subject of this chapter. As is true with estimation, the *purpose of hypothesis testing is to aid the clinician, researcher, or administrator in reaching a conclusion concerning a population by examining a sample from that population*. Estimation and hypothesis testing are not as different as they are made to appear by the fact that most textbooks devote a separate chapter to each. As we will explain later, one may use confidence intervals to arrive at the same conclusions that are reached by using the hypothesis testing procedures discussed in this chapter.

Basic Concepts In this section some of the basic concepts essential to an understanding of hypothesis testing are presented. The specific details of particular tests will be given in succeeding sections.

DEFINITION

A hypothesis may be defined simply as a statement about one or more populations.

The hypothesis is frequently concerned with the parameters of the populations about which the statement is made. A hospital administrator may hypothesize that the average length of stay of patients admitted to the hospital is 5 days; a public health nurse may hypothesize that a particular educational program will result in improved communication between nurse and patient; a physician may hypothesize that a certain drug will be effective in 90 percent of the cases for which it is used. By means of hypothesis testing one determines whether or not such statements are compatible with the available data.

Types of Hypotheses Researchers are concerned with two types of hypotheses—research hypotheses and statistical hypotheses.

DEFINITION

The research hypothesis is the conjecture or supposition that motivates the research.

It may be the result of years of observation on the part of the researcher. A public health nurse, for example, may have noted that certain clients responded more readily to a particular type of health education program. A physician may recall numerous instances in which certain combinations of therapeutic measures were more effective than any one of them alone. Research projects often result from the desire of such health practitioners to determine whether or not their theories or suspicions can be supported when subjected to the rigors of scientific investigation.

Research hypotheses lead directly to statistical hypotheses.

DEFINITION

Statistical hypotheses are hypotheses that are stated in such a way that they may be evaluated by appropriate statistical techniques.

In this book the hypotheses that we will focus on are statistical hypotheses. We will assume that the research hypotheses for the examples and exercises have already been considered.

Hypothesis Testing Steps For convenience, hypothesis testing will be presented as a ten-step procedure. There is nothing magical or sacred about this particular format. It merely breaks the process down into a logical sequence of actions and decisions.

1. **Data.** The nature of the data that form the basis of the testing procedures must be understood, since this determines the particular test to be employed. Whether the data consist of counts or measurements, for example, must be determined.
2. **Assumptions.** As we learned in the chapter on estimation, different assumptions lead to modifications of confidence intervals. The same is true in hypothesis testing: A general procedure is modified depending on the assumptions. In fact, the same assumptions that are of importance in estimation are important in hypothesis testing. We have seen that these include assumptions about the normality of the population distribution, equality of variances, and independence of samples.
3. **Hypotheses.** There are two statistical hypotheses involved in hypothesis testing, and these should be stated explicitly. The *null hypothesis* is the *hypothesis to be tested*. It is designated by the symbol H_0 . The null hypothesis is sometimes referred to as a *hypothesis of no difference*, since it is a statement of agreement with (or no difference from) conditions presumed to be true in the population of interest. In general, the null hypothesis is set up for the express purpose of being discredited. Consequently, the complement of the conclusion that the researcher is seeking to reach becomes the statement of the null hypothesis. In the testing process the null hypothesis either is rejected or is not rejected. If the null hypothesis is not rejected, we will say that the data on which the test is based do not provide sufficient evidence to cause rejection. If the testing procedure leads to rejection, we will say that the data at hand are not compatible with the null hypothesis, but are supportive of some other hypothesis. The *alternative hypothesis* is a statement of what we will believe is true if our sample data cause us to reject the null hypothesis. Usually the alternative hypothesis and the research hypothesis are the same, and in fact the two terms are used interchangeably. We shall designate the alternative hypothesis by the symbol H_A .

Rules for Stating Statistical Hypotheses When hypotheses are of the type considered in this chapter an indication of equality (either $=$, \leq , or \geq) must appear in the null hypothesis. Suppose, for example, that we want to answer the

question: Can we conclude that a certain population mean is not 50? The null hypothesis is

$$H_0: \mu = 50$$

and the alternative is

$$H_A: \mu \neq 50$$

Suppose we want to know if we can conclude that the population mean is greater than 50. Our hypotheses are

$$H_0: \mu \leq 50 \quad H_A: \mu > 50$$

If we want to know if we can conclude that the population mean is less than 50, the hypotheses are

$$H_0: \mu \geq 50 \quad H_A: \mu < 50$$

In summary, we may state the following rules of thumb for deciding what statement goes in the null hypothesis and what statement goes in the alternative hypothesis:

- (a) What you hope or expect to be able to conclude as a result of the test usually should be placed in the alternative hypothesis.
- (b) The null hypothesis should contain a statement of equality, either $=$, \leq , or \geq .
- (c) The null hypothesis is the hypothesis that is tested.
- (d) The null and alternative hypotheses are complementary. That is, the two together exhaust all possibilities regarding the value that the hypothesized parameter can assume.

A Precaution It should be pointed out that neither hypothesis testing nor statistical inference, in general, leads to the proof of a hypothesis; it merely indicates whether the hypothesis is supported or is not supported by the available data. When we fail to reject a null hypothesis, therefore, we do not say that it is true, but that it may be true. When we speak of accepting a null hypothesis, we have this limitation in mind and do not wish to convey the idea that accepting implies proof.

4. Test statistic. The test statistic is some statistic that may be computed from the data of the sample. As a rule, there are many possible values that the test statistic may assume, the particular value observed depending on the particular sample drawn. As we will see, the test statistic serves as a decision maker, since the decision to reject or not to reject the null hypothesis depends on the magnitude of the test statistic. An example of a test statistic is the quantity

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad (7.1.1)$$

where μ_0 is a hypothesized value of a population mean. This test statistic is related to the statistic

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (7.1.2)$$

with which we are already familiar.

General Formula for Test Statistic The following is a general formula for a test statistic that will be applicable in many of the hypothesis tests discussed in this book:

$$\text{test statistic} = \frac{\text{relevant statistic} - \text{hypothesized parameter}}{\text{standard error of the relevant statistic}}$$

In Equation 7.1.1, \bar{x} is the relevant statistic, μ_0 is the hypothesized parameter, and σ/\sqrt{n} is the standard error of \bar{x} , the relevant statistic.

5. Distribution of test statistic. It has been pointed out that the key to statistical inference is the sampling distribution. We are reminded of this again when it becomes necessary to specify the probability distribution of the test statistic. The distribution of the test statistic

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

for example, follows the standard normal distribution if the null hypothesis is true and the assumptions are met.

6. Decision rule. All possible values that the test statistic can assume are points on the horizontal axis of the graph of the distribution of the test statistic and are divided into two groups; one group constitutes what is known as the *rejection region* and the other group makes up the *nonrejection region*. The values of the test statistic forming the rejection region are those values that are less likely to occur if the null hypothesis is true, while the values making up the acceptance region are more likely to occur if the null hypothesis is true. *The decision rule tells us to reject the null hypothesis if the value of the test statistic that we compute from our sample is one of the values in the rejection region and to not reject the null hypothesis if the computed value of the test statistic is one of the values in the nonrejection region.*

Significance Level The decision as to which values go into the rejection region and which ones go into the nonrejection region is made on the basis of the desired *level of significance*, designated by α . The term *level of significance* reflects the fact that hypothesis tests are sometimes called significance tests, and a computed value of the test statistic that falls in the rejection region is said to be *significant*. The level of significance, α , specifies the area under the curve of the distribution of the test statistic that is above the values on the horizontal axis constituting the rejection region.

DEFINITION

The level of significance α is a probability and, in fact, is the probability of rejecting a true null hypothesis.

Since to reject a true null hypothesis would constitute an error, it seems only reasonable that we should make the probability of rejecting a true null hypothesis small and, in fact, that is what is done. We select a small value of α in order to make the probability of rejecting a true null hypothesis small. The more frequently encountered values of α are .01, .05, and .10.

Types of Errors The error committed when a true null hypothesis is rejected is called the *type I error*. The *type II error* is the error committed when a false null hypothesis is not rejected. The probability of committing a type II error is designated by β .

Whenever we reject a null hypothesis there is always the concomitant risk of committing a type I error, rejecting a true null hypothesis. Whenever we fail to reject a null hypothesis the risk of failing to reject a false null hypothesis is always present. We make α small, but we generally exercise no control over β , although we know that in most practical situations it is larger than α .

We never know whether we have committed one of these errors when we reject or fail to reject a null hypothesis, since the true state of affairs is unknown. If the testing procedure leads to rejection of the null hypothesis, we can take comfort from the fact that we made α small and, therefore, the probability of committing a type I error was small. If we fail to reject the null hypothesis, we do not know the concurrent risk of committing a type II error, since β is usually unknown but, as has been pointed out, we do know that, in most practical situations, it is larger than α .

Figure 7.1.1 shows for various conditions of a hypothesis test the possible actions that an investigator may take and the conditions under which each of the two types of error will be made. The table shown in this figure is an example of what is generally referred to as a *confusion matrix*.

- 7. Calculation of test statistic.** From the data contained in the sample we compute a value of the test statistic and compare it with the rejection and nonrejection regions that have already been specified.
- 8. Statistical decision.** The statistical decision consists of rejecting or of not rejecting the null hypothesis. It is rejected if the computed value of the test statistic falls in the

		Condition of Null Hypothesis	
		True	False
Possible Action	Fail to reject H_0	Correct action	Type II error
	Reject H_0	Type I error	Correct action

FIGURE 7.1.1 Conditions under which type I and type II errors may be committed.

rejection region, and it is not rejected if the computed value of the test statistic falls in the nonrejection region.

9. **Conclusion.** If H_0 is rejected, we conclude that H_A is true. If H_0 is not rejected, we conclude that H_0 may be true.
10. **p values.** The p value is a number that tells us how unusual our sample results are, given that the null hypothesis is true. A p value indicating that the sample results are not likely to have occurred, if the null hypothesis is true, provides justification for doubting the truth of the null hypothesis.

DEFINITION

A p value is the probability that the computed value of a test statistic is at least as extreme as a specified value of the test statistic when the null hypothesis is true. Thus, the p value is the smallest value of α for which we can reject a null hypothesis.

We emphasize that when the null hypothesis is not rejected one should not say that the null hypothesis is accepted. We should say that the null hypothesis is “not rejected.” We avoid using the word “accept” in this case because we may have committed a type II error. Since, frequently, the probability of committing a type II error can be quite high, we do not wish to commit ourselves to accepting the null hypothesis.

Figure 7.1.2 is a flowchart of the steps that we follow when we perform a hypothesis test.

Purpose of Hypothesis Testing The purpose of hypothesis testing is to assist administrators and clinicians in making decisions. The administrative or clinical decision usually depends on the statistical decision. If the null hypothesis is rejected, the administrative or clinical decision usually reflects this, in that the decision is compatible with the alternative hypothesis. The reverse is usually true if the null hypothesis is not rejected. The administrative or clinical decision, however, may take other forms, such as a decision to gather more data.

We also emphasize that the hypothesis testing procedures highlighted in the remainder of this chapter generally examine the case of normally distributed data or cases where the procedures are appropriate because the central limit theorem applies. In practice, it is not uncommon for samples to be small relative to the size of the population, or to have samples that are highly skewed, and hence the assumption of normality is violated. Methods to handle this situation, that is *distribution-free* or *nonparametric methods*, are examined in detail in Chapter 13. Most computer packages include an analytical procedure (for example, the Shapiro-Wilk or Anderson-Darling test) for testing normality. It is important that such tests are carried out prior to analysis of data. Further, when testing two samples, there is an implicit assumption that the variances are equal. Tests for this assumption are provided in Section 7.8. Finally, it should be noted that hypothesis tests, just like confidence intervals, are relatively

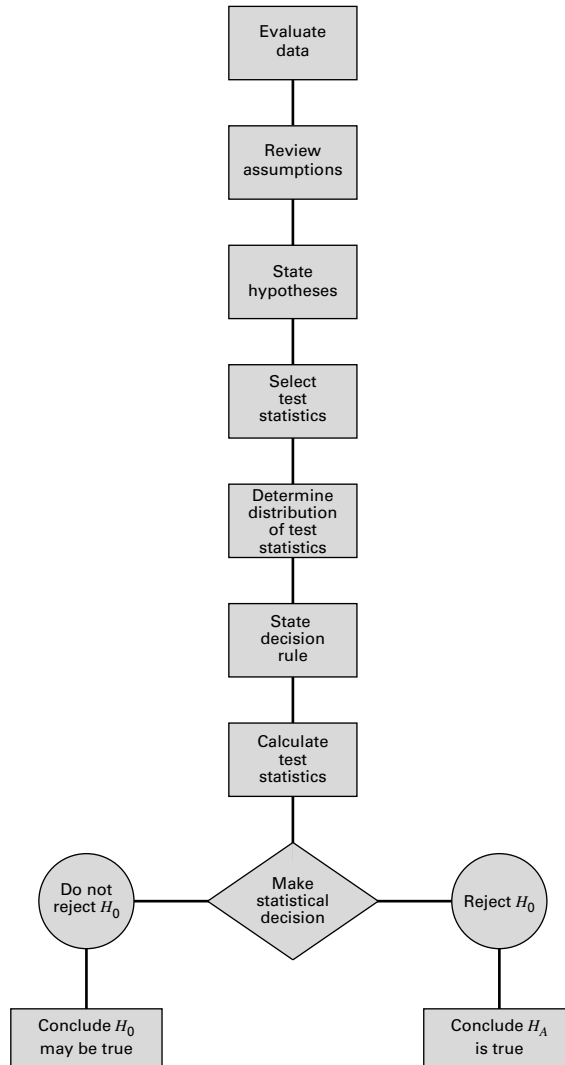


FIGURE 7.1.2 Steps in the hypothesis testing procedure.

sensitive to the size of the samples being tested, and caution should be taken when interpreting results involving very small sample sizes.

We must emphasize at this point, however, that the outcome of the statistical test is only one piece of evidence that influences the administrative or clinical decision. The statistical decision should not be interpreted as definitive but should be considered along with all the other relevant information available to the experimenter.

With these general comments as background, we now discuss specific hypothesis tests.

7.2 HYPOTHESIS TESTING: A SINGLE POPULATION MEAN

In this section we consider the testing of a hypothesis about a population mean under three different conditions: (1) when sampling is from a normally distributed population of values with known variance; (2) when sampling is from a normally distributed population with unknown variance, and (3) when sampling is from a population that is not normally distributed. Although the theory for conditions 1 and 2 depends on normally distributed populations, it is common practice to make use of the theory when relevant populations are only approximately normally distributed. This is satisfactory as long as the departure from normality is not drastic. When sampling is from a normally distributed population and the population variance is known, the test statistic for testing $H_0: \mu = \mu_0$ is

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \quad (7.2.1)$$

which, when H_0 is true, is distributed as the standard normal. Examples 7.2.1 and 7.2.2 illustrate hypothesis testing under these conditions.

Sampling from Normally Distributed Populations: Population Variances Known

As we did in Chapter 6, we again emphasize that situations in which the variable of interest is normally distributed with a known variance are rare. The following example, however, will serve to illustrate the procedure.

EXAMPLE 7.2.1

Researchers are interested in the mean age of a certain population. Let us say that they are asking the following question: Can we conclude that the mean age of this population is different from 30 years?

Solution: Based on our knowledge of hypothesis testing, we reply that they can conclude that the mean age is different from 30 if they can reject the null hypothesis that the mean is equal to 30. Let us use the ten-step hypothesis testing procedure given in the previous section to help the researchers reach a conclusion.

1. **Data.** The data available to the researchers are the ages of a simple random sample of 10 individuals drawn from the population of interest. From this sample a mean of $\bar{x} = 27$ has been computed.
2. **Assumptions.** It is assumed that the sample comes from a population whose ages are approximately normally distributed. Let us also assume that the population has a known variance of $\sigma^2 = 20$.
3. **Hypotheses.** The hypothesis to be tested, or null hypothesis, is that the mean age of the population is equal to 30. The alternative hypothesis is

that the mean age of the population is not equal to 30. Note that we are identifying with the alternative hypothesis the conclusion the researchers wish to reach, so that if the data permit rejection of the null hypothesis, the researchers' conclusion will carry more weight, since the accompanying probability of rejecting a true null hypothesis will be small. We will make sure of this by assigning a small value to α , the probability of committing a type I error. We may present the relevant hypotheses in compact form as follows:

$$H_0: \mu = 30$$

$$H_A: \mu \neq 30$$

4. **Test statistic.** Since we are testing a hypothesis about a population mean, since we assume that the population is normally distributed, and since the population variance is known, our test statistic is given by Equation 7.2.1.
5. **Distribution of test statistic.** Based on our knowledge of sampling distributions and the normal distribution, we know that the test statistic is normally distributed with a mean of 0 and a variance of 1, if H_0 is true. There are many possible values of the test statistic that the present situation can generate; one for every possible sample of size 10 that can be drawn from the population. Since we draw only one sample, we have only one of these possible values on which to base a decision.
6. **Decision rule.** The decision rule tells us to reject H_0 if the computed value of the test statistic falls in the rejection region and to fail to reject H_0 if it falls in the nonrejection region. We must now specify the rejection and nonrejection regions. We can begin by asking ourselves what magnitude of values of the test statistic will cause rejection of H_0 . If the null hypothesis is false, it may be so either because the population mean is less than 30 or because the population mean is greater than 30. Therefore, either sufficiently small values or sufficiently large values of the test statistic will cause rejection of the null hypothesis. We want these extreme values to constitute the rejection region. How extreme must a possible value of the test statistic be to qualify for the rejection region? The answer depends on the significance level we choose, that is, the size of the probability of committing a type I error. Let us say that we want the probability of rejecting a true null hypothesis to be $\alpha = .05$. Since our rejection region is to consist of two parts, sufficiently small values and sufficiently large values of the test statistic, part of α will have to be associated with the large values and part with the small values. It seems reasonable that we should divide α equally and let $\alpha/2 = .025$ be associated with small values and $\alpha/2 = .025$ be associated with large values.

Critical Value of Test Statistic

What value of the test statistic is so large that, when the null hypothesis is true, the probability of obtaining a value this large or larger is .025? In other words, what is the value of z to the right of which lies .025 of the area under the standard normal distribution? The value of z to the right of which lies .025 of the area is the same value that has .975 of the area between it and $-\infty$. We look in the body of Appendix Table D until we find .975 or its closest value and read the corresponding marginal entries to obtain our z value. In the present example the value of z is 1.96. Similar reasoning will lead us to find -1.96 as the value of the test statistic so small that when the null hypothesis is true, the probability of obtaining a value this small or smaller is .025. Our rejection region, then, consists of all values of the test statistic equal to or greater than 1.96 and less than or equal to -1.96 . The nonrejection region consists of all values in between. We may state the decision rule for this test as follows: *reject H_0 if the computed value of the test statistic is either ≥ 1.96 or ≤ -1.96 . Otherwise, do not reject H_0 .* The rejection and nonrejection regions are shown in Figure 7.2.1. The values of the test statistic that separate the rejection and nonrejection regions are called *critical values* of the test statistic, and the rejection region is sometimes referred to as the *critical region*.

The decision rule tells us to compute a value of the test statistic from the data of our sample and to reject H_0 if we get a value that is either equal to or greater than 1.96 or equal to or less than -1.96 and to fail to reject H_0 if we get any other value. The value of α and, hence, the decision rule should be decided on before gathering the data. This prevents our being accused of allowing the sample results to influence our choice of α . This condition of objectivity is highly desirable and should be preserved in all tests.

7. Calculation of test statistic. From our sample we compute

$$z = \frac{27 - 30}{\sqrt{20/10}} = \frac{-3}{1.4142} = -2.12$$

8. Statistical decision. Abiding by the decision rule, we are able to reject the null hypothesis since -2.12 is in the rejection region. We

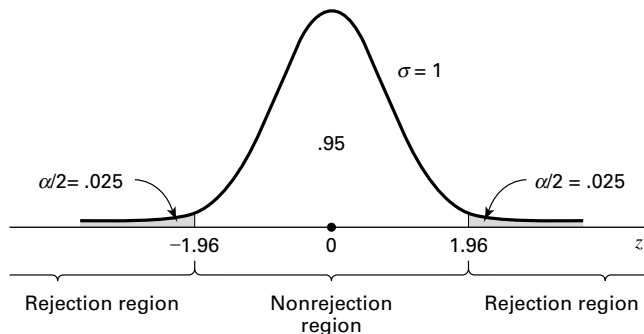


FIGURE 7.2.1 Rejection and nonrejection regions for Example 7.2.1.

can say that the computed value of the test statistic is significant at the .05 level.

- 9. Conclusion.** We conclude that μ is not equal to 30 and let our administrative or clinical actions be in accordance with this conclusion.
- 10. p values.** Instead of saying that an observed value of the test statistic is significant or is not significant, most writers in the research literature prefer to report the exact probability of getting a value as extreme as or more extreme than that observed if the null hypothesis is true. In the present instance these writers would give the computed value of the test statistic along with the statement $p = .0340$. The statement $p = .0340$ means that the probability of getting a value as extreme as 2.12 in either direction, when the null hypothesis is true, is .0340. The value .0340 is obtained from Appendix Table D and is the probability of observing a $z \geq 2.12$ or a $z \leq -2.12$ when the null hypothesis is true. That is, when H_0 is true, the probability of obtaining a value of z as large as or larger than 2.12 is .0170, and the probability of observing a value of z as small as or smaller than -2.12 is .0170. The probability of one or the other of these events occurring, when H_0 is true, is equal to the sum of the two individual probabilities, and hence, in the present example, we say that $p = .0170 + .0170 = .0340$.

Recall that the p value for a test may be defined also as the smallest value of α for which the null hypothesis can be rejected. Since, in Example 7.2.1, our p value is .0340, we know that we could have chosen an α value as small as .0340 and still have rejected the null hypothesis. If we had chosen an α smaller than .0340, we would not have been able to reject the null hypothesis. A general rule worth remembering, then, is this: *if the p value is less than or equal to α , we reject the null hypothesis; if the p value is greater than α , we do not reject the null hypothesis.*

The reporting of p values as part of the results of an investigation is more informative to the reader than such statements as “the null hypothesis is rejected at the .05 level of significance” or “the results were not significant at the .05 level.” Reporting the p value associated with a test lets the reader know just how common or how rare is the computed value of the test statistic given that H_0 is true. ■

Testing H_0 by Means of a Confidence Interval Earlier, we stated that one can use confidence intervals to test hypotheses. In Example 7.2.1 we used a hypothesis testing procedure to test $H_0: \mu = 30$ against the alternative, $H_A: \mu \neq 30$. We were able to reject H_0 because the computed value of the test statistic fell in the rejection region.

Let us see how we might have arrived at this same conclusion by using a $100(1 - \alpha)$ percent confidence interval. The 95 percent confidence interval for μ is

$$\begin{aligned} &27 \pm 1.96\sqrt{20/10} \\ &27 \pm 1.96(1.414) \\ &27 \pm 2.7714 \\ &(24.2286, 29.7714) \end{aligned}$$

Since this interval does not include 30, we say 30 is not a candidate for the mean we are estimating and, therefore, μ is not equal to 30 and H_0 is rejected. This is the same conclusion reached by means of the hypothesis testing procedure.

If the hypothesized parameter, 30, had been within the 95 percent confidence interval, we would have said that H_0 is not rejected at the .05 level of significance. In general, *when testing a null hypothesis by means of a two-sided confidence interval, we reject H_0 at the α level of significance if the hypothesized parameter is not contained within the $100(1 - \alpha)$ percent confidence interval. If the hypothesized parameter is contained within the interval, H_0 cannot be rejected at the α level of significance.*

One-Sided Hypothesis Tests The hypothesis test illustrated by Example 7.2.1 is an example of a *two-sided test*, so called because the rejection region is split between the two sides or tails of the distribution of the test statistic. A hypothesis test may be *one-sided*, in which case all the rejection region is in one or the other tail of the distribution. Whether a one-sided or a two-sided test is used depends on the nature of the question being asked by the researcher.

If both large and small values will cause rejection of the null hypothesis, a two-sided test is indicated. When either sufficiently “small” values only or sufficiently “large” values only will cause rejection of the null hypothesis, a one-sided test is indicated.

EXAMPLE 7.2.2

Refer to Example 7.2.1. Suppose, instead of asking if they could conclude that $\mu \neq 30$, the researchers had asked: Can we conclude that $\mu < 30$? To this question we would reply that they can so conclude if they can reject the null hypothesis that $\mu \geq 30$.

Solution: Let us go through the ten-step procedure to reach a decision based on a one-sided test.

1. **Data.** See the previous example.
2. **Assumptions.** See the previous example.
3. **Hypotheses.**

$$\begin{aligned} H_0: \mu &\geq 30 \\ H_A: \mu &< 30 \end{aligned}$$

The inequality in the null hypothesis implies that the null hypothesis consists of an infinite number of hypotheses. The test will be made only

at the point of equality, since it can be shown that if H_0 is rejected when the test is made at the point of equality it would be rejected if the test were done for any other value of μ indicated in the null hypothesis.

4. Test statistic.

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

5. Distribution of test statistic. See the previous example.

6. Decision rule. Let us again use $\alpha = .05$. To determine where to place the rejection region, let us ask ourselves what magnitude of values would cause rejection of the null hypothesis. If we look at the hypotheses, we see that sufficiently small values would cause rejection and that large values would tend to reinforce the null hypothesis. We will want our rejection region to be where the small values are—at the lower tail of the distribution. This time, since we have a one-sided test, all of α will go in the one tail of the distribution. By consulting Appendix Table D, we find that the value of z to the left of which lies .05 of the area under the standard normal curve is -1.645 after interpolating. Our rejection and nonrejection regions are now specified and are shown in Figure 7.2.2.

Our decision rule tells us to reject H_0 if the computed value of the test statistic is less than or equal to -1.645 .

7. Calculation of test statistic. From our data we compute

$$z = \frac{27 - 30}{\sqrt{20/10}} = -2.12$$

8. Statistical decision. We are able to reject the null hypothesis since $-2.12 < -1.645$.

9. Conclusion. We conclude that the population mean is smaller than 30 and act accordingly.

10. p value. The p value for this test is .0170, since $P(z \leq -2.12)$, when H_0 is true, is .0170 as given by Appendix Table D when we determine the

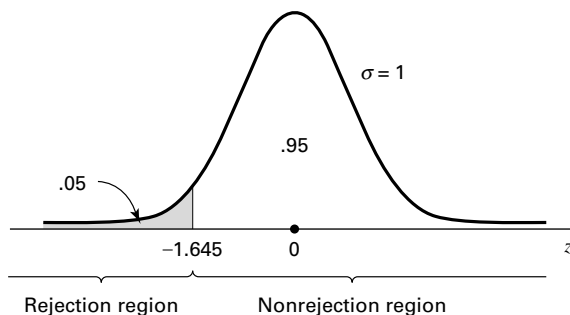


FIGURE 7.2.2 Rejection and nonrejection regions for Example 7.2.2.

magnitude of the area to the left of -2.12 under the standard normal curve. One can test a one-sided null hypothesis by means of a one-sided confidence interval. However, we will not cover the construction and interpretation of this type of confidence interval in this book.

If the researcher's question had been, "Can we conclude that the mean is greater than 30?", following the above ten-step procedure would have led to a one-sided test with all the rejection region at the upper tail of the distribution of the test statistic and a critical value of $+1.645$. ■

Sampling from a Normally Distributed Population: Population Variance Unknown

As we have already noted, the population variance is usually unknown in actual situations involving statistical inference about a population mean. When sampling is from an approximately normal population with an unknown variance, the test statistic for testing $H_0: \mu = \mu_0$ is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (7.2.2)$$

which, when H_0 is true, is distributed as Student's t with $n - 1$ degrees of freedom. The following example illustrates the hypothesis testing procedure when the population is assumed to be normally distributed and its variance is unknown. This is the usual situation encountered in practice.

EXAMPLE 7.2.3

Nakamura et al. (A-1) studied subjects with medial collateral ligament (MCL) and anterior cruciate ligament (ACL) tears. Between February 1995 and December 1997, 17 consecutive patients with combined acute ACL and grade III MCL injuries were treated by the same physician at the research center. One of the variables of interest was the length of time in days between the occurrence of the injury and the first magnetic resonance imaging (MRI). The data are shown in Table 7.2.1. We wish to know if we can conclude that the mean number of days between injury and initial MRI is not 15 days in a population presumed to be represented by these sample data.

TABLE 7.2.1 Number of Days Until MRI for Subjects with MCL and ACL Tears

Subject	Days	Subject	Days	Subject	Days	Subject	Days
1	14	6	0	11	28	16	14
2	9	7	10	12	24	17	9
3	18	8	4	13	24		
4	26	9	8	14	2		
5	12	10	21	15	3		

Source: Norimasa Nakamura, Shuji Horibe, Yukyoshi Toritsuka, Tomoki Mitsuoka, Hideki Yoshikawa, and Konsei Shino, "Acute Grade III Medial Collateral Ligament Injury of the Knee Associated with Anterior Cruciate Ligament Tear," *American Journal of Sports Medicine*, 31 (2003), 261–267.

Solution: We will be able to conclude that the mean number of days for the population is not 15 if we can reject the null hypothesis that the population mean is equal to 15.

- Data.** The data consist of number of days until MRI on 17 subjects as previously described.
- Assumptions.** The 17 subjects constitute a simple random sample from a population of similar subjects. We assume that the number of days until MRI in this population is approximately normally distributed.
- Hypotheses.**

$$H_0: \mu = 15$$

$$H_A: \mu \neq 15$$

- Test statistic.** Since the population variance is unknown, our test statistic is given by Equation 7.2.2.
- Distribution of test statistic.** Our test statistic is distributed as Student's t with $n - 1 = 17 - 1 = 16$ degrees of freedom if H_0 is true.
- Decision rule.** Let $\alpha = .05$. Since we have a two-sided test, we put $\alpha/2 = .025$ in each tail of the distribution of our test statistic. The t values to the right and left of which .025 of the area lies are 2.1199 and -2.1199 . These values are obtained from Appendix Table E. The rejection and nonrejection regions are shown in Figure 7.2.3.

The decision rule tells us to compute a value of the test statistic and reject H_0 if the computed t is either greater than or equal to 2.1199 or less than or equal to -2.1199 .

- Calculation of test statistic.** From our sample data we compute a sample mean of 13.2941 and a sample standard deviation of 8.88654. Substituting these statistics into Equation 7.2.2 gives

$$t = \frac{13.2941 - 15}{8.88654/\sqrt{17}} = \frac{-1.7059}{2.1553} = -.791$$

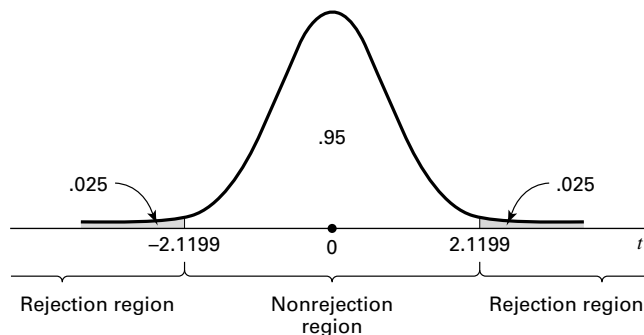


FIGURE 7.2.3 Rejection and nonrejection regions for Example 7.2.3.

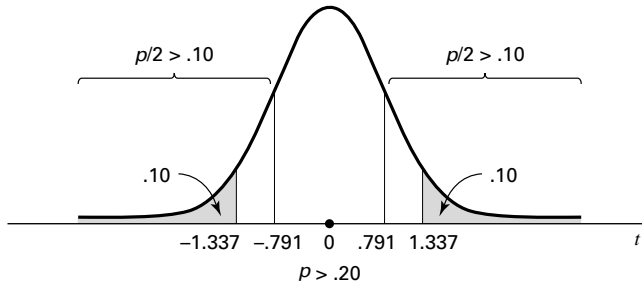


FIGURE 7.2.4 Determination of p value for Example 7.2.3.

8. **Statistical decision.** Do not reject H_0 , since $-.791$ falls in the non-rejection region.
9. **Conclusion.** Our conclusion, based on these data, is that the mean of the population from which the sample came may be 15.
10. **p value.** The exact p value for this test cannot be obtained from Appendix Table E since it gives t values only for selected percentiles. The p value can be stated as an interval, however. We find that $-.791$ is less than -1.337 , the value of t to the left of which lies .10 of the area under the t with 16 degrees of freedom. Consequently, when H_0 is true, the probability of obtaining a value of t as small as or smaller than $-.791$ is greater than .10. That is $P(t \leq -.791) > .10$. Since the test was two-sided, we must allow for the possibility of a computed value of the test statistic as large in the opposite direction as that observed. Appendix Table E reveals that $P(t \geq .791) > .10$ also. The p value, then, is $p > .20$. In fact, Excel calculates the p value to be .4403. Figure 7.2.4 shows the p value for this example.

If in the previous example the hypotheses had been

$$\begin{aligned} H_0: \mu &\geq 15 \\ H_A: \mu &< 15 \end{aligned}$$

the testing procedure would have led to a one-sided test with all the rejection region at the lower tail of the distribution, and if the hypotheses had been

$$\begin{aligned} H_0: \mu &\leq 15 \\ H_A: \mu &> 15 \end{aligned}$$

we would have had a one-sided test with all the rejection region at the upper tail of the distribution. ■

Sampling from a Population That Is Not Normally Distributed

If, as is frequently the case, the sample on which we base our hypothesis test about a population mean comes from a population that is not normally distributed, we may, if our sample is large (greater than or equal to 30), take advantage of the central limit theorem and use $z = (\bar{x} - \mu_0)/(\sigma/\sqrt{n})$ as the test statistic. If the population standard deviation is not

known, the usual practice is to use the sample standard deviation as an estimate. The test statistic for testing $H_0: \mu = \mu_0$, then, is

$$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (7.2.3)$$

which, when H_0 is true, is distributed approximately as the standard normal distribution if n is large. The rationale for using s to replace σ is that the large sample, necessary for the central limit theorem to apply, will yield a sample standard deviation that closely approximates σ .

EXAMPLE 7.2.4

The goal of a study by Klingler et al. (A-2) was to determine how symptom recognition and perception influence clinical presentation as a function of race. They characterized symptoms and care-seeking behavior in African-American patients with chest pain seen in the emergency department. One of the presenting vital signs was systolic blood pressure. Among 157 African-American men, the mean systolic blood pressure was 146 mm Hg with a standard deviation of 27. We wish to know if, on the basis of these data, we may conclude that the mean systolic blood pressure for a population of African-American men is greater than 140.

Solution: We will say that the data do provide sufficient evidence to conclude that the population mean is greater than 140 if we can reject the null hypothesis that the mean is less than or equal to 140. The following test may be carried out:

1. **Data.** The data consist of systolic blood pressure scores for 157 African-American men with $\bar{x} = 146$ and $s = 27$.
2. **Assumptions.** The data constitute a simple random sample from a population of African-American men who report to an emergency department with symptoms similar to those in the sample. We are unwilling to assume that systolic blood pressure values are normally distributed in such a population.
3. **Hypotheses.**

$$H_0: \mu \leq 140$$

$$H_A: \mu > 140$$

4. **Test statistic.** The test statistic is given by Equation 7.2.3, since s is unknown.
5. **Distribution of test statistic.** Because of the central limit theorem, the test statistic is at worst approximately normally distributed with $\mu = 0$ if H_0 is true.
6. **Decision rule.** Let $\alpha = .05$. The critical value of the test statistic is 1.645. The rejection and nonrejection regions are shown in Figure 7.2.5. Reject H_0 if computed $z \geq 1.645$.

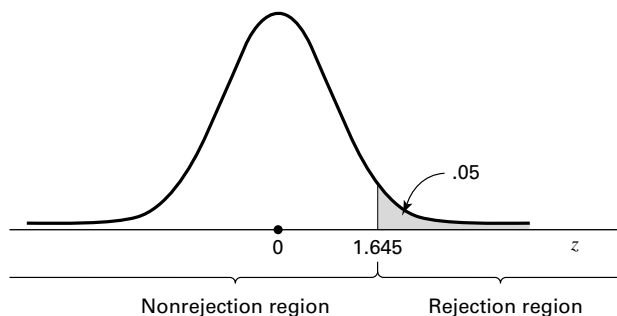


FIGURE 7.2.5 Rejection and nonrejection regions for Example 7.2.4.

7. Calculation of test statistic.

$$z = \frac{146 - 140}{27/\sqrt{157}} = \frac{6}{2.1548} = 2.78$$

8. Statistical decision. Reject H_0 since $2.78 > 1.645$.

9. Conclusion. Conclude that the mean systolic blood pressure for the sampled population is greater than 140.

10. p value. The p value for this test is $1 - .9973 = .0027$, since as shown in Appendix Table D, the area (.0027) to the right of 2.78 is less than .05, the area to the right of 1.645. ■

Procedures for Other Conditions If the population variance had been known, the procedure would have been identical to the above except that the known value of σ , instead of the sample value s , would have been used in the denominator of the computed test statistic.

Depending on what the investigators wished to conclude, either a two-sided test or a one-sided test, with the rejection region at the lower tail of the distribution, could have been made using the above data.

When testing a hypothesis about a single population mean, we may use Figure 6.3.3 to decide quickly whether the test statistic is z or t .

Computer Analysis To illustrate the use of computers in testing hypotheses, we consider the following example.

EXAMPLE 7.2.5

The following are the head circumferences (centimeters) at birth of 15 infants:

33.38	32.15	33.99	34.10	33.97
34.34	33.95	33.85	34.23	32.73
33.46	34.13	34.45	34.19	34.05

We wish to test $H_0: \mu = 34.5$ against $H_A: \mu \neq 34.5$.

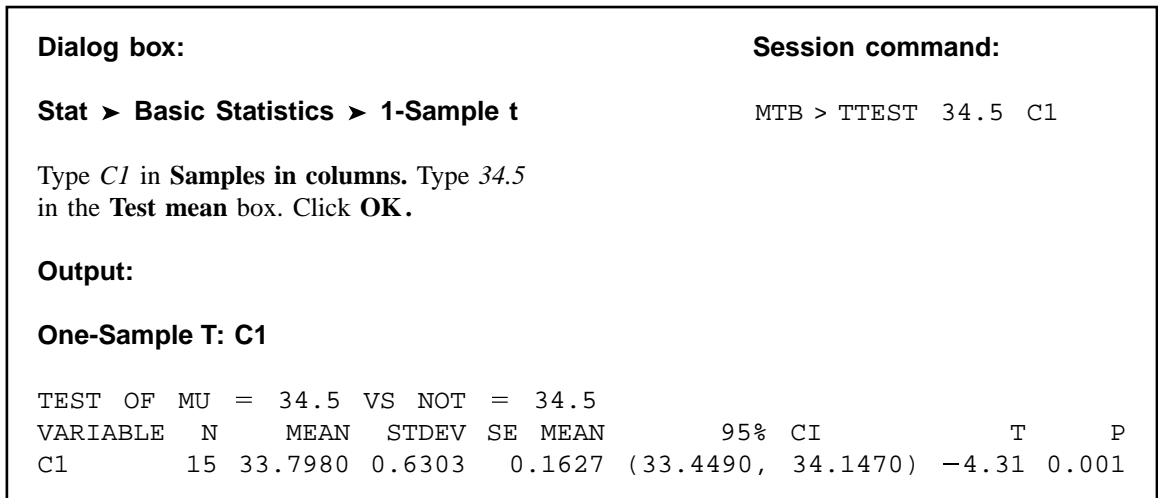


FIGURE 7.2.6 MINITAB procedure and output for Example 7.2.5.

Solution: We assume that the assumptions for use of the t statistic are met. We enter the data into Column 1 and proceed as shown in Figure 7.2.6.

To indicate that a test is one-sided when in Windows, click on the **Options** button and then choose “less than” or “greater than” as appropriate in the **Alternative** box. If z is the appropriate test statistic, we choose 1-Sample z from the Basic Statistics menu. The remainder of the commands are the same as for the t test.

We learn from the printout that the computed value of the test statistic is -4.31 and the p value for the test is $.0007$. SAS[®] users may use the output from PROC MEANS or PROC UNIVARIATE to perform hypothesis tests.

When both the z statistic and the t statistic are inappropriate test statistics for use with the available data, one may wish to use a non-parametric technique to test a hypothesis about a single population measure of central tendency. One such procedure, the sign test, is discussed in Chapter 13. ■

EXERCISES

For each of the following exercises carry out the ten-step hypothesis testing procedure for the given significance level. For each exercise, as appropriate, explain why you chose a one-sided test or a two-sided test. Discuss how you think researchers and/or clinicians might use the results of your hypothesis test. What clinical and/or research decisions and/or actions do you think would be appropriate in light of the results of your test?

- 7.2.1** Escobar et al. (A-3) performed a study to validate a translated version of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) questionnaire used with Spanish-speaking patients with hip or knee osteoarthritis. For the 76 women classified with severe hip pain, the

WOMAC mean function score (on a scale from 0 to 100 with a higher number indicating less function) was 70.7 with a standard deviation of 14.6. We wish to know if we may conclude that the mean function score for a population of similar women subjects with severe hip pain is less than 75. Let $\alpha = .01$.

- 7.2.2** A study by Thienprasiddhi et al. (A-4) examined a sample of 16 subjects with open-angle glaucoma and unilateral hemifield defects. The ages (years) of the subjects were:

62	62	68	48	51	60	51	57
57	41	62	50	53	34	62	61

Source: Phamornsak Thienprasiddhi, Vivienne C. Greenstein, Candice S. Chen, Jeffrey M. Liebmann, Robert Ritch, and Donald C. Hood, "Multifocal Visual Evoked Potential Responses in Glaucoma Patients with Unilateral Hemifield Defects," *American Journal of Ophthalmology*, 136 (2003), 34–40.

Can we conclude that the mean age of the population from which the sample may be presumed to have been drawn is less than 60 years? Let $\alpha = .05$.

- 7.2.3** The purpose of a study by Lugliè et al. (A-5) was to investigate the oral status of a group of patients diagnosed with thalassemia major (TM). One of the outcome measures was the decayed, missing, and filled teeth index (DMFT). In a sample of 18 patients the mean DMFT index value was 10.3 with a standard deviation of 7.3. Is this sufficient evidence to allow us to conclude that the mean DMFT index is greater than 9.0 in a population of similar subjects? Let $\alpha = .10$.
- 7.2.4** A study was made of a sample of 25 records of patients seen at a chronic disease hospital on an outpatient basis. The mean number of outpatient visits per patient was 4.8, and the sample standard deviation was 2. Can it be concluded from these data that the population mean is greater than four visits per patient? Let the probability of committing a type I error be $.05$. What assumptions are necessary?
- 7.2.5** In a sample of 49 adolescents who served as the subjects in an immunologic study, one variable of interest was the diameter of skin test reaction to an antigen. The sample mean and standard deviation were 21 and 11 mm erythema, respectively. Can it be concluded from these data that the population mean is less than 30? Let $\alpha = .05$.
- 7.2.6** Nine laboratory animals were infected with a certain bacterium and then immunosuppressed. The mean number of organisms later recovered from tissue specimens was 6.5 (coded data) with a standard deviation of $.6$. Can one conclude from these data that the population mean is greater than 6? Let $\alpha = .05$. What assumptions are necessary?
- 7.2.7** A sample of 25 freshman nursing students made a mean score of 77 on a test designed to measure attitude toward the dying patient. The sample standard deviation was 10. Do these data provide sufficient evidence to indicate, at the $.05$ level of significance, that the population mean is less than 80? What assumptions are necessary?
- 7.2.8** We wish to know if we can conclude that the mean daily caloric intake in the adult rural population of a developing country is less than 2000. A sample of 500 had a mean of 1985 and a standard deviation of 210. Let $\alpha = .05$.
- 7.2.9** A survey of 100 similar-sized hospitals revealed a mean daily census in the pediatrics service of 27 with a standard deviation of 6.5. Do these data provide sufficient evidence to indicate that the population mean is greater than 25? Let $\alpha = .05$.

- 7.2.10** Following a week-long hospital supervisory training program, 16 assistant hospital administrators made a mean score of 74 on a test administered as part of the evaluation of the training program. The sample standard deviation was 12. Can it be concluded from these data that the population mean is greater than 70? Let $\alpha = .05$. What assumptions are necessary?
- 7.2.11** A random sample of 16 emergency reports was selected from the files of an ambulance service. The mean time (computed from the sample data) required for ambulances to reach their destinations was 13 minutes. Assume that the population of times is normally distributed with a variance of 9. Can we conclude at the .05 level of significance that the population mean is greater than 10 minutes?
- 7.2.12** The following data are the oxygen uptakes (milliliters) during incubation of a random sample of 15 cell suspensions:

14.0, 14.1, 14.5, 13.2, 11.2, 14.0, 14.1, 12.2,
11.1, 13.7, 13.2, 16.0, 12.8, 14.4, 12.9

Do these data provide sufficient evidence at the .05 level of significance that the population mean is not 12 ml? What assumptions are necessary?

- 7.2.13** Can we conclude that the mean maximum voluntary ventilation value for apparently healthy college seniors is not 110 liters per minute? A sample of 20 yielded the following values:

132, 33, 91, 108, 67, 169, 54, 203, 190, 133,
96, 30, 187, 21, 63, 166, 84, 110, 157, 138

Let $\alpha = .01$. What assumptions are necessary?

- 7.2.14** The following are the systolic blood pressures (mm Hg) of 12 patients undergoing drug therapy for hypertension:

183, 152, 178, 157, 194, 163, 144, 114, 178, 152, 118, 158

Can we conclude on the basis of these data that the population mean is less than 165? Let $\alpha = .05$. What assumptions are necessary?

- 7.2.15** Can we conclude that the mean age at death of patients with homozygous sickle-cell disease is less than 30 years? A sample of 50 patients yielded the following ages in years:

15.5	2.0	45.1	1.7	.8	1.1	18.2	9.7	28.1	18.2
27.6	45.0	1.0	66.4	2.0	67.4	2.5	61.7	16.2	31.7
6.9	13.5	1.9	31.2	9.0	2.6	29.7	13.5	2.6	14.4
20.7	30.9	36.6	1.1	23.6	.9	7.6	23.5	6.3	40.2
23.7	4.8	33.2	27.1	36.7	3.2	38.0	3.5	21.8	2.4

Let $\alpha = .05$. What assumptions are necessary?

- 7.2.16** The following are intraocular pressure (mm Hg) values recorded for a sample of 21 elderly subjects:

14.5	12.9	14.0	16.1	12.0	17.5	14.1	12.9	17.9	12.0
16.4	24.2	12.2	14.4	17.0	10.0	18.5	20.8	16.2	14.9
19.6									

Can we conclude from these data that the mean of the population from which the sample was drawn is greater than 14? Let $\alpha = .05$. What assumptions are necessary?

- 7.2.17** Suppose it is known that the IQ scores of a certain population of adults are approximately normally distributed with a standard deviation of 15. A simple random sample of 25 adults drawn from this population had a mean IQ score of 105. On the basis of these data can we conclude that the mean IQ score for the population is not 100? Let the probability of committing a type I error be .05.
- 7.2.18** A research team is willing to assume that systolic blood pressures in a certain population of males are approximately normally distributed with a standard deviation of 16. A simple random sample of 64 males from the population had a mean systolic blood pressure reading of 133. At the .05 level of significance, do these data provide sufficient evidence for us to conclude that the population mean is greater than 130?
- 7.2.19** A simple random sample of 16 adults drawn from a certain population of adults yielded a mean weight of 63 kg. Assume that weights in the population are approximately normally distributed with a variance of 49. Do the sample data provide sufficient evidence for us to conclude that the mean weight for the population is less than 70 kg? Let the probability of committing a type I error be .01.

7.3 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION MEANS

Hypothesis testing involving the difference between two population means is most frequently employed to determine whether or not it is reasonable to conclude that the two population means are unequal. In such cases, one or the other of the following hypotheses may be formulated:

1. $H_0: \mu_1 - \mu_2 = 0, \quad H_A: \mu_1 - \mu_2 \neq 0$
2. $H_0: \mu_1 - \mu_2 \geq 0, \quad H_A: \mu_1 - \mu_2 < 0$
3. $H_0: \mu_1 - \mu_2 \leq 0, \quad H_A: \mu_1 - \mu_2 > 0$

It is possible, however, to test the hypothesis that the difference is equal to, greater than or equal to, or less than or equal to some value other than zero.

As was done in the previous section, hypothesis testing involving the difference between two population means will be discussed in three different contexts: (1) when sampling is from normally distributed populations with known population variances, (2) when sampling is from normally distributed populations with unknown population variances, and (3) when sampling is from populations that are not normally distributed.

Sampling from Normally Distributed Populations: Population Variances Known When each of two independent simple random samples has been drawn from a normally distributed population with a known variance, the test statistic for testing the null hypothesis of equal population means is

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (7.3.1)$$

where the subscript 0 indicates that the difference is a hypothesized parameter. When H_0 is true the test statistic of Equation 7.3.1 is distributed as the standard normal.

EXAMPLE 7.3.1

Researchers wish to know if the data they have collected provide sufficient evidence to indicate a difference in mean serum uric acid levels between normal individuals and individuals with Down's syndrome. The data consist of serum uric acid readings on 12 individuals with Down's syndrome and 15 normal individuals. The means are $\bar{x}_1 = 4.5$ mg/100 ml and $\bar{x}_2 = 3.4$ mg/100 ml.

Solution: We will say that the sample data do provide evidence that the population means are not equal if we can reject the null hypothesis that the population means are equal. Let us reach a conclusion by means of the ten-step hypothesis testing procedure.

1. **Data.** See problem statement.
2. **Assumptions.** The data constitute two independent simple random samples each drawn from a normally distributed population with a variance equal to 1 for the Down's syndrome population and 1.5 for the normal population.
3. **Hypotheses.**

$$\begin{aligned}H_0: \mu_1 - \mu_2 &= 0 \\H_A: \mu_1 - \mu_2 &\neq 0\end{aligned}$$

An alternative way of stating the hypotheses is as follows:

$$\begin{aligned}H_0: \mu_1 &= \mu_2 \\H_A: \mu_1 &\neq \mu_2\end{aligned}$$

4. **Test statistic.** The test statistic is given by Equation 7.3.1.
5. **Distribution of test statistic.** When the null hypothesis is true, the test statistic follows the standard normal distribution.
6. **Decision rule.** Let $\alpha = .05$. The critical values of z are ± 1.96 . Reject H_0 unless $-1.96 < z_{\text{computed}} < 1.96$. The rejection and nonrejection regions are shown in Figure 7.3.1.
7. **Calculation of test statistic.**

$$z = \frac{(4.5 - 3.4) - 0}{\sqrt{1/12 + 1.5/15}} = \frac{1.1}{.4282} = 2.57$$

8. **Statistical decision.** Reject H_0 , since $2.57 > 1.96$.
9. **Conclusion.** Conclude that, on the basis of these data, there is an indication that the two population means are not equal.
10. **p value.** For this test, $p = .0102$.

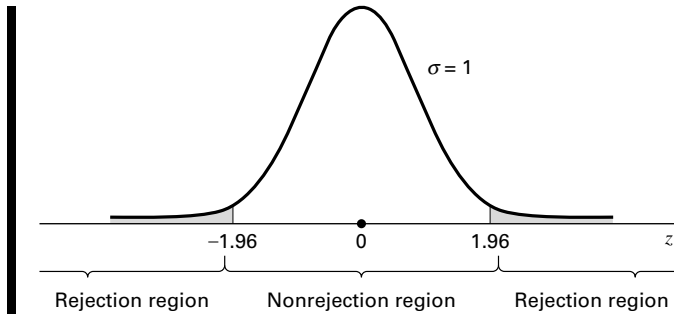


FIGURE 7.3.1 Rejection and nonrejection regions for Example 7.3.1

A 95 Percent Confidence Interval for $\mu_1 - \mu_2$ In the previous chapter the 95 percent confidence interval for $\mu_1 - \mu_2$, computed from the same data, was found to be .26 to 1.94. Since this interval does not include 0, we say that 0 is not a candidate for the difference between population means, and we conclude that the difference is not zero. Thus we arrive at the same conclusion by means of a confidence interval.

Sampling from Normally Distributed Populations: Population Variances Unknown As we have learned, when the population variances are unknown, two possibilities exist. The two population variances may be equal or they may be unequal. We consider first the case where it is known, or it is reasonable to assume, that they are equal. A test of the hypothesis that two population variances are equal is described in Section 7.8.

Population Variances Equal When the population variances are unknown, but assumed to be equal, we recall from Chapter 6 that it is appropriate to pool the sample variances by means of the following formula:

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

When each of two independent simple random samples has been drawn from a normally distributed population and the two populations have equal but unknown variances, the test statistic for testing $H_0: \mu_1 = \mu_2$ is given by

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}} \quad (7.3.2)$$

which, when H_0 is true, is distributed as Student's t with $n_1 + n_2 - 2$ degrees of freedom.

EXAMPLE 7.3.2

The purpose of a study by Tam et al. (A-6) was to investigate wheelchair maneuvering in individuals with lower-level spinal cord injury (SCI) and healthy controls (C). Subjects used a modified wheelchair to incorporate a rigid seat surface to facilitate the specified experimental measurements. Interface pressure measurement was recorded by using a high-resolution pressure-sensitive mat with a spatial resolution of four sensors per square centimeter taped on the rigid seat support. During static sitting conditions, average pressures were recorded under the ischial tuberosities (the bottom part of the pelvic bones). The data for measurements of the left ischial tuberosity (in mm Hg) for the SCI and control groups are shown in Table 7.3.1. We wish to know if we may conclude, on the basis of these data, that, in general, healthy subjects exhibit lower pressure than SCI subjects.

Solution:

- 1. Data.** See statement of problem.
- 2. Assumptions.** The data constitute two independent simple random samples of pressure measurements, one sample from a population of control subjects and the other sample from a population with lower-level spinal cord injury. We shall assume that the pressure measurements in both populations are approximately normally distributed. The population variances are unknown but are assumed to be equal.
- 3. Hypotheses.** $H_0: \mu_C \geq \mu_{SCI}$, $H_A: \mu_C < \mu_{SCI}$.
- 4. Test statistic.** The test statistic is given by Equation 7.3.2.
- 5. Distribution of test statistic.** When the null hypothesis is true, the test statistic follows Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom.
- 6. Decision rule.** Let $\alpha = .05$. The critical value of t is -1.7341 . Reject H_0 unless $t_{\text{computed}} > -1.7341$.
- 7. Calculation of test statistic.** From the sample data we compute

$$\bar{x}_C = 126.1, \quad s_C = 21.8, \quad \bar{x}_{SCI} = 133.1, \quad s_{SCI} = 32.2$$

Next, we pool the sample variances to obtain

$$s_p^2 = \frac{9(21.8)^2 + 9(32.2)^2}{9 + 9} = 756.04$$

TABLE 7.3.1 Pressures (mm Hg) Under the Pelvis during Static Conditions for Example 7.3.2

Control	131	115	124	131	122	117	88	114	150	169
SCI	60	150	130	180	163	130	121	119	130	148

Source: Eric W. Tam, Arthur F. Mak, Wai Nga Lam, John H. Evans, and York Y. Chow, "Pelvic Movement and Interface Pressure Distribution During Manual Wheelchair Propulsion," *Archives of Physical Medicine and Rehabilitation*, 84 (2003), 1466–1472.

We now compute

$$t = \frac{(126.1 - 133.1) - 0}{\sqrt{\frac{756.04}{10} + \frac{756.04}{10}}} = -.569$$

8. **Statistical decision.** We fail to reject H_0 , since $-1.7341 < -.569$; that is, $-.569$ falls in the nonrejection region.
9. **Conclusion.** On the basis of these data, we cannot conclude that the population mean pressure is less for healthy subjects than for SCI subjects.
10. **p value.** For this test, $p > .10$ using Table E, or $.5764$ using a computer since $-1.330 < -.569$. ■

Population Variances Unequal When two independent simple random samples have been drawn from normally distributed populations with unknown and unequal variances, the test statistic for testing $H_0: \mu_1 = \mu_2$ is

$$t' = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (7.3.3)$$

The critical value of t' for an α level of significance and a two-sided test is approximately

$$t'_{1-(\alpha/2)} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2} \quad (7.3.4)$$

where $w_1 = s_1^2/n_1$, $w_2 = s_2^2/n_2$, $t_1 = t_{1-(\alpha/2)}$ for $n_1 - 1$ degrees of freedom, and $t_2 = t_{1-(\alpha/2)}$ for $n_2 - 1$ degrees of freedom. The critical value of t' for a one-sided test is found by computing $t'_{1-\alpha}$ by Equation 7.3.4, using $t_1 = t_{1-\alpha}$ for $n_1 - 1$ degrees of freedom and $t_2 = t_{1-\alpha}$ for $n_2 - 1$ degrees of freedom.

For a two-sided test, reject H_0 if the computed value of t' is either greater than or equal to the critical value given by Equation 7.3.4 or less than or equal to the negative of that value.

For a one-sided test with the rejection region in the right tail of the sampling distribution, reject H_0 if the computed t' is equal to or greater than the critical t' . For a one-sided test with a left-tail rejection region, reject H_0 if the computed value of t' is equal to or smaller than the negative of the critical t' computed by the indicated adaptation of Equation 7.3.4.

EXAMPLE 7.3.3

Dernellis and Panaretou (A-7) examined subjects with hypertension and healthy control subjects. One of the variables of interest was the aortic stiffness index. Measures of this variable were calculated from the aortic diameter evaluated by M-mode echocardiography and blood pressure measured by a sphygmomanometer. Generally, physicians wish to

reduce aortic stiffness. In the 15 patients with hypertension (group 1), the mean aortic stiffness index was 19.16 with a standard deviation of 5.29. In the 30 control subjects (group 2), the mean aortic stiffness index was 9.53 with a standard deviation of 2.69. We wish to determine if the two populations represented by these samples differ with respect to mean aortic stiffness index.

Solution:

1. **Data.** The sample sizes, means, and sample standard deviations are:

$$\begin{aligned} n_1 &= 15, & \bar{x}_1 &= 19.16, & s_1 &= 5.29 \\ n_2 &= 30, & \bar{x}_2 &= 9.53, & s_2 &= 2.69 \end{aligned}$$

2. **Assumptions.** The data constitute two independent random samples, one from a population of subjects with hypertension and the other from a control population. We assume that aortic stiffness values are approximately normally distributed in both populations. The population variances are unknown and unequal.

3. **Hypotheses.**

$$\begin{aligned} H_0: \mu_1 - \mu_2 &= 0 \\ H_A: \mu_1 - \mu_2 &\neq 0 \end{aligned}$$

4. **Test statistic.** The test statistic is given by Equation 7.3.3.
5. **Distribution of test statistic.** The statistic given by Equation 7.3.3 does not follow Student's t distribution. We, therefore, obtain its critical values by Equation 7.3.4.
6. **Decision rule.** Let $\alpha = .05$. Before computing t' we calculate $w_1 = (5.29)^2/15 = 1.8656$ and $w_2 = (2.69)^2/30 = .2412$. In Appendix Table E we find that $t_1 = 2.1448$ and $t_2 = 2.0452$. By Equation 7.3.4 we compute

$$t' = \frac{1.8656(2.1448) + .2412(2.0452)}{1.8656 + .2412} = 2.133$$

Our decision rule, then, is reject H_0 if the computed t is either ≥ 2.133 or ≤ -2.133 .

7. **Calculation of test statistic.** By Equation 7.3.3 we compute

$$t' = \frac{(19.16 - 9.53) - 0}{\sqrt{\frac{(5.29)^2}{15} + \frac{(2.69)^2}{30}}} = \frac{9.63}{1.4515} = 6.63$$

8. **Statistical decision.** Since $6.63 > 2.133$, we reject H_0 .
9. **Conclusion.** On the basis of these results we conclude that the two population means are different.
10. **p value.** For this test $p < .05$; program R calculates this value to be $< .00001$. ■

Sampling from Populations That Are Not Normally Distributed

When sampling is from populations that are not normally distributed, the results of the central limit theorem may be employed if sample sizes are large (say, ≥ 30). This will allow the use of normal theory since the distribution of the difference between sample means will be approximately normal. When each of two large independent simple random samples has been drawn from a population that is not normally distributed, the test statistic for testing $H_0: \mu_1 = \mu_2$ is

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad (7.3.5)$$

which, when H_0 is true, follows the standard normal distribution. If the population variances are known, they are used; but if they are unknown, as is the usual case, the sample variances, which are necessarily based on large samples, are used as estimates. Sample variances are not pooled, since equality of population variances is not a necessary assumption when the z statistic is used.

EXAMPLE 7.3.4

The objective of a study by Sairam et al. (A-8) was to identify the role of various disease states and additional risk factors in the development of thrombosis. One focus of the study was to determine if there were differing levels of the anticardiolipin antibody IgG in subjects with and without thrombosis. Table 7.3.2 summarizes the researchers' findings:

TABLE 7.3.2 IgG Levels for Subjects With and Without Thrombosis for Example 7.3.4

Group	Mean IgG Level (ml/unit)	Sample Size	Standard Deviation
Thrombosis	59.01	53	44.89
No thrombosis	46.61	54	34.85

Source: S. Sairam, B. A. Baethge and T. McNearney, "Analysis of Risk Factors and Comorbid Diseases in the Development of Thrombosis in Patients with Anticardiolipin Antibodies," *Clinical Rheumatology*, 22 (2003), 24–29.

We wish to know if we may conclude, on the basis of these results, that, in general, persons with thrombosis have, on the average, higher IgG levels than persons without thrombosis.

Solution:

1. **Data.** See statement of example.
2. **Assumptions.** The statistics were computed from two independent samples that behave as simple random samples from a population of persons with thrombosis and a population of persons who do not have thrombosis. Since the population variances are unknown, we will use the sample variances in the calculation of the test statistic.
3. **Hypotheses.**

$$\begin{aligned} H_0: \mu_T - \mu_{NT} &\leq 0 \\ H_A: \mu_T - \mu_{NT} &> 0 \end{aligned}$$

or, alternatively,

$$\begin{aligned} H_0: \mu_T &\leq \mu_{NT} \\ H_A: \mu_T &> \mu_{NT} \end{aligned}$$

4. **Test statistic.** Since we have large samples, the central limit theorem allows us to use Equation 7.3.5 as the test statistic.
5. **Distribution of test statistic.** When the null hypothesis is true, the test statistic is distributed approximately as the standard normal.
6. **Decision rule.** Let $\alpha = .01$. This is a one-sided test with a critical value of z equal to 2.33. Reject H_0 if $z_{\text{computed}} \geq 2.33$.
7. **Calculation of test statistic.**

$$z = \frac{59.01 - 46.61}{\sqrt{\frac{44.89^2}{53} + \frac{34.85^2}{54}}} = 1.59$$

8. **Statistical decision.** Fail to reject H_0 , since $z = 1.59$ is in the non-rejection region.
9. **Conclusion.** These data indicate that on the average, persons with thrombosis and persons without thrombosis may not have differing IgG levels.
10. **p value.** For this test, $p = .0559$. When testing a hypothesis about the difference between two populations means, we may use Figure 6.4.1 to decide quickly whether the test statistic should be z or t . ■

We may use MINITAB to perform two-sample t tests. To illustrate, let us refer to the data in Table 7.3.1. We put the data for control subjects and spinal cord

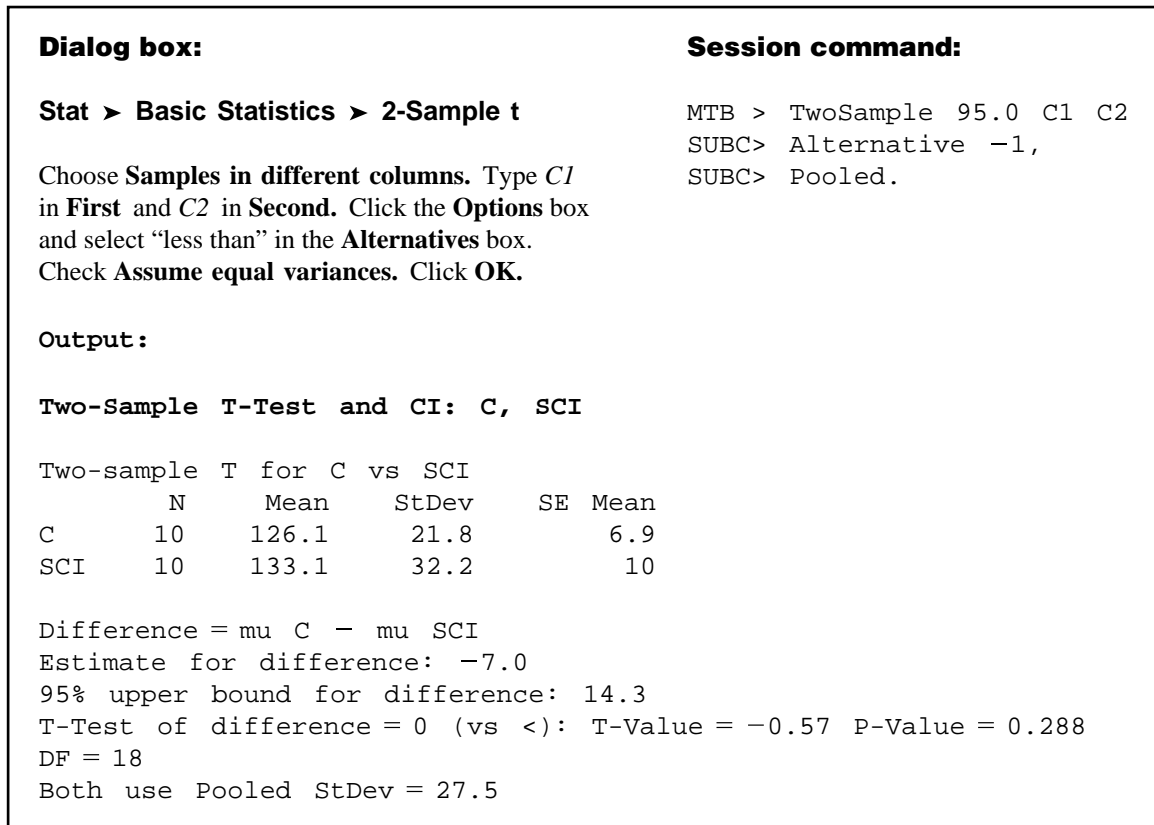


FIGURE 7.3.2 MINITAB procedure and output for two-sample t test, Example 7.3.2 (data in Table 7.3.1).

injury subjects in Column 1 and Column 2, respectively, and proceed as shown in Figure 7.3.2.

The SAS[®] statistical package performs the t test for equality of population means under both assumptions regarding population variances: that they are equal and that they are not equal. Note that SAS[®] designates the p value as $\text{Pr} > |t|$. The default output is a p value for a two-sided test. The researcher using SAS[®] must divide this quantity in half when the hypothesis test is one-sided. The SAS[®] package also tests for equality of population variances as described in Section 7.8. Figure 7.3.3 shows the SAS[®] output for Example 7.3.2.

Alternatives to z and t Sometimes neither the z statistic nor the t statistic is an appropriate test statistic for use with the available data. When such is the case, one may wish to use a nonparametric technique for testing a hypothesis about the difference between two population measures of central tendency. The Mann-Whitney test statistic and the median test, discussed in Chapter 13, are frequently used alternatives to the z and t statistics.

```

The SAS System
The TTEST Procedure

                                Statistics
Variable group          Lower CL  Upper CL  Lower CL  Upper CL
                        N    Mean    Mean    Std      Std      Std      Std
-----
pressure C              10  110.49  126.1   141.71  15.008   21.82   39.834   6.9
pressure SCI            10  110.08  133.1   156.12  22.133   32.178  58.745  10.176
pressure Diff (1-2)    -32.83   -7    18.83  20.773  27.491  40.655  12.294

                                T-Tests
-----
Variable  Method          Variances    DF      t Value    Pr > |t|
pressure  Pooled            Equal        18      -0.57      0.5761
pressure  Satterthwaite      Unequal      15.8    -0.57      0.5771

                                Equality of Variances
-----
Variable  Method          Num DF      Den DF      F Value    Pr > F
pressure  Folded F          9           9           2.17      0.2626

```

FIGURE 7.3.3 SAS® output for Example 7.3.2 (data in Table 7.3.1).

EXERCISES

In each of the following exercises, complete the ten-step hypothesis testing procedure. State the assumptions that are necessary for your procedure to be valid. For each exercise, as appropriate, explain why you chose a one-sided test or a two-sided test. Discuss how you think researchers or clinicians might use the results of your hypothesis test. What clinical or research decisions or actions do you think would be appropriate in light of the results of your test?

- 7.3.1** Subjects in a study by Dabonneville et al. (A-9) included a sample of 40 men who claimed to engage in a variety of sports activities (multisport). The mean body mass index (BMI) for these men was 22.41 with a standard deviation of 1.27. A sample of 24 male rugby players had a mean BMI of 27.75 with a standard deviation of 2.64. Is there sufficient evidence for one to claim that, in general, rugby players have a higher BMI than the multisport men? Let $\alpha = .01$.
- 7.3.2** The purpose of a study by Ingle and Eastell (A-10) was to examine the bone mineral density (BMD) and ultrasound properties of women with ankle fractures. The investigators recruited 31 postmenopausal women with ankle fractures and 31 healthy postmenopausal women to serve as controls. One of the baseline measurements was the stiffness index of the lunar Achilles. The mean stiffness index for the ankle fracture group was 76.9 with a standard deviation of 12.6. In the control group, the mean was 90.9 with a standard deviation of 12.5. Do these data provide sufficient evidence to allow you to conclude that, in general, the mean stiffness index is higher in

healthy postmenopausal women than in postmenopausal women with ankle fractures? Let $\alpha = .05$.

- 7.3.3** Hoekema et al. (A-11) studied the craniofacial morphology of 26 male patients with obstructive sleep apnea syndrome (OSAS) and 37 healthy male subjects (non-OSAS). One of the variables of interest was the length from the most superoanterior point of the body of the hyoid bone to the Frankfort horizontal (measured in millimeters).

Length (mm) Non-OSAS				Length (mm) OSAS		
96.80	97.00	101.00	88.95	105.95	114.90	113.70
100.70	97.70	88.25	101.05	114.90	114.35	116.30
94.55	97.00	92.60	92.60	110.35	112.25	108.75
99.65	94.55	98.25	97.00	123.10	106.15	113.30
109.15	106.45	90.85	91.95	119.30	102.60	106.00
102.75	94.55	95.25	88.95	110.00	102.40	101.75
97.70	94.05	88.80	95.75	98.95	105.05	
92.10	89.45	101.40		114.20	112.65	
91.90	89.85	90.55		108.95	128.95	
89.50	98.20	109.80		105.05	117.70	

Source: Data provided courtesy of A. Hoekema, D.D.S.

Do these data provide sufficient evidence to allow us to conclude that the two sampled populations differ with respect to length from the hyoid bone to the Frankfort horizontal? Let $\alpha = .01$.

- 7.3.4** Can we conclude that patients with primary hypertension (PH), on the average, have higher total cholesterol levels than normotensive (NT) patients? This was one of the inquiries of interest for Rossi et al. (A-12). In the following table are total cholesterol measurements (mg/dl) for 133 PH patients and 41 NT patients. Can we conclude that PH patients have, on average, higher total cholesterol levels than NT patients? Let $\alpha = .05$.

Total Cholesterol (mg/dl)						
Primary Hypertensive Patients				Normotensive Patients		
207	221	212	220	190	286	189
172	223	260	214	245	226	196
191	181	210	215	171	187	142
221	217	265	206	261	204	179
203	208	206	247	182	203	212
241	202	198	221	162	206	163
208	218	210	199	182	196	196
199	216	211	196	225	168	189
185	168	274	239	203	229	142
235	168	223	199	195	184	168
214	214	175	244	178	186	121
134	203	203	214	240	281	
226	280	168	236	222	203	

(Continued)

Total Cholesterol (mg/dl)						
Primary Hypertensive Patients				Normotensive Patients		
222	203	178	249	117	177	135
213	225	217	212	252	179	161
272	227	200	259	203	194	
185	239	226	189	245	206	
181	265	207	235	218	219	
238	228	232	239	152	173	
141	226	182	239	231	189	
203	236	215	210	237	194	
222	195	239	203		196	
221	284	210	188		212	
180	183	207	237		168	
276	266	224	231		188	
226	258	251	222		232	
224	214	212	174		242	
206	260	201	219		200	

Source: Data provided courtesy of Gian Paolo Rossi, M.D., F.A.C.C., F.A.H.A.

- 7.3.5** Garção and Cabrita (A-13) wanted to evaluate the community pharmacist's capacity to positively influence the results of antihypertensive drug therapy through a pharmaceutical care program in Portugal. Eighty-two subjects with essential hypertension were randomly assigned to an intervention or a control group. The intervention group received monthly monitoring by a research pharmacist to monitor blood pressure, assess adherence to treatment, prevent, detect, and resolve drug-related problems, and encourage nonpharmacologic measures for blood pressure control. The changes after 6 months in diastolic blood pressure (pre – post, mm Hg) are given in the following table for patients in each of the two groups.

Intervention Group				Control Group			
20	4	12	16	0	4	12	0
2	24	6	10	12	2	2	8
36	6	24	16	18	2	0	10
26	-2	42	10	0	8	0	14
2	8	20	6	8	10	-4	8
20	8	14	6	10	0	12	0
2	16	-2	2	8	6	4	2
14	14	10	8	14	10	28	-8
30	8	2	16	4	-2	-18	16
18	20	18	-12	-2	2	12	12
6				-6			

Source: Data provided courtesy of José Garção, M.S., Pharm.D.

On the basis of these data, what should the researcher conclude? Let $\alpha = .05$.

- 7.3.6** A test designed to measure mothers' attitudes toward their labor and delivery experiences was given to two groups of new mothers. Sample 1 (attenders) had attended prenatal classes held at the local

health department. Sample 2 (nonattenders) did not attend the classes. The sample sizes and means and standard deviations of the test scores were as follows:

Sample	n	\bar{x}	s
1	15	4.75	1.0
2	22	3.00	1.5

Do these data provide sufficient evidence to indicate that attenders, on the average, score higher than nonattenders? Let $\alpha = .05$.

- 7.3.7** Cortisol level determinations were made on two samples of women at childbirth. Group 1 subjects underwent emergency cesarean section following induced labor. Group 2 subjects delivered by either cesarean section or the vaginal route following spontaneous labor. The sample sizes, mean cortisol levels, and standard deviations were as follows:

Sample	n	\bar{x}	s
1	10	435	65
2	12	645	80

Do these data provide sufficient evidence to indicate a difference in the mean cortisol levels in the populations represented? Let $\alpha = .05$.

- 7.3.8** Protoporphyrin levels were measured in two samples of subjects. Sample 1 consisted of 50 adult male alcoholics with ring sideroblasts in the bone marrow. Sample 2 consisted of 40 apparently healthy adult nonalcoholic males. The mean protoporphyrin levels and standard deviations for the two samples were as follows:

Sample	\bar{x}	s
1	340	250
2	45	25

Can one conclude on the basis of these data that protoporphyrin levels are higher in the represented alcoholic population than in the nonalcoholic population? Let $\alpha = .01$.

- 7.3.9** A researcher was interested in knowing if preterm infants with late metabolic acidosis and preterm infants without the condition differ with respect to urine levels of a certain chemical. The mean levels, standard deviations, and sample sizes for the two samples studied were as follows:

Sample	n	\bar{x}	s
With condition	35	8.5	5.5
Without condition	40	4.8	3.6

What should the researcher conclude on the basis of these results? Let $\alpha = .05$.

- 7.3.10** Researchers wished to know if they could conclude that two populations of infants differ with respect to mean age at which they walked alone. The following data (ages in months) were collected:

Sample from population A: 9.5, 10.5, 9.0, 9.75, 10.0, 13.0,
10.0, 13.5, 10.0, 9.5, 10.0, 9.75
Sample from population B: 12.5, 9.5, 13.5, 13.75, 12.0, 13.75,
12.5, 9.5, 12.0, 13.5, 12.0, 12.0

What should the researchers conclude? Let $\alpha = .05$.

- 7.3.11** Does sensory deprivation have an effect on a person's alpha-wave frequency? Twenty volunteer subjects were randomly divided into two groups. Subjects in group A were subjected to a 10-day period of sensory deprivation, while subjects in group B served as controls. At the end of the experimental period, the alpha-wave frequency component of subjects' electroencephalograms was measured. The results were as follows:

Group A: 10.2, 9.5, 10.1, 10.0, 9.8, 10.9, 11.4, 10.8, 9.7, 10.4
Group B: 11.0, 11.2, 10.1, 11.4, 11.7, 11.2, 10.8, 11.6, 10.9, 10.9

Let $\alpha = .05$.

- 7.3.12** Can we conclude that, on the average, lymphocytes and tumor cells differ in size? The following are the cell diameters (μm) of 40 lymphocytes and 50 tumor cells obtained from biopsies of tissue from patients with melanoma:

Lymphocytes							
9.0	9.4	4.7	4.8	8.9	4.9	8.4	5.9
6.3	5.7	5.0	3.5	7.8	10.4	8.0	8.0
8.6	7.0	6.8	7.1	5.7	7.6	6.2	7.1
7.4	8.7	4.9	7.4	6.4	7.1	6.3	8.8
8.8	5.2	7.1	5.3	4.7	8.4	6.4	8.3

Tumor Cells									
12.6	14.6	16.2	23.9	23.3	17.1	20.0	21.0	19.1	19.4
16.7	15.9	15.8	16.0	17.9	13.4	19.1	16.6	18.9	18.7
20.0	17.8	13.9	22.1	13.9	18.3	22.8	13.0	17.9	15.2
17.7	15.1	16.9	16.4	22.8	19.4	19.6	18.4	18.2	20.7
16.3	17.7	18.1	24.3	11.2	19.5	18.6	16.4	16.1	21.5

Let $\alpha = .05$.

7.4 PAIRED COMPARISONS

In our previous discussion involving the difference between two population means, it was assumed that the samples were independent. A method frequently employed for assessing the effectiveness of a treatment or experimental procedure is one that makes use of related

observations resulting from nonindependent samples. A hypothesis test based on this type of data is known as a *paired comparisons* test.

Reasons for Pairing It frequently happens that true differences do not exist between two populations with respect to the variable of interest, but the presence of extraneous sources of variation may cause rejection of the null hypothesis of no difference. On the other hand, true differences also may be masked by the presence of extraneous factors.

Suppose, for example, that we wish to compare two sunscreens. There are at least two ways in which the experiment may be carried out. One method would be to select a simple random sample of subjects to receive sunscreen A and an independent simple random sample of subjects to receive sunscreen B. We send the subjects out into the sunshine for a specified length of time, after which we will measure the amount of damage from the rays of the sun. Suppose we employ this method, but inadvertently, most of the subjects receiving sunscreen A have darker complexions that are naturally less sensitive to sunlight. Let us say that after the experiment has been completed we find that subjects receiving sunscreen A had less sun damage. We would not know if they had less sun damage because sunscreen A was more protective than sunscreen B or because the subjects were naturally less sensitive to the sun.

A better way to design the experiment would be to select just one simple random sample of subjects and let each member of the sample receive both sunscreens. We could, for example, randomly assign the sunscreens to the left or the right side of each subject's back with each subject receiving both sunscreens. After a specified length of exposure to the sun, we would measure the amount of sun damage to each half of the back. If the half of the back receiving sunscreen A tended to be less damaged, we could more confidently attribute the result to the sunscreen, since in each instance both sunscreens were applied to equally pigmented skin.

The objective in paired comparisons tests is to eliminate a maximum number of sources of extraneous variation by making the pairs similar with respect to as many variables as possible.

Related or paired observations may be obtained in a number of ways. The same subjects may be measured before and after receiving some treatment. Litter mates of the same sex may be assigned randomly to receive either a treatment or a placebo. Pairs of twins or siblings may be assigned randomly to two treatments in such a way that members of a single pair receive different treatments. In comparing two methods of analysis, the material to be analyzed may be divided equally so that one-half is analyzed by one method and one-half is analyzed by the other. Or pairs may be formed by matching individuals on some characteristic, for example, digital dexterity, which is closely related to the measurement of interest, say, posttreatment scores on some test requiring digital manipulation.

Instead of performing the analysis with individual observations, we use d_i , the difference between pairs of observations, as the variable of interest.

When the n sample differences computed from the n pairs of measurements constitute a simple random sample from a normally distributed population of differences, the test statistic for testing hypotheses about the population mean difference μ_d is

$$t = \frac{\bar{d} - \mu_{d_0}}{s_{\bar{d}}} \quad (7.4.1)$$

where \bar{d} is the sample mean difference, μ_{d_0} is the hypothesized population mean difference, $s_{\bar{d}} = s_d/\sqrt{n}$, n is the number of sample differences, and s_d is the standard deviation of the sample differences. When H_0 is true, the test statistic is distributed as Student's t with $n - 1$ degrees of freedom.

Although to begin with we have two samples—say, before levels and after levels—we do not have to worry about equality of variances, as with independent samples, since our variable is the difference between readings in the same individual, or matched individuals, and, hence, only one variable is involved. The arithmetic involved in performing a paired comparisons test, therefore, is the same as for performing a test involving a single sample as described in Section 7.2.

The following example illustrates the procedures involved in a paired comparisons test.

EXAMPLE 7.4.1

John M. Morton et al. (A-14) examined gallbladder function before and after fundoplication—a surgery used to stop stomach contents from flowing back into the esophagus (reflux)—in patients with gastroesophageal reflux disease. The authors measured gallbladder functionality by calculating the gallbladder ejection fraction (GBEF) before and after fundoplication. The goal of fundoplication is to increase GBEF, which is measured as a percent. The data are shown in Table 7.4.1. We wish to know if these data provide sufficient evidence to allow us to conclude that fundoplication increases GBEF functioning.

Solution: We will say that sufficient evidence is provided for us to conclude that the fundoplication is effective if we can reject the null hypothesis that the population mean change μ_d is different from zero in the appropriate direction. We may reach a conclusion by means of the ten-step hypothesis testing procedure.

- 1. Data.** The data consist of the GBEF for 12 individuals, before and after fundoplication. We shall perform the statistical analysis on the differences in preop and postop GBEF. We may obtain the differences in one of two ways: by subtracting the preop percents from the postop percents or by subtracting the postop percents from the preop percents. Let us

TABLE 7.4.1 Gallbladder Function in Patients with Presentations of Gastroesophageal Reflux Disease Before and After Treatment

Preop (%)	22	63.3	96	9.2	3.1	50	33	69	64	18.8	0	34
Postop (%)	63.5	91.5	59	37.8	10.1	19.6	41	87.8	86	55	88	40

Source: John M. Morton, Steven P. Bowers, Tananchai A. Lucktong, Samer Mattar, W. Alan Bradshaw, Kevin E. Behrns, Mark J. Koruda, Charles A. Herbst, William McCartney, Raghuvveer K. Halkar, C. Daniel Smith, and Timothy M. Farrell, "Gallbladder Function Before and After Fundoplication," *Journal of Gastrointestinal Surgery*, 6 (2002), 806–811.

obtain the differences by subtracting the preop percents from the postop percents. The $d_i = \text{postop} - \text{preop}$ differences are:

41.5, 28.2, -37.0, 28.6, 7.0, -30.4, 8.0, 18.8, 22.0, 36.2, 88.0, 6.0

2. **Assumptions.** The observed differences constitute a simple random sample from a normally distributed population of differences that could be generated under the same circumstances.
3. **Hypotheses.** The way we state our null and alternative hypotheses must be consistent with the way in which we subtract measurements to obtain the differences. In the present example, we want to know if we can conclude that the fundoplication is useful in increasing GBEF percentage. If it is effective in improving GBEF, we would expect the postop percents to tend to be higher than the preop percents. If, therefore, we subtract the preop percents from the postop percents (postop - preop), we would expect the differences to tend to be positive. Furthermore, we would expect the mean of a population of such differences to be positive. So, under these conditions, asking if we can conclude that the fundoplication is effective is the same as asking if we can conclude that the population mean difference is positive (greater than zero).

The null and alternative hypotheses are as follows:

$$H_0: \mu_d \leq 0$$

$$H_A: \mu_d > 0$$

If we had obtained the differences by subtracting the postop percents from the preop weights (preop - postop), our hypotheses would have been

$$H_0: \mu_d \geq 0$$

$$H_A: \mu_d < 0$$

If the question had been such that a two-sided test was indicated, the hypotheses would have been

$$H_0: \mu_d = 0$$

$$H_A: \mu_d \neq 0$$

regardless of the way we subtracted to obtain the differences.

4. **Test statistic.** The appropriate test statistic is given by Equation 7.4.1.
5. **Distribution of test statistic.** If the null hypothesis is true, the test statistic is distributed as Student's t with $n - 1$ degrees of freedom.
6. **Decision rule.** Let $\alpha = .05$. The critical value of t is 1.7959. Reject H_0 if computed t is greater than or equal to the critical value. The rejection and nonrejection regions are shown in Figure 7.4.1.

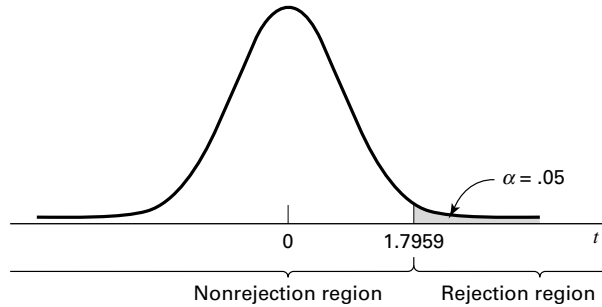


FIGURE 7.4.1 Rejection and nonrejection regions for Example 7.4.1.

7. Calculation of test statistic. From the $n = 12$ differences d_i , we compute the following descriptive measures:

$$\bar{d} = \frac{\sum d_i}{n} = \frac{(41.5) + (28.2) + (-37.0) + \cdots + (6.0)}{12} = \frac{216.9}{12} = 18.075$$

$$s_d^2 = \frac{\sum (d_i - \bar{d})^2}{n - 1} = \frac{n \sum d_i^2 - (\sum d_i)^2}{n(n - 1)} = \frac{12(15669.49) - (216.9)^2}{(12)(11)} = 1068.0930$$

$$t = \frac{18.075 - 0}{\sqrt{1068.0930/12}} = \frac{18.075}{9.4344} = 1.9159$$

8. Statistical decision. Reject H_0 , since 1.9159 is in the rejection region.

Paired T-Test and CI: C2, C1

Paired T for C2 - C1

	N	Mean	StDev	SE Mean
C2	12	56.6083	27.8001	8.0252
C1	12	38.5333	30.0587	8.6772
Difference	12	18.0750	32.6817	9.4344

95% lower bound for mean difference: 1.1319

T-Test of mean difference = 0 (vs > 0): T-Value = 1.92 P-Value = 0.041

FIGURE 7.4.2 MINITAB procedure and output for paired comparisons test, Example 7.4.1 (data in Table 7.4.1).

9. Conclusion. We may conclude that the fundoplication procedure increases GBEF functioning.

10. p value. For this test, $.025 < p < .05$, since $1.7959 < 1.9159 < 2.2010$. MINITAB provides the exact p value as .041 (Figure 7.4.2). ■

A Confidence Interval for μ_d A 95 percent confidence interval for μ_d may be obtained as follows:

$$\begin{aligned} \bar{d} \pm t_{1-(\alpha/2)}s_{\bar{d}} \\ 18.075 \pm 2.2010\sqrt{1068.0930/12} \\ 18.075 \pm 20.765 \\ (-2.690, 38.840) \end{aligned}$$

The Use of z If, in the analysis of paired data, the population variance of the differences is known, the appropriate test statistic is

$$z = \frac{\bar{d} - \mu_d}{\sigma_d/\sqrt{n}} \quad (7.4.2)$$

It is unlikely that σ_d will be known in practice.

If the assumption of normally distributed d_i 's cannot be made, the central limit theorem may be employed if n is large. In such cases, the test statistic is Equation 7.4.2, with s_d used to estimate σ_d when, as is generally the case, the latter is unknown.

Disadvantages The use of the paired comparisons test is not without its problems. If different subjects are used and randomly assigned to two treatments, considerable time and expense may be involved in our trying to match individuals on one or more relevant variables. A further price we pay for using paired comparisons is a loss of degrees of freedom. If we do not use paired observations, we have $2n - 2$ degrees of freedom available as compared to $n - 1$ when we use the paired comparisons procedure.

In general, in deciding whether or not to use the paired comparisons procedure, one should be guided by the economics involved as well as by a consideration of the gains to be realized in terms of controlling extraneous variation.

Alternatives If neither z nor t is an appropriate test statistic for use with available data, one may wish to consider using some nonparametric technique to test a hypothesis about a median difference. The sign test, discussed in Chapter 13, is a candidate for use in such cases.

EXERCISES

In the following exercises, carry out the ten-step hypothesis testing procedure at the specified significance level. For each exercise, as appropriate, explain why you chose a one-sided test or a two-sided test. Discuss how you think researchers or clinicians might use the results of your hypothesis test. What clinical or research decisions or actions do you think would be appropriate in light of the results of your test?

- 7.4.1** Ellen Davis Jones (A-15) studied the effects of reminiscence therapy for older women with depression. She studied 15 women 60 years or older residing for 3 months or longer in an assisted living long-term care facility. For this study, depression was measured by the Geriatric Depression Scale (GDS). Higher scores indicate more severe depression symptoms. The participants received reminiscence therapy for long-term care, which uses family photographs, scrapbooks, and personal memorabilia to stimulate memory and conversation among group members. Pre-treatment and post-treatment depression scores are given in the following table. Can we conclude, based on these data, that subjects who participate in reminiscence therapy experience, on average, a decline in GDS depression scores? Let $\alpha = .01$.

Pre-GDS: 12 10 16 2 12 18 11 16 16 10 14 21 9 19 20
 Post-GDS: 11 10 11 3 9 13 8 14 16 10 12 22 9 16 18

Source: Data provided courtesy of Ellen Davis Jones, N.D., R.N., FNP-C.

- 7.4.2** Beney et al. (A-16) evaluated the effect of telephone follow-up on the physical well-being dimension of health-related quality of life in patients with cancer. One of the main outcome variables was measured by the physical well-being subscale of the Functional Assessment of Cancer Therapy Scale-General (FACT-G). A higher score indicates higher physical well-being. The following table shows the baseline FACT-G score and the follow-up score to evaluate the physical well-being during the 7 days after discharge from hospital to home for 66 patients who received a phone call 48–72 hours after discharge that gave patients the opportunity to discuss medications, problems, and advice. Is there sufficient evidence to indicate that quality of physical well-being significantly decreases in the first week of discharge among patients who receive a phone call? Let $\alpha = .05$.

Subject	Baseline FACT-G	Follow-up FACT-G	Subject	Baseline FACT-G	Follow-up FACT-G
1	16	19	34	25	14
2	26	19	35	21	17
3	13	9	36	14	22
4	20	23	37	23	22
5	22	25	38	19	16
6	21	20	39	19	15
7	20	10	40	18	23
8	15	20	41	20	21
9	25	22	42	18	11
10	20	18	43	22	22
11	11	6	44	7	17
12	22	21	45	23	9
13	18	17	46	19	16
14	21	13	47	17	16
15	25	25	48	22	20
16	17	21	49	19	23
17	26	22	50	5	17
18	18	22	51	22	17
19	7	9	52	12	6
20	25	24	53	19	19
21	22	15	54	17	20
22	15	9	55	7	6

(Continued)

Subject	Baseline FACT-G	Follow-up FACT-G	Subject	Baseline FACT-G	Follow-up FACT-G
23	19	7	56	27	10
24	23	20	57	22	16
25	19	19	58	16	14
26	21	24	59	26	24
27	24	23	60	17	19
28	21	15	61	23	22
29	28	27	62	23	23
30	18	26	63	13	3
31	25	26	64	24	22
32	25	26	65	17	21
33	28	28	66	22	21

Source: Data provided courtesy of Johnny Beney, Ph.D. and E. Beth Devine, Pharm.D., M.B.A. et al.

- 7.4.3** The purpose of an investigation by Morley et al. (A-17) was to evaluate the analgesic effectiveness of a daily dose of oral methadone in patients with chronic neuropathic pain syndromes. The researchers used a visual analogue scale (0–100 mm, higher number indicates higher pain) ratings for maximum pain intensity over the course of the day. Each subject took either 20 mg of methadone or a placebo each day for 5 days. Subjects did not know which treatment they were taking. The following table gives the mean maximum pain intensity scores for the 5 days on methadone and the 5 days on placebo. Do these data provide sufficient evidence, at the .05 level of significance, to indicate that in general the maximum pain intensity is lower on days when methadone is taken?

Subject	Methadone	Placebo
1	29.8	57.2
2	73.0	69.8
3	98.6	98.2
4	58.8	62.4
5	60.6	67.2
6	57.2	70.6
7	57.2	67.8
8	89.2	95.6
9	97.0	98.4
10	49.8	63.2
11	37.0	63.6

Source: John S. Morley, John Bridson, Tim P. Nash, John B. Miles, Sarah White, and Matthew K. Makin, "Low-Dose Methadone Has an Analgesic Effect in Neuropathic Pain: A Double-Blind Randomized Controlled Crossover Trial," *Palliative Medicine*, 17 (2003), 576–587.

- 7.4.4** Woo and McKenna (A-18) investigated the effect of broadband ultraviolet B (UVB) therapy and topical calcipotriol cream used together on areas of psoriasis. One of the outcome variables is the Psoriasis Area and Severity Index (PASI). The following table gives the PASI scores for 20 subjects measured at baseline and after eight treatments. Do these data provide sufficient evidence, at the .01 level of significance, to indicate that the combination therapy reduces PASI scores?

Subject	Baseline	After 8 Treatments
1	5.9	5.2
2	7.6	12.2
3	12.8	4.6
4	16.5	4.0
5	6.1	0.4
6	14.4	3.8
7	6.6	1.2
8	5.4	3.1
9	9.6	3.5
10	11.6	4.9
11	11.1	11.1
12	15.6	8.4
13	6.9	5.8
14	15.2	5.0
15	21.0	6.4
16	5.9	0.0
17	10.0	2.7
18	12.2	5.1
19	20.2	4.8
20	6.2	4.2

Source: Data provided courtesy of W. K. Woo, M.D.

- 7.4.5** One of the purposes of an investigation by Porcellini et al. (A-19) was to investigate the effect on CD4 T cell count of administration of intermittent interleukin (IL-2) in addition to highly active antiretroviral therapy (HAART). The following table shows the CD4 T cell count at baseline and then again after 12 months of HAART therapy with IL-2. Do the data show, at the .05 level, a significant change in CD4 T cell count?

Subject	1	2	3	4	5	6	7
CD4 T cell count at entry ($\times 10^6/L$)	173	58	103	181	105	301	169
CD4 T cell count at end of follow-up ($\times 10^6/L$)	257	108	315	362	141	549	369

Source: Simona Procclini, Giuliana Vallanti, Silvia Nozza, Guido Poli, Adriano Lazzarin, Guiseppe Tabussi, and Antonio Grassia. "Improved Thymopoietic Potential in Aviremic HIV-Infected Individuals with HAART by Intermittent IL-2 Administration," *AIDS*, 17 (2003), 1621–1630.

7.5 HYPOTHESIS TESTING: A SINGLE POPULATION PROPORTION

Testing hypotheses about population proportions is carried out in much the same way as for means when the conditions necessary for using the normal curve are met. One-sided or two-sided tests may be made, depending on the question being asked. When a sample

sufficiently large for application of the central limit theorem as discussed in Section 5.5 is available for analysis, the test statistic is

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \quad (7.5.1)$$

which, when H_0 is true, is distributed approximately as the standard normal.

EXAMPLE 7.5.1

Wagenknecht et al. (A-20) collected data on a sample of 301 Hispanic women living in San Antonio, Texas. One variable of interest was the percentage of subjects with impaired fasting glucose (IFG). IFG refers to a metabolic stage intermediate between normal glucose homeostasis and diabetes. In the study, 24 women were classified in the IFG stage. The article cites population estimates for IFG among Hispanic women in Texas as 6.3 percent. Is there sufficient evidence to indicate that the population of Hispanic women in San Antonio has a prevalence of IFG higher than 6.3 percent?

Solution:

- Data.** The data are obtained from the responses of 301 individuals of which 24 possessed the characteristic of interest; that is, $\hat{p} = 24/301 = .080$.
- Assumptions.** The study subjects may be treated as a simple random sample from a population of similar subjects, and the sampling distribution of \hat{p} is approximately normally distributed in accordance with the central limit theorem.
- Hypotheses.**

$$H_0: p \leq .063$$

$$H_A: p > .063$$

We conduct the test at the point of equality. The conclusion we reach will be the same as we would reach if we conducted the test using any other hypothesized value of p greater than .063. If H_0 is true, $p = .063$ and the standard error $\sigma_{\hat{p}} = \sqrt{(.063)(.937)/301}$. Note that we use the hypothesized value of p in computing $\sigma_{\hat{p}}$. We do this because the entire test is based on the assumption that the null hypothesis is true. To use the sample proportion, \hat{p} , in computing $\sigma_{\hat{p}}$ would not be consistent with this concept.

- Test statistic.** The test statistic is given by Equation 7.5.1.
- Distribution of test statistic.** If the null hypothesis is true, the test statistic is approximately normally distributed with a mean of zero.
- Decision rule.** Let $\alpha = .05$. The critical value of z is 1.645. Reject H_0 if the computed z is ≥ 1.645 .

7. Calculation of test statistic.

$$z = \frac{.080 - .063}{\sqrt{\frac{(.063)(.937)}{301}}} = 1.21$$

8. Statistical decision. Do not reject H_0 since $1.21 < 1.645$.

9. Conclusion. We cannot conclude that in the sampled population the proportion who are IFG is higher than 6.3 percent.

10. p value. $p = .1131$. ■

Tests involving a single proportion can be carried out using a variety of computer programs. Outputs from MINITAB and NCSS, using the data from Example 7.5.1, are shown in Figure 7.5.1. It should be noted that the results will vary slightly, because of rounding errors, if calculations are done by hand. It should also be noted that some programs, such as NCSS, use a continuity correction in calculating the z -value, and therefore the test statistic values and corresponding p values differ slightly from the MINITAB output.

MINITAB Output						
Test and CI for One Proportion						
Test of $p = 0.063$ vs $p > 0.063$						
			95% Lower			
Sample	X	N	Sample p	Bound	Z-Value	P-Value
1	24	301	0.079734	0.054053	1.19	0.116
Using the normal approximation.						
NCSS Output						
Normal Approximation using (P0)						
Alternative Hypothesis	Z-Value	Prob Level	Decision (5%)			
$P <> P_0$	1.0763	0.281780	Accept H_0			
$P < P_0$	1.0763	0.859110	Accept H_0			
$P > P_0$	1.0763	0.140890	Accept H_0			

FIGURE 7.5.1 MINITAB and partial NCSS output for the data in Example 7.5.1.

EXERCISES

For each of the following exercises, carry out the ten-step hypothesis testing procedure at the designated level of significance. For each exercise, as appropriate, explain why you chose a one-sided test or a two-sided test. Discuss how you think researchers or clinicians might use the results of your hypothesis test. What clinical or research decisions or actions do you think would be appropriate in light of the results of your test?

- 7.5.1** Jacquemyn et al. (A-21) conducted a survey among gynecologists-obstetricians in the Flanders region and obtained 295 responses. Of those responding, 90 indicated that they had performed at least one cesarean section on demand every year. Does this study provide sufficient evidence for us to conclude that less than 35 percent of the gynecologists-obstetricians in the Flanders region perform at least one cesarean section on demand each year? Let $\alpha = .05$.
- 7.5.2** In an article in the journal *Health and Place*, Hui and Bell (A-22) found that among 2428 boys ages 7 to 12 years, 461 were overweight or obese. On the basis of this study, can we conclude that more than 15 percent of the boys ages 7 to 12 in the sampled population are obese or overweight? Let $\alpha = .05$.
- 7.5.3** Becker et al. (A-23) conducted a study using a sample of 50 ethnic Fijian women. The women completed a self-report questionnaire on dieting and attitudes toward body shape and change. The researchers found that five of the respondents reported at least weekly episodes of binge eating during the previous 6 months. Is this sufficient evidence to conclude that less than 20 percent of the population of Fijian women engage in at least weekly episodes of binge eating? Let $\alpha = .05$.
- 7.5.4** The following questionnaire was completed by a simple random sample of 250 gynecologists. The number checking each response is shown in the appropriate box.
- When you have a choice, which procedure do you prefer for obtaining samples of endometrium?
 - Dilation and curettage 175
 - Vobra aspiration 75
 - Have you seen one or more pregnant women during the past year whom you knew to have elevated blood lead levels?
 - Yes 25
 - No 225
 - Do you routinely acquaint your pregnant patients who smoke with the suspected hazards of smoking to the fetus?
 - Yes 238
 - No 12

Can we conclude from these data that in the sampled population more than 60 percent prefer dilation and curettage for obtaining samples of endometrium? Let $\alpha = .01$.

- 7.5.5** Refer to Exercise 7.5.4. Can we conclude from these data that in the sampled population fewer than 15 percent have seen (during the past year) one or more pregnant women with elevated blood lead levels? Let $\alpha = .05$.
- 7.5.6** Refer to Exercise 7.5.4. Can we conclude from these data that more than 90 percent acquaint their pregnant patients who smoke with the suspected hazards of smoking to the fetus? Let $\alpha = .05$.

7.6 HYPOTHESIS TESTING: THE DIFFERENCE BETWEEN TWO POPULATION PROPORTIONS

The most frequent test employed relative to the difference between two population proportions is that their difference is zero. It is possible, however, to test that the difference is equal to some other value. Both one-sided and two-sided tests may be made.

When the null hypothesis to be tested is $p_1 - p_2 = 0$, we are hypothesizing that the two population proportions are equal. We use this as justification for combining the results of the two samples to come up with a pooled estimate of the hypothesized common proportion. If this procedure is adopted, one computes

$$\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}, \quad \text{and} \quad \bar{q} = 1 - \bar{p}$$

where x_1 and x_2 are the numbers in the first and second samples, respectively, possessing the characteristic of interest. This pooled estimate of $p = p_1 = p_2$ is used in computing $\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}$, the estimated standard error of the estimator, as follows:

$$\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_1} + \frac{\bar{p}(1 - \bar{p})}{n_2}} \quad (7.6.1)$$

The test statistic becomes

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}} \quad (7.6.2)$$

which is distributed approximately as the standard normal if the null hypothesis is true.

EXAMPLE 7.6.1

Noonan syndrome is a genetic condition that can affect the heart, growth, blood clotting, and mental and physical development. Noonan et al. (A-24) examined the stature of men and women with Noonan syndrome. The study contained 29 male and 44 female adults. One of the cut-off values used to assess stature was the third percentile of adult height. Eleven of the males fell below the third percentile of adult male height, while 24 of the females fell below the third percentile of female adult height. Does this study provide sufficient evidence for us to conclude that among subjects with Noonan syndrome, females are more likely than males to fall below the respective third percentile of adult height? Let $\alpha = .05$.

Solution:

- Data.** The data consist of information regarding the height status of Noonan syndrome males and females as described in the statement of the example.
- Assumptions.** We assume that the patients in the study constitute independent simple random samples from populations of males and females with Noonan syndrome.
- Hypotheses.**

$$H_0: p_F \leq p_M \quad \text{or} \quad p_F - p_M \leq 0$$

$$H_A: p_F > p_M \quad \text{or} \quad p_F - p_M > 0$$

where p_F is the proportion of females below the third percentile of female adult height and p_M is the proportion of males below the third percentile of male adult height.

- Test statistic.** The test statistic is given by Equation 7.6.2.
- Distribution of test statistic.** If the null hypothesis is true, the test statistic is distributed approximately as the standard normal.
- Decision rule.** Let $\alpha = .05$. The critical value of z is 1.645. Reject H_0 if computed z is greater than 1.645.
- Calculation of test statistic.** From the sample data we compute $\hat{p}_F = 24/44 = .545$, $\hat{p}_M = 11/29 = .379$, and $\bar{p} = (24 + 11)/(44 + 29) = .479$. The computed value of the test statistic, then, is

$$z = \frac{(.545 - .379)}{\sqrt{\frac{(.479)(.521)}{44} + \frac{(.479)(.521)}{29}}} = 1.39$$

- Statistical decision.** Fail to reject H_0 since $1.39 < 1.645$.
- Conclusion.** In the general population of adults with Noonan syndrome there may be no difference in the proportion of males and females who have heights below the third percentile of adult height.
- p value.** For this test $p = .0823$. ■

Tests involving two proportions, using the data from Example 7.6.1, can be carried out with a variety of computer programs. Outputs from MINITAB and NCSS are shown in Figure 7.6.1. Again, it should be noted that, because of rounding errors, the results will vary slightly if calculations are done by hand.

MINITAB Output**Test and CI for Two Proportions**

Sample	X	N	Sample p
1	24	44	0.545455
2	11	29	0.379310

Difference = $p(1) - p(2)$

Estimate for difference: 0.166144

95% lower bound for difference: -0.0267550

Test for difference = 0 (vs > 0): $Z = 1.39$ P-Value = 0.082

NCSS Output

Test Name	Test Statistic's Distribution	Test Statistic Value	Prob Level	Conclude H1 at 5% Significance?
Z-Test	Normal	1.390	0.0822	No

FIGURE 7.6.1 MINITAB and partial NCSS output for the data in Example 7.6.1.

EXERCISES

In each of the following exercises use the ten-step hypothesis testing procedure. For each exercise, as appropriate, explain why you chose a one-sided test or a two-sided test. Discuss how you think researchers or clinicians might use the results of your hypothesis test. What clinical or research decisions or actions do you think would be appropriate in light of the results of your test?

- 7.6.1** Ho et al. (A-25) used telephone interviews of randomly selected respondents in Hong Kong to obtain information regarding individuals' perceptions of health and smoking history. Among 1222 current male smokers, 72 reported that they had "poor" or "very poor" health, while 30 among 282 former male smokers reported that they had "poor" or "very poor" health. Is this sufficient evidence to allow one to conclude that among Hong Kong men there is a difference between current and former smokers with respect to the proportion who perceive themselves as having "poor" and "very poor" health? Let $\alpha = .01$.
- 7.6.2** Landolt et al. (A-26) examined rates of posttraumatic stress disorder (PTSD) in mothers and fathers. Parents were interviewed 5 to 6 weeks after an accident or a new diagnosis of cancer or diabetes mellitus type I for their child. Twenty-eight of the 175 fathers interviewed and 43 of the 180 mothers

interviewed met the criteria for current PTSD. Is there sufficient evidence for us to conclude that fathers are less likely to develop PTSD than mothers when a child is traumatized by an accident, cancer diagnosis, or diabetes diagnosis? Let $\alpha = .05$.

- 7.6.3** In a *Kidney International* article, Avram et al. (A-27) reported on a study involving 529 hemodialysis patients and 326 peritoneal dialysis patients. They found that at baseline 249 subjects in the hemodialysis treatment group were diabetic, while at baseline 134 of the subjects in the peritoneal dialysis group were diabetic. Is there a significant difference in diabetes prevalence at baseline between the two groups of this study? Let $\alpha = .05$. What does your finding regarding sample significance imply about the populations of subjects?
- 7.6.4** In a study of obesity the following results were obtained from samples of males and females between the ages of 20 and 75:

	<i>n</i>	Number Overweight
Males	150	21
Females	200	48

Can we conclude from these data that in the sampled populations there is a difference in the proportions who are overweight? Let $\alpha = .05$.

7.7 HYPOTHESIS TESTING: A SINGLE POPULATION VARIANCE

In Section 6.9 we examined how it is possible to construct a confidence interval for the variance of a normally distributed population. The general principles presented in that section may be employed to test a hypothesis about a population variance. When the data available for analysis consist of a simple random sample drawn from a normally distributed population, the test statistic for testing hypotheses about a population variance is

$$\chi^2 = (n - 1)s^2/\sigma^2 \quad (7.7.1)$$

which, when H_0 is true, is distributed as χ^2 with $n - 1$ degrees of freedom.

EXAMPLE 7.7.1

The purpose of a study by Wilkins et al. (A-28) was to measure the effectiveness of recombinant human growth hormone (rhGH) on children with total body surface area burns > 40 percent. In this study, 16 subjects received daily injections at home of rhGH. At baseline, the researchers wanted to know the current levels of insulin-like growth factor (IGF-I) prior to administration of rhGH. The sample variance of IGF-I levels (in ng/ml) was 670.81. We wish to know if we may conclude from these data that the population variance is not 600.

Solution:

- Data.** See statement in the example.
- Assumptions.** The study sample constitutes a simple random sample from a population of similar children. The IGF-I levels are normally distributed.
- Hypotheses.**

$$H_0: \sigma^2 = 600$$

$$H_A: \sigma^2 \neq 600$$

- Test statistic.** The test statistic is given by Equation 7.7.1.
- Distribution of test statistic.** When the null hypothesis is true, the test statistic is distributed as χ^2 with $n - 1$ degrees of freedom.
- Decision rule.** Let $\alpha = .05$. Critical values of χ^2 are 6.262 and 27.488. Reject H_0 unless the computed value of the test statistic is between 6.262 and 27.488. The rejection and nonrejection regions are shown in Figure 7.7.1.
- Calculation of test statistic.**

$$\chi^2 = \frac{15(670.81)}{600} = 16.77$$

- Statistical decision.** Do not reject H_0 since $6.262 < 16.77 < 27.488$.
- Conclusion.** Based on these data we are unable to conclude that the population variance is not 600.
- p value.** The determination of the p value for this test is complicated by the fact that we have a two-sided test and an asymmetric sampling distribution. When we have a two-sided test and a symmetric sampling distribution such as the standard normal or t , we may, as we have seen, double the one-sided p value. Problems arise when we attempt to

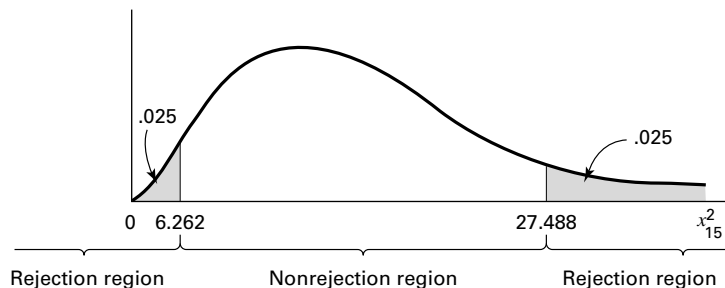


FIGURE 7.7.1 Rejection and nonrejection regions for Example 7.7.1.

do this with an asymmetric sampling distribution such as the chi-square distribution. In this situation the one-sided p value is reported along with the direction of the observed departure from the null hypothesis. In fact, this procedure may be followed in the case of symmetric sampling distributions. Precedent, however, seems to favor doubling the one-sided p value when the test is two-sided and involves a symmetric sampling distribution.

For the present example, then, we may report the p value as follows: $p > .05$ (two-sided test). A population variance greater than 600 is suggested by the sample data, but this hypothesis is not strongly supported by the test.

If the problem is stated in terms of the population standard deviation, one may square the sample standard deviation and perform the test as indicated above. ■

One-Sided Tests Although this was an example of a two-sided test, one-sided tests may also be made by logical modification of the procedure given here.

For $H_A: \sigma^2 > \sigma_0^2$, reject H_0 if computed $\chi^2 \geq \chi_{1-\alpha}^2$

For $H_A: \sigma^2 < \sigma_0^2$, reject H_0 if computed $\chi^2 \leq \chi_{\alpha}^2$

Tests involving a single population variance can be carried out using MINITAB software. Most other statistical computer programs lack procedures for carrying out these tests directly. The output from MINITAB, using the data from Example 7.7.1, is shown in Figure 7.7.2.

Test and CI for One Variance			
Statistics			
N	StDev	Variance	
16	25.9	671	
95% Confidence Intervals			
Method	CI for StDev	CI for Variance	
Standard	(19.1, 40.1)	(366, 1607)	
Tests			
Method	Chi-Square	DF	P-Value
Standard	16.77	15	0.666

FIGURE 7.7.2 MINITAB output for the data in Example 7.7.1.

EXERCISES

In each of the following exercises, carry out the ten-step testing procedure. For each exercise, as appropriate, explain why you chose a one-sided test or a two-sided test. Discuss how you think researchers or clinicians might use the results of your hypothesis test. What clinical or research decisions or actions do you think would be appropriate in light of the results of your test?

- 7.7.1 Recall Example 7.2.3, where Nakamura et al. (A-1) studied subjects with acute medial collateral ligament injury (MCL) with anterior cruciate ligament tear (ACL). The ages of the 17 subjects were:

31, 26, 21, 15, 26, 16, 19, 21, 28, 27, 22, 20, 25, 31, 20, 25, 15

Use these data to determine if there is sufficient evidence for us to conclude that in a population of similar subjects, the variance of the ages of the subjects is not 20 years. Let $\alpha = .01$.

- 7.7.2 Robinson et al. (A-29) studied nine subjects who underwent baffle procedure for transposition of the great arteries (TGA). At baseline, the systemic vascular resistance (SVR) (measured in $\text{WU} \times \text{m}^2$) values at rest yielded a standard deviation of 28. Can we conclude from these data that the SVR variance of a population of similar subjects with TGA is not 700? Let $\alpha = .10$.
- 7.7.3 Vital capacity values were recorded for a sample of 10 patients with severe chronic airway obstruction. The variance of the 10 observations was .75. Test the null hypothesis that the population variance is 1.00. Let $\alpha = .05$.
- 7.7.4 Hemoglobin (g percent) values were recorded for a sample of 20 children who were part of a study of acute leukemia. The variance of the observations was 5. Do these data provide sufficient evidence to indicate that the population variance is greater than 4? Let $\alpha = .05$.
- 7.7.5 A sample of 25 administrators of large hospitals participated in a study to investigate the nature and extent of frustration and emotional tension associated with the job. Each participant was given a test designed to measure the extent of emotional tension he or she experienced as a result of the duties and responsibilities associated with the job. The variance of the scores was 30. Can it be concluded from these data that the population variance is greater than 25? Let $\alpha = .05$.
- 7.7.6 In a study in which the subjects were 15 patients suffering from pulmonary sarcoid disease, blood gas determinations were made. The variance of the Pao_2 (mm Hg) values was 450. Test the null hypothesis that the population variance is greater than 250. Let $\alpha = .05$.
- 7.7.7 Analysis of the amniotic fluid from a simple random sample of 15 pregnant women yielded the following measurements on total protein (grams per 100 ml) present:

.69, 1.04, .39, .37, .64, .73, .69, 1.04,
.83, 1.00, .19, .61, .42, .20, .79

Do these data provide sufficient evidence to indicate that the population variance is greater than .05? Let $\alpha = .05$. What assumptions are necessary?

7.8 HYPOTHESIS TESTING: THE RATIO OF TWO POPULATION VARIANCES

As we have seen, the use of the t distribution in constructing confidence intervals and in testing hypotheses for the difference between two population means assumes that the population variances are equal. As a rule, the only hints available about the magnitudes of

the respective variances are the variances computed from samples taken from the populations. We would like to know if the difference that, undoubtedly, will exist between the sample variances is indicative of a real difference in population variances, or if the difference is of such magnitude that it could have come about as a result of chance alone when the population variances are equal.

Two methods of chemical analysis may give the same results on the average. It may be, however, that the results produced by one method are more variable than the results of the other. We would like some method of determining whether this is likely to be true.

Variance Ratio Test Decisions regarding the comparability of two population variances are usually based on the *variance ratio test*, which is a test of the null hypothesis that two population variances are equal. When we test the hypothesis that two population variances are equal, we are, in effect, testing the hypothesis that their ratio is equal to 1.

We learned in the preceding chapter that, when certain assumptions are met, the quantity $(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$ is distributed as F with $n_1 - 1$ numerator degrees of freedom and $n_2 - 1$ denominator degrees of freedom. If we are hypothesizing that $\sigma_1^2 = \sigma_2^2$, we assume that the hypothesis is true, and the two variances cancel out in the above expression leaving s_1^2/s_2^2 , which follows the same F distribution. The ratio s_1^2/s_2^2 will be designated V.R. for variance ratio.

For a two-sided test, we follow the convention of placing the larger sample variance in the numerator and obtaining the critical value of F for $\alpha/2$ and the appropriate degrees of freedom. However, for a one-sided test, which of the two sample variances is to be placed in the numerator is predetermined by the statement of the null hypothesis. For example, for the null hypothesis that σ_1^2/σ_2^2 , the appropriate test statistic is V.R. = s_1^2/s_2^2 . The critical value of F is obtained for α (not $\alpha/2$) and the appropriate degrees of freedom. In like manner, if the null hypothesis is that $\sigma_1^2 \geq \sigma_2^2$, the appropriate test statistic is V.R. = s_2^2/s_1^2 . In all cases, the decision rule is to reject the null hypothesis if the computed V.R. is equal to or greater than the critical value of F .

EXAMPLE 7.8.1

Borden et al. (A-30) compared meniscal repair techniques using cadaveric knee specimens. One of the variables of interest was the load at failure (in newtons) for knees fixed with the FasT-FIX technique (group 1) and the vertical suture method (group 2). Each technique was applied to six specimens. The standard deviation for the FasT-FIX method was 30.62, and the standard deviation for the vertical suture method was 11.37. Can we conclude that, in general, the variance of load at failure is higher for the FasT-FIX technique than the vertical suture method?

Solution:

- 1. Data.** See the statement of the example.
- 2. Assumptions.** Each sample constitutes a simple random sample of a population of similar subjects. The samples are independent. We assume the loads at failure in both populations are approximately normally distributed.

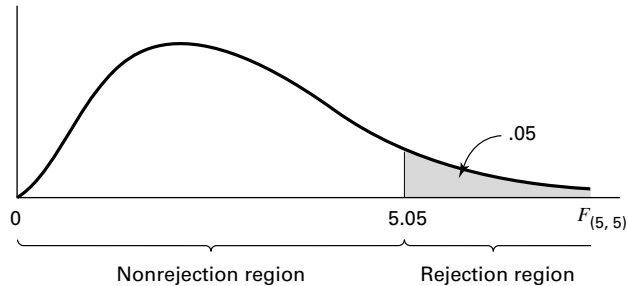


FIGURE 7.8.1 Rejection and nonrejection regions, Example 7.8.1.

3. Hypotheses.

$$H_0: \sigma_1^2 \leq \sigma_2^2$$

$$H_A: \sigma_1^2 > \sigma_2^2$$

4. Test statistic.

$$\text{V.R.} = \frac{s_1^2}{s_2^2} \quad (7.8.1)$$

5. Distribution of test statistic. When the null hypothesis is true, the test statistic is distributed as F with $n_1 - 1$ numerator and $n_2 - 1$ denominator degrees of freedom.

6. Decision rule. Let $\alpha = .05$. The critical value of F , from Appendix Table G, is 5.05. Note that if Table G does not contain an entry for the given numerator degrees of freedom, we use the column closest in value to the given numerator degrees of freedom. Reject H_0 if $\text{V.R.} \geq 5.05$. The rejection and nonrejection regions are shown in Figure 7.8.1.

7. Calculation of test statistic.

$$\text{V.R.} = \frac{(30.62)^2}{(11.37)^2} = 7.25$$

8. Statistical decision. We reject H_0 , since $7.25 > 5.05$; that is, the computed ratio falls in the rejection region.

9. Conclusion. The failure load variability is higher when using the FasT-FIX method than the vertical suture method.

10. p value. Because the computed V.R. of 7.25 is greater than 5.05, the p value for this test is less than 0.05. Excel calculates this p value to be .0243. ■

Several computer programs can be used to test the equality of two variances. Outputs from these programs will differ depending on the test that is used. We saw in Figure 7.3.3,

for example, that the SAS system uses a folded F-test procedure. MINITAB uses two different tests. The first is an F-test under the assumption of normality, and the other is a modified Levene's test (1) that is used when normality cannot be assumed. SPSS uses an unmodified Levene's test (2). Regardless of the options, these tests are generally considered superior to the variance ratio test that is presented in Example 7.8.1. Discussion of the mathematics behind these tests is beyond the scope of this book, but an example is given to illustrate these procedures, since results from these tests are often provided automatically as outputs when a computer program is used to carry out a t -test.

EXAMPLE 7.8.2

Using the data from Example 7.3.2, we are interested in testing whether the assumption of the equality of variances can be assumed prior to performing a t -test. For ease of discussion, the data are reproduced below (Table 7.8.1):

TABLE 7.8.1 Pressures (mm Hg) Under the Pelvis During Static Conditions for Example 7.3.2

Control	131	115	124	131	122	117	88	114	150	169
SCI	60	150	130	180	163	130	121	119	130	148

Partial outputs for MINITAB, SAS, and SPSS are shown in Figure 7.8.2. Regardless of the test or program that is used, we fail to reject the null hypothesis of equal variances ($H_0: \sigma_1^2 = \sigma_2^2$) because all p values > 0.05 . We may now proceed with a t -test under the assumption of equal variances. ■

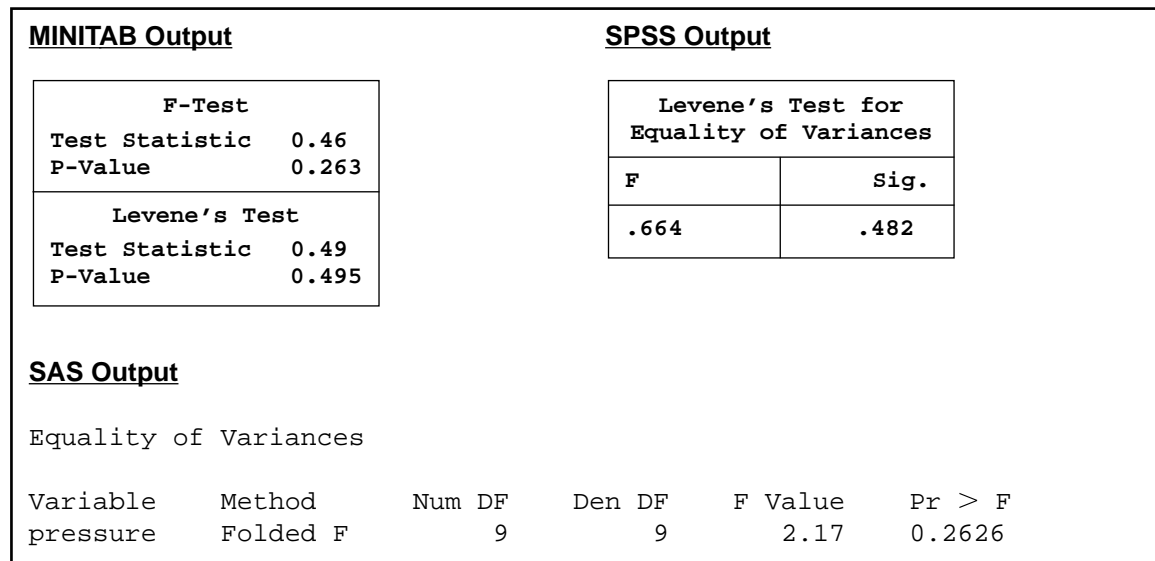


FIGURE 7.8.2 Partial MINITAB, SPSS, and SAS outputs for testing the equality of two population variances.

EXERCISES

In the following exercises perform the ten-step test. For each exercise, as appropriate, explain why you chose a one-sided test or a two-sided test. Discuss how you think researchers or clinicians might use the results of your hypothesis test. What clinical or research decisions or actions do you think would be appropriate in light of the results of your test?

7.8.1 Dora et al. (A-31) investigated spinal canal dimensions in 30 subjects symptomatic with disc herniation selected for a discectomy and 45 asymptomatic individuals. The researchers wanted to know if spinal canal dimensions are a significant risk factor for the development of sciatica. Toward that end, they measured the spinal canal dimension between vertebrae L3 and L4 and obtained a mean of 17.8 mm in the discectomy group with a standard deviation of 3.1. In the control group, the mean was 18.5 mm with a standard deviation of 2.8 mm. Is there sufficient evidence to indicate that in relevant populations the variance for subjects symptomatic with disc herniation is larger than the variance for control subjects? Let $\alpha = .05$.

7.8.2 Nagy et al. (A-32) studied 50 stable patients who were admitted for a gunshot wound that traversed the mediastinum. Of these, eight were deemed to have a mediastinal injury and 42 did not. The standard deviation for the ages of the eight subjects with mediastinal injury was 4.7 years, and the standard deviation of ages for the 42 without injury was 11.6 years. Can we conclude from these data that the variance of age is larger for a population of similar subjects without injury compared to a population with mediastinal injury? Let $\alpha = .05$.

7.8.3 A test designed to measure level of anxiety was administered to a sample of male and a sample of female patients just prior to undergoing the same surgical procedure. The sample sizes and the variances computed from the scores were as follows:

$$\text{Males: } n = 16, s^2 = 150$$

$$\text{Females: } n = 21, s^2 = 275$$

Do these data provide sufficient evidence to indicate that in the represented populations the scores made by females are more variable than those made by males? Let $\alpha = .05$.

7.8.4 In an experiment to assess the effects on rats of exposure to cigarette smoke, 11 animals were exposed and 11 control animals were not exposed to smoke from unfiltered cigarettes. At the end of the experiment, measurements were made of the frequency of the ciliary beat (beats/min at 20°C) in each animal. The variance for the exposed group was 3400 and 1200 for the unexposed group. Do these data indicate that in the populations represented the variances are different? Let $\alpha = .05$.

7.8.5 Two pain-relieving drugs were compared for effectiveness on the basis of length of time elapsing between administration of the drug and cessation of pain. Thirteen patients received drug 1, and 13 received drug 2. The sample variances were $s_1^2 = 64$ and $s_2^2 = 16$. Test the null hypothesis that the two populations variances are equal. Let $\alpha = .05$.

7.8.6 Packed cell volume determinations were made on two groups of children with cyanotic congenital heart disease. The sample sizes and variances were as follows:

Group	n	s^2
1	10	40
2	16	84

Do these data provide sufficient evidence to indicate that the variance of population 2 is larger than the variance of population 1? Let $\alpha = .05$.

- 7.8.7 Independent simple random samples from two strains of mice used in an experiment yielded the following measurements on plasma glucose levels following a traumatic experience:

Strain A: 54, 99, 105, 46, 70, 87, 55, 58, 139, 91

Strain B: 93, 91, 93, 150, 80, 104, 128, 83, 88, 95, 94, 97

Do these data provide sufficient evidence to indicate that the variance is larger in the population of strain A mice than in the population of strain B mice? Let $\alpha = .05$. What assumptions are necessary?

7.9 THE TYPE II ERROR AND THE POWER OF A TEST

In our discussion of hypothesis testing our focus has been on α , the probability of committing a type I error (rejecting a true null hypothesis). We have paid scant attention to β , the probability of committing a type II error (failing to reject a false null hypothesis). There is a reason for this difference in emphasis. For a given test, α is a single number assigned by the investigator in advance of performing the test. It is a measure of the acceptable risk of rejecting a true null hypothesis. On the other hand, β may assume one of many values. Suppose we wish to test the null hypothesis that some population parameter is equal to some specified value. If H_0 is false and we fail to reject it, we commit a type II error. If the hypothesized value of the parameter is not the true value, the value of β (the probability of committing a type II error) depends on several factors: (1) the true value of the parameter of interest, (2) the hypothesized value of the parameter, (3) the value of α , and (4) the sample size, n . For fixed α and n , then, we may, before performing a hypothesis test, compute many values of β by postulating many values for the parameter of interest given that the hypothesized value is false.

For a given hypothesis test it is of interest to know how well the test controls type II errors. If H_0 is in fact false, we would like to know the probability that we will reject it. The *power* of a test, designated $1 - \beta$, provides this desired information. The quantity $1 - \beta$ is the probability that we will reject a false null hypothesis; it may be computed for any alternative value of the parameter about which we are testing a hypothesis. Therefore, $1 - \beta$ is the probability that we will take the correct action when H_0 is false because the true parameter value is equal to the one for which we computed $1 - \beta$. For a given test we may specify any number of possible values of the parameter of interest and for each compute the value of $1 - \beta$. The result is called a *power function*. The graph of a power function, called a *power curve*, is a helpful device for quickly assessing the nature of the power of a given test. The following example illustrates the procedures we use to analyze the power of a test.

EXAMPLE 7.9.1

Suppose we have a variable whose values yield a population standard deviation of 3.6. From the population we select a simple random sample of size $n = 100$. We select a value of $\alpha = .05$ for the following hypotheses:

$$H_0: \mu = 17.5, \quad H_A: \mu \neq 17.5$$

Solution: When we study the power of a test, we locate the rejection and nonrejection regions on the \bar{x} scale rather than the z scale. We find the critical values of \bar{x} for a two-sided test using the following formulas:

$$\bar{x}_U = \mu_0 + z \frac{\sigma}{\sqrt{n}} \quad (7.9.1)$$

and

$$\bar{x}_L = \mu_0 - z \frac{\sigma}{\sqrt{n}} \quad (7.9.2)$$

where \bar{x}_U and \bar{x}_L are the upper and lower critical values, respectively, of \bar{x} ; $+z$ and $-z$ are the critical values of z ; and μ_0 is the hypothesized value of μ . For our example, we have

$$\begin{aligned} \bar{x}_U &= 17.50 + 1.96 \frac{(3.6)}{(10)} = 17.50 + 1.96(.36) \\ &= 17.50 + .7056 = 18.21 \end{aligned}$$

and

$$\bar{x}_L = 17.50 - 1.96(.36) = 17.50 - .7056 = 16.79$$

Suppose that H_0 is false, that is, that μ is not equal to 17.5. In that case, μ is equal to some value other than 17.5. We do not know the actual value of μ . But if H_0 is false, μ is one of the many values that are greater than or smaller than 17.5. Suppose that the true population mean is $\mu_1 = 16.5$. Then the sampling distribution of \bar{x}_1 is also approximately normal, with $\mu_{\bar{x}} = \mu = 16.5$. We call this sampling distribution $f(\bar{x}_1)$, and we call the sampling distribution under the null hypothesis $f(\bar{x}_0)$.

β , the probability of the type II error of failing to reject a false null hypothesis, is the area under the curve of $f(\bar{x}_1)$ that overlaps the nonrejection region specified under H_0 . To determine the value of β , we find the area under $f(\bar{x}_1)$, above the \bar{x} axis, and between $\bar{x} = 16.79$ and $\bar{x} = 18.21$. The value of β is equal to $P(16.79 \leq \bar{x} \leq 18.21)$ when $\mu = 16.5$. This is the same as

$$\begin{aligned} P\left(\frac{16.79 - 16.5}{.36} \leq z \leq \frac{18.21 - 16.5}{.36}\right) &= P\left(\frac{.29}{.36} \leq z \leq \frac{1.71}{.36}\right) \\ &= P(.81 \leq z \leq 4.75) \\ &\approx 1 - .7910 = .2090 \end{aligned}$$

Thus, the probability of taking an appropriate action (that is, rejecting H_0) when the null hypothesis states that $\mu = 17.5$, but in fact $\mu = 16.5$, is

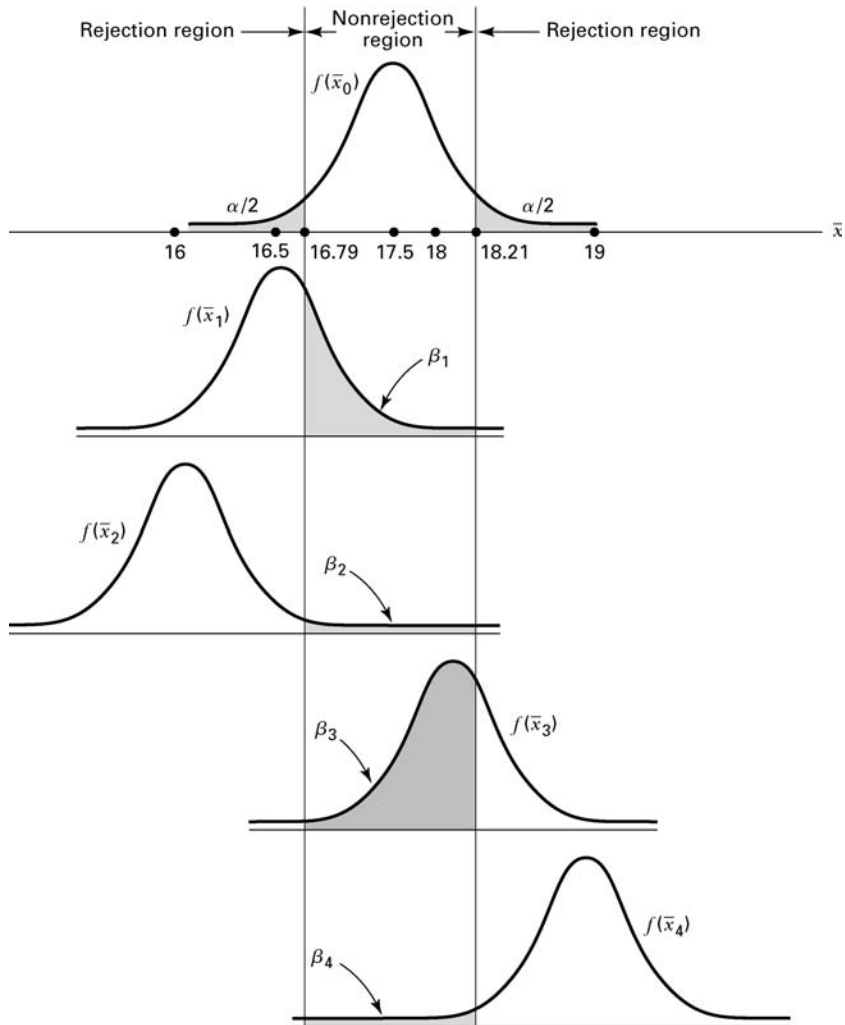


FIGURE 7.9.1 Size of β for selected values for H_1 for Example 7.9.1.

$1 - .2090 = .7910$. As we noted, μ may be one of a large number of possible values when H_0 is false. Figure 7.9.1 shows a graph of several such possibilities. Table 7.9.1 shows the corresponding values of β and $1 - \beta$ (which are approximate), along with the values of β for some additional alternatives.

Note that in Figure 7.9.1 and Table 7.9.1 those values of μ under the alternative hypothesis that are closer to the value of μ specified by H_0 have larger associated β values. For example, when $\mu = 18$ under the alternative hypothesis, $\beta = .7190$; and when $\mu = 19.0$ under H_A , $\beta = .0143$. The power of the test for these two alternatives, then, is $1 - .7190 = .2810$ and $1 - .0143 = .9857$, respectively. We show the power of the test graphically

TABLE 7.9.1 Values of β and $1 - \beta$ for Selected Alternative Values of μ_1 , Example 7.9.1

Possible Values of μ Under H_A When H_0 is False	β	$1 - \beta$
16.0	0.0143	0.9857
16.5	0.2090	0.7910
17.0	0.7190	0.2810
18.0	0.7190	0.2810
18.5	0.2090	0.7910
19.0	0.0143	0.9857

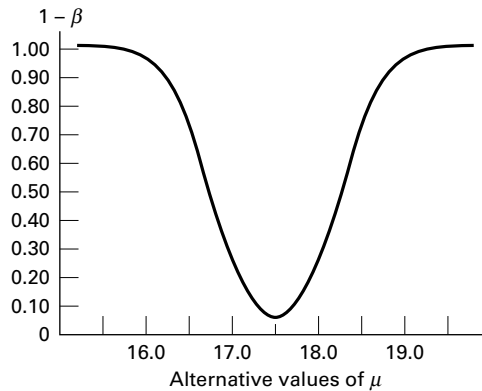


FIGURE 7.9.2 Power curve for Example 7.9.1.

in a power curve, as in Figure 7.9.2. Note that the higher the curve, the greater the power. ■

Although only one value of α is associated with a given hypothesis test, there are many values of β , one for each possible value of μ if μ_0 is not the true value of μ as hypothesized. Unless alternative values of μ are much larger or smaller than μ_0 , β is relatively large compared with α . Typically, we use hypothesis-testing procedures more often in those cases in which, when H_0 is false, the true value of the parameter is fairly close to the hypothesized value. In most cases, β , the computed probability of failing to reject a false null hypothesis, is larger than α , the probability of rejecting a true null hypothesis. These facts are compatible with our statement that a decision based on a rejected null hypothesis is more conclusive than a decision based on a null hypothesis that is not rejected. The probability of being wrong in the latter case is generally larger than the probability of being wrong in the former case.

Figure 7.9.2 shows the V-shaped appearance of a power curve for a two-sided test. In general, a two-sided test that discriminates well between the value of the parameter in H_0 and values in H_1 results in a narrow V-shaped power curve. A wide V-shaped curve

indicates that the test discriminates poorly over a relatively wide interval of alternative values of the parameter.

Power Curves for One-Sided Tests The shape of a power curve for a one-sided test with the rejection region in the upper tail is an elongated S. If the rejection region of a one-sided test is located in the lower tail of the distribution, the power curve takes the form of a reverse elongated S. The following example shows the nature of the power curve for a one-sided test.

EXAMPLE 7.9.2

The mean time laboratory employees now take to do a certain task on a machine is 65 seconds, with a standard deviation of 15 seconds. The times are approximately normally distributed. The manufacturers of a new machine claim that their machine will reduce the mean time required to perform the task. The quality-control supervisor designs a test to determine whether or not she should believe the claim of the makers of the new machine. She chooses a significance level of $\alpha = 0.01$ and randomly selects 20 employees to perform the task on the new machine. The hypotheses are

$$H_0: \mu \geq 65, \quad H_A: \mu < 65$$

The quality-control supervisor also wishes to construct a power curve for the test.

Solution: The quality-control supervisor computes, for example, the following value of $1 - \beta$ for the alternative $\mu = 55$. The critical value of $1 - \beta$ for the test is

$$65 - 2.33 \left(\frac{15}{\sqrt{20}} \right) = 57$$

We find β as follows:

$$\begin{aligned} \beta &= P(\bar{x} > 57 \mid \mu = 55) = P\left(z > \frac{57 - 55}{15/\sqrt{20}}\right) = P(z > .60) \\ &= 1 - .7257 = .2743 \end{aligned}$$

Consequently, $1 - \beta = 1 - .2743 = .7257$. Figure 7.9.3 shows the calculation of β . Similar calculations for other alternative values of μ

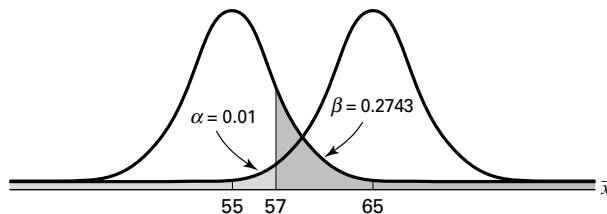


FIGURE 7.9.3 β calculated for $\mu = 55$.

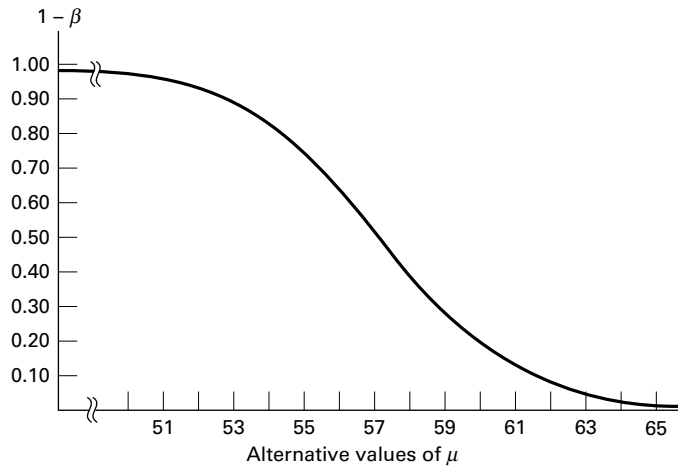


FIGURE 7.9.4 Power curve for Example 7.9.2.

also yield values of $1 - \beta$. When plotted against the values of μ , these give the power curve shown in Figure 7.9.4. ■

Operating Characteristic Curves Another way of evaluating a test is to look at its *operating characteristic (OC) curve*. To construct an OC curve, we plot values of β , rather than $1 - \beta$, along the vertical axis. Thus, an OC curve is the complement of the corresponding power curve.

EXERCISES

Construct and graph the power function for each of the following situations.

7.9.1 $H_0: \mu \leq 516, \quad H_A: \mu > 516, \quad n = 16, \quad \sigma = 32, \quad \alpha = 0.05.$

7.9.2 $H_0: \mu = 3, \quad H_A: \mu \neq 3, \quad n = 100, \quad \sigma = 1, \quad \alpha = 0.05.$

7.9.3 $H_0: \mu \leq 4.25, \quad H_A: \mu > 4.25, \quad n = 81, \quad \sigma = 1.8, \quad \alpha = 0.01.$

7.10 DETERMINING SAMPLE SIZE TO CONTROL TYPE II ERRORS

You learned in Chapter 6 how to find the sample sizes needed to construct confidence intervals for population means and proportions for specified levels of confidence. You learned in Chapter 7 that confidence intervals may be used to test hypotheses. The method of determining sample size presented in Chapter 6 takes into account the probability of a type I error, but not a type II error since the level of confidence is determined by the confidence coefficient, $1 - \alpha$.

In many statistical inference procedures, the investigator wishes to consider the type II error as well as the type I error when determining the sample size. To illustrate the procedure, we refer again to Example 7.9.2.

EXAMPLE 7.10.1

In Example 7.9.2, the hypotheses are

$$H_0: \mu \geq 65, \quad H_A: \mu < 65$$

The population standard deviation is 15, and the probability of a type I error is set at .01. Suppose that we want the probability of failing to reject H_0 (β) to be .05 if H_0 is false because the true mean is 55 rather than the hypothesized 65. How large a sample do we need in order to realize, simultaneously, the desired levels of α and β ?

Solution: For $\alpha = .01$ and $n = 20$, β is equal to .2743. The critical value is 57. Under the new conditions, the critical value is unknown. Let us call this new critical value C . Let μ_0 be the hypothesized mean and μ_1 the mean under the alternative hypothesis. We can transform each of the relevant sampling distributions of \bar{x} , the one with a mean of μ_0 and the one with a mean of μ_1 to a z distribution. Therefore, we can convert C to a z value on the horizontal scale of each of the two standard normal distributions. When we transform the sampling distribution of \bar{x} that has a mean of μ_0 to the standard normal distribution, we call the z that results z_0 . When we transform the sampling distribution \bar{x} that has a mean of μ_1 to the standard normal distribution, we call the z that results z_1 . Figure 7.10.1 represents the situation described so far.

We can express the critical value C as a function of z_0 and μ_0 and also as a function of z_1 and μ_1 . This gives the following equations:

$$C = \mu_0 - z_0 \frac{\sigma}{\sqrt{n}} \quad (7.10.1)$$

$$C = \mu_1 + z_1 \frac{\sigma}{\sqrt{n}} \quad (7.10.2)$$

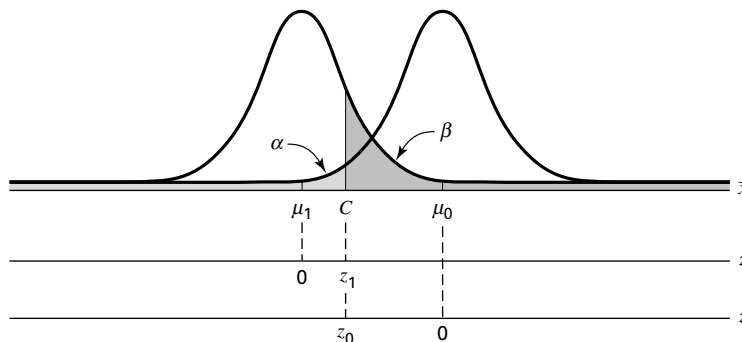


FIGURE 7.10.1 Graphic representation of relationships in determination of sample size to control both type I and type II errors.

We set the right-hand sides of these equations equal to each other and solve for n , to obtain

$$n = \left[\frac{(z_0 + z_1)\sigma}{(\mu_0 - \mu_1)} \right]^2 \quad (7.10.3)$$

To find n for our illustrative example, we substitute appropriate quantities into Equation 7.10.3. We have $\mu_0 = 65$, $\mu_1 = 55$, and $\sigma = 15$. From Appendix Table D, the value of z that has .01 of the area to its left is -2.33 . The value of z that has .05 of the area to its right is 1.645. Both z_0 and z_1 are taken as positive. We determine whether C lies above or below either μ_0 or μ_1 when we substitute into Equations 7.10.1 and 7.10.2. Thus, we compute

$$n = \left[\frac{(2.33 + 1.645)(15)}{(65 - 55)} \right]^2 = 35.55$$

We would need a sample of size 36 to achieve the desired levels of α and β when we choose $\mu_1 = 55$ as the alternative value of μ .

We now compute C , the critical value for the test, and state an appropriate decision rule. To find C , we may substitute known numerical values into either Equation 7.10.1 or Equation 7.10.2. For illustrative purposes, we solve both equations for C . First we have

$$C = 65 - 2.33 \left(\frac{15}{\sqrt{36}} \right) = 59.175$$

From Equation 7.10.2, we have

$$C = 55 - 1.645 \left(\frac{15}{\sqrt{36}} \right) = 59.1125$$

The difference between the two results is due to rounding error.

The decision rule, when we use the first value of C , is as follows:

Select a sample of size 36 and compute \bar{x} , if $\bar{x} \leq 59.175$, reject H_0 . If $\bar{x} > 59.175$, do not reject H_0 .

We have limited our discussion of the type II error and the power of a test to the case involving a population mean. The concepts extend to cases involving other parameters. ■

EXERCISES

- 7.10.1** Given $H_0: \mu = 516$, $H_A: \mu > 516$, $n = 16$, $\sigma = 32$, $\alpha = .05$. Let $\beta = .10$ and $\mu_1 = 520$, and find n and C . State the appropriate decision rule.
- 7.10.2** Given $H_0: \mu \leq 4.500$, $H_A: \mu > 4.500$, $n = 16$, $\sigma = .020$, $\alpha = .01$. Let $\beta = .05$ and $\mu_1 = 4.52$, and find n and C . State the appropriate decision rule.
- 7.10.3** Given $H_0: \mu \leq 4.25$, $H_A: \mu > 4.25$, $n = 81$, $\sigma = 1.8$, $\alpha = .01$. Let $\beta = .03$ and $\mu_1 = 5.00$, and find n and C . State the appropriate decision rule.

7.11 SUMMARY

In this chapter the general concepts of hypothesis testing are discussed. A general procedure for carrying out a hypothesis test consisting of the following ten steps is suggested.

1. Description of data.
2. Statement of necessary assumptions.
3. Statement of null and alternative hypotheses.
4. Specification of the test statistic.
5. Specification of the distribution of the test statistic.
6. Statement of the decision rule.
7. Calculation of test statistic from sample data.
8. The statistical decision based on sample results.
9. Conclusion.
10. Determination of p value.

A number of specific hypothesis tests are described in detail and illustrated with appropriate examples. These include tests concerning population means, the difference between two population means, paired comparisons, population proportions, the difference between two population proportions, a population variance, and the ratio of two population variances. In addition we discuss the power of a test and the determination of sample size for controlling both type I and type II errors.

SUMMARY OF FORMULAS FOR CHAPTER 7

Formula Number	Name	Formula
7.1.1, 7.1.2, 7.2.1	z -transformation (using either μ or μ_0)	$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$
7.2.2	t -transformation	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
7.2.3	Test statistic when sampling from a population that is not normally distributed	$z = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$
7.3.1	Test statistic when sampling from normally distributed populations: population variances known	$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

7.3.2	Test statistic when sampling from normally distributed populations: population variances unknown and equal	$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$, where $s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$
7.3.3, 7.3.4	Test statistic when sampling from normally distributed populations: population variances unknown and unequal	$t' = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$, where $t'_{1-(\alpha/2)} = \frac{w_1 t_1 + w_2 t_2}{w_1 + w_2}$
7.3.5	Sampling from populations that are not normally distributed	$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
7.4.1	Test statistic for paired differences when the population variance is unknown	$t = \frac{\bar{d} - \mu_{d_0}}{s_{\bar{d}}}$
7.4.2	Test statistic for paired differences when the population variance is known	$z = \frac{\bar{d} - \mu_d}{\sigma_d / \sqrt{n}}$
7.5.1	Test statistic for a single population proportion	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$
7.6.1, 7.6.2	Test statistic for the difference between two population proportions	$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)_0}{\hat{\sigma}_{\hat{p}_1 - \hat{p}_2}}$, where $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$, and $\hat{\sigma}_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\bar{p}(1 - \bar{p})}{n_1} + \frac{\bar{p}(1 - \bar{p})}{n_2}}$
7.7.1	Test statistic for a single population variance	$\chi^2 = \frac{(n - 1)s^2}{\sigma^2}$
7.8.1	Variance ratio	$\text{V.R.} = \frac{s_1^2}{s_2^2}$

(Continued)

7.9.1, 7.9.2	Upper and lower critical values for \bar{x}	$\bar{x}_U = \mu_0 + z \frac{\sigma}{\sqrt{n}}$ $\bar{x}_L = \mu_0 - z \frac{\sigma}{\sqrt{n}}$
7.10.1, 7.10.2	Critical value for determining sample size to control type II errors	$C = \mu_0 - z_0 \frac{\sigma}{\sqrt{n}} = \mu_1 + z_1 \frac{\sigma}{\sqrt{n}}$
7.10.3	Sample size to control type II errors	$n = \left[\frac{(z_0 + z_1)\sigma}{(\mu_0 - \mu_1)} \right]^2$
Symbol Key	<ul style="list-style-type: none"> • α = type I error rate • C = critical value • χ^2 = chi-square distribution • \bar{d} = average difference • μ = mean of population • μ_0 = hypothesized mean • n = sample size • p = proportion for population • \bar{p} = average proportion • $q = (1 - p)$ • \hat{p} = estimated proportion for sample • σ^2 = population variance • σ = population standard deviation • $\sigma_{\bar{d}}$ = standard error of difference • $\sigma_{\bar{x}}$ = standard error • s = standard deviation of sample • $s_{\bar{d}}$ = standard deviation of the difference • s_p = pooled standard deviation • t = Student's t-transformation • t' = Cochran's correction to t • \bar{x} = mean of sample • \bar{x}_L = lower limit of critical value for \bar{x} • \bar{x}_U = upper limit of critical value for \bar{x} • z = standard normal transformation 	

REVIEW QUESTIONS AND EXERCISES

1. What is the purpose of hypothesis testing?
2. What is a hypothesis?
3. List and explain each step in the ten-step hypothesis testing procedure.

4. Define:
 - (a) Type I error
 - (b) Type II error
 - (c) The power of a test
 - (d) Power function
 - (e) Power curve
 - (f) Operating characteristic curve
5. Explain the difference between the power curves for one-sided tests and two-sided tests.
6. Explain how one decides what statement goes into the null hypothesis and what statement goes into the alternative hypothesis.
7. What are the assumptions underlying the use of the t statistic in testing hypotheses about a single mean? The difference between two means?
8. When may the z statistic be used in testing hypotheses about
 - (a) a single population mean?
 - (b) the difference between two population means?
 - (c) a single population proportion?
 - (d) the difference between two population proportions?
9. In testing a hypothesis about the difference between two population means, what is the rationale behind pooling the sample variances?
10. Explain the rationale behind the use of the paired comparisons test.
11. Give an example from your field of interest where a paired comparisons test would be appropriate. Use real or realistic data and perform an appropriate hypothesis test.
12. Give an example from your field of interest where it would be appropriate to test a hypothesis about the difference between two population means. Use real or realistic data and carry out the ten-step hypothesis testing procedure.
13. Do Exercise 12 for a single population mean.
14. Do Exercise 12 for a single population proportion.
15. Do Exercise 12 for the difference between two population proportions.
16. Do Exercise 12 for a population variance.
17. Do Exercise 12 for the ratio of two population variances.
18. Ochsenkühn et al. (A-33) studied birth as a result of in vitro fertilization (IVF) and birth from spontaneous conception. In the sample, there were 163 singleton births resulting from IVF with a mean birth weight of 3071 g and sample standard deviation of 761 g. Among the 321 singleton births resulting from spontaneous conception, the mean birth weight was 3172 g with a standard deviation of 702 g. Determine if these data provide sufficient evidence for us to conclude that the mean birth weight in grams of singleton births resulting from IVF is lower, in general, than the mean birth weight of singleton births resulting from spontaneous conception. Let $\alpha = .10$.
19. William Tindall (A-34) performed a retrospective study of the records of patients receiving care for hypercholesterolemia. The following table gives measurements of total cholesterol for patients before and 6 weeks after taking a statin drug. Is there sufficient evidence at the $\alpha = .01$ level of significance for us to conclude that the drug would result in reduction in total cholesterol in a population of similar hypercholesterolemia patients?

Id. No.	Before	After	Id. No.	Before	After	Id. No.	Before	After
1	195	125	37	221	191	73	205	151
2	208	164	38	245	164	74	298	163
3	254	152	39	250	162	75	305	171
4	226	144	40	266	180	76	262	129
5	290	212	41	240	161	77	320	191
6	239	171	42	218	168	78	271	167
7	216	164	43	278	200	79	195	158
8	286	200	44	185	139	80	345	192
9	243	190	45	280	207	81	223	117
10	217	130	46	278	200	82	220	114
11	245	170	47	223	134	83	279	181
12	257	182	48	205	133	84	252	167
13	199	153	49	285	161	85	246	158
14	277	204	50	314	203	86	304	190
15	249	174	51	235	152	87	292	177
16	197	160	52	248	198	88	276	148
17	279	205	53	291	193	89	250	169
18	226	159	54	231	158	90	236	185
19	262	170	55	208	148	91	256	172
20	231	180	56	263	203	92	269	188
21	234	161	57	205	156	93	235	172
22	170	139	58	230	161	94	184	151
23	242	159	59	250	150	95	253	156
24	186	114	60	209	181	96	352	219
25	223	134	61	269	186	97	266	186
26	220	166	62	261	164	98	321	206
27	277	170	63	255	164	99	233	173
28	235	136	64	275	195	100	224	109
29	216	134	65	239	169	101	274	109
30	197	138	66	298	177	102	222	136
31	253	181	67	265	217	103	194	131
32	209	147	68	220	191	104	293	228
33	245	164	69	196	129	105	262	211
34	217	159	70	177	142	106	306	192
35	187	139	71	211	138	107	239	174
36	265	171	72	244	166			

Source: Data provided courtesy of William Tindall, Ph.D. and the Wright State University Consulting Center.

20. The objective of a study by van Vollenhoven et al. (A-35) was to examine the effectiveness of Etanercept alone and Etanercept in combination with methotrexate in the treatment of rheumatoid arthritis. They performed a retrospective study using data from the STURE database, which collects efficacy and safety data for all patients starting biological treatments at the major hospitals in Stockholm, Sweden. The researchers identified 40 subjects who were prescribed Etanercept only and 57 who were given Etanercept with methotrexate. One of the outcome measures was the number of swollen joints. The following table gives the mean number of swollen joints in the two groups as well as the standard error of the mean. Is there sufficient evidence at the

$\alpha = .05$ level of significance for us to conclude that there is a difference in mean swollen joint counts in the relevant populations?

Treatment	Mean	Standard Error of Mean
Etanercept	5.56	0.84
Etanercept plus methotrexate	4.40	0.57

21. Miyazaki et al. (A-36) examined the recurrence-free rates of stripping with varicectomy and stripping with sclerotherapy for the treatment of primary varicose veins. The varicectomy group consisted of 122 limbs for which the procedure was done, and the sclerotherapy group consisted of 98 limbs for which that procedure was done. After 3 years, 115 limbs of the varicectomy group and 87 limbs of the sclerotherapy group were recurrence-free. Is this sufficient evidence for us to conclude there is no difference, in general, in the recurrence-free rate between the two procedures for treating varicose veins? Let $\alpha = .05$.
22. Recall the study, reported in Exercise 7.8.1, in which Dora et al. (A-37) investigated spinal canal dimensions in 30 subjects symptomatic with disc herniation selected for a discectomy and 45 asymptomatic individuals (control group). One of the areas of interest was determining if there is a difference between the two groups in the spinal canal cross-sectional area (cm²) between vertebrae L5/S1. The data in the following table are simulated to be consistent with the results reported in the paper. Do these simulated data provide evidence for us to conclude that a difference in the spinal canal cross-sectional area exists between a population of subjects with disc herniations and a population of those who do not have disc herniations? Let $\alpha = .05$.

Herniated Disc Group					Control Group					
2.62	2.57	1.98	3.21	3.59	3.72	4.30	2.87	3.87	2.73	5.28
1.60	1.80	3.91	2.56	1.53	1.33	2.36	3.67	1.64	3.54	3.63
2.39	2.67	3.53	2.26	2.82	4.26	3.08	3.32	4.00	2.76	3.58
2.05	1.19	3.01	2.39	3.61	3.11	3.94	4.39	3.73	2.22	2.73
2.09	3.79	2.45	2.55	2.10	5.02	3.62	3.02	3.15	3.57	2.37
2.28	2.33	2.81	3.70	2.61	5.42	3.35	2.62	3.72	4.37	5.28
					4.97	2.58	2.25	3.12	3.43	
					3.95	2.98	4.11	3.08	2.22	

Source: Simulated data.

23. Iannello et al. (A-38) investigated differences between triglyceride levels in healthy obese (control) subjects and obese subjects with chronic active B or C hepatitis. Triglyceride levels of 208 obese controls had a mean value of 1.81 with a standard error of the mean of .07 mmol/L. The 19 obese hepatitis subjects had a mean of .71 with a standard error of the mean of .05. Is this sufficient evidence for us to conclude that, in general, a difference exists in average triglyceride levels between obese healthy subjects and obese subjects with hepatitis B or C? Let $\alpha = .01$.
24. Kindergarten students were the participants in a study conducted by Susan Bazyk et al. (A-39). The researchers studied the fine motor skills of 37 children receiving occupational therapy. They used an index of fine motor skills that measured hand use, eye-hand coordination, and manual dexterity

before and after 7 months of occupational therapy. Higher values indicate stronger fine motor skills. The scores appear in the following table.

Subject	Pre	Post	Subject	Pre	Post
1	91	94	20	76	112
2	61	94	21	79	91
3	85	103	22	97	100
4	88	112	23	109	112
5	94	91	24	70	70
6	112	112	25	58	76
7	109	112	26	97	97
8	79	97	27	112	112
9	109	100	28	97	112
10	115	106	29	112	106
11	46	46	30	85	112
12	45	41	31	112	112
13	106	112	32	103	106
14	112	112	33	100	100
15	91	94	34	88	88
16	115	112	35	109	112
17	59	94	36	85	112
18	85	109	37	88	97
19	112	112			

Source: Data provided courtesy of Susan Bazyk, M.H.S.

Can one conclude on the basis of these data that after 7 months, the fine motor skills in a population of similar subjects would be stronger? Let $\alpha = .05$. Determine the p value.

25. A survey of 90 recently delivered women on the rolls of a county welfare department revealed that 27 had a history of intrapartum or postpartum infection. Test the null hypothesis that the population proportion with a history of intrapartum or postpartum infection is less than or equal to .25. Let $\alpha = .05$. Determine the p value.
26. In a sample of 150 hospital emergency admissions with a certain diagnosis, 128 listed vomiting as a presenting symptom. Do these data provide sufficient evidence to indicate, at the .01 level of significance, that the population proportion is less than .92? Determine the p value.
27. A research team measured tidal volume in 15 experimental animals. The mean and standard deviation were 45 and 5 cc, respectively. Do these data provide sufficient evidence to indicate that the population mean is greater than 40 cc? Let $\alpha = .05$.
28. A sample of eight patients admitted to a hospital with a diagnosis of biliary cirrhosis had a mean IgM level of 160.55 units per milliliter. The sample standard deviation was 50. Do these data provide sufficient evidence to indicate that the population mean is greater than 150? Let $\alpha = .05$. Determine the p value.
29. Some researchers have observed a greater airway resistance in smokers than in nonsmokers. Suppose a study, conducted to compare the percent of tracheobronchial retention of particles in smoking-discordant monozygotic twins, yielded the following results:

Percent Retention		Percent Retention	
Smoking Twin	Nonsmoking Twin	Smoking Twin	Nonsmoking Twin
60.6	47.5	57.2	54.3
12.0	13.3	62.7	13.9
56.0	33.0	28.7	8.9
75.2	55.2	66.0	46.1
12.5	21.9	25.2	29.8
29.7	27.9	40.1	36.2

Do these data support the hypothesis that tracheobronchial clearance is slower in smokers? Let $\alpha = .05$. Determine the p value for this test.

30. Circulating levels of estrone were measured in a sample of 25 postmenopausal women following estrogen treatment. The sample mean and standard deviation were 73 and 16, respectively. At the .05 significance level can one conclude on the basis of these data that the population mean is higher than 70?
31. Systemic vascular resistance determinations were made on a sample of 16 patients with chronic, congestive heart failure while receiving a particular treatment. The sample mean and standard deviation were 1600 and 700, respectively. At the .05 level of significance do these data provide sufficient evidence to indicate that the population mean is less than 2000?
32. The mean length at birth of 14 male infants was 53 cm with a standard deviation of 9 cm. Can one conclude on the basis of these data that the population mean is not 50 cm? Let the probability of committing a type I error be .10.

For each of the studies described in Exercises 33 through 38, answer as many of the following questions as possible: (a) What is the variable of interest? (b) Is the parameter of interest a mean, the difference between two means (independent samples), a mean difference (paired data), a proportion, or the difference between two proportions (independent samples)? (c) What is the sampled population? (d) What is the target population? (e) What are the null and alternative hypotheses? (f) Is the alternative one-sided (left tail), one-sided (right tail), or two-sided? (g) What type I and type II errors are possible? (h) Do you think the null hypothesis was rejected? Explain why or why not.

33. During a one-year period, Hong et al. (A-40) studied all patients who presented to the surgical service with possible appendicitis. One hundred eighty-two patients with possible appendicitis were randomized to either clinical assessment (CA) alone or clinical evaluation and abdominal/pelvic CT. A true-positive case resulted in a laparotomy that revealed a lesion requiring operation. A true-negative case did not require an operation at one-week follow-up evaluation. At the close of the study, they found no significant difference in the hospital length of stay for the two treatment groups.
34. Recall the study reported in Exercise 7.8.2 in which Nagy et al. (A-32) studied 50 stable patients admitted for a gunshot wound that traversed the mediastinum. They found that eight of the subjects had a mediastinal injury, while 42 did not have such an injury. They performed a student's t test to determine if there was a difference in mean age (years) between the two groups. The reported p value was .59.
35. Dykstra et al. (A-41) studied 15 female patients with urinary frequency with or without incontinence. The women were treated with botulinum toxin type B (BTX-B). A t test of the

- pre/post-difference in frequency indicated that these 15 patients experienced an average of 5.27 fewer frequency episodes per day after treatment with BTX-B. The p value for the test was less than 0.001.
36. Recall the study reported in Exercise 6.10.2 in which Horesh et al. (A-42) investigated suicidal behavior among adolescents. In addition to impulsivity, the researchers studied hopelessness among the 33 subjects in the suicidal group and the 32 subjects in the nonsuicidal group. The means for the two groups on the Beck Hopelessness Scale were 11.6 and 5.2, respectively, and the t value for the test was 5.13.
 37. Mauksch et al. (A-43) surveyed 500 consecutive patients (ages 18 to 64 years) in a primary care clinic serving only uninsured, low-income patients. They used self-report questions about why patients were coming to the clinic, and other tools to classify subjects as either having or not having major mental illness. Compared with patients without current major mental illness, patients with a current major mental illness reported significantly ($p < .001$) more concerns, chronic illnesses, stressors, forms of maltreatment, and physical symptoms.
 38. A study by Hosking et al. (A-44) was designed to compare the effects of alendronate and risedronate on bone mineral density (BMD). One of the outcome measures was the percent increase in BMD at 12 months. Alendronate produced a significantly higher percent change (4.8 percent) in BMD than risedronate (2.8 percent) with a p value $< .001$.
 39. For each of the following situations, identify the type I and type II errors and the correct actions.
 - (a) H_0 : A new treatment is not more effective than the traditional one.
 - (1) Adopt the new treatment when the new one is more effective.
 - (2) Continue with the traditional treatment when the new one is more effective.
 - (3) Continue with the traditional treatment when the new one is not more effective.
 - (4) Adopt the new treatment when the new one is not more effective.
 - (b) H_0 : A new physical therapy procedure is satisfactory.
 - (1) Employ a new procedure when it is unsatisfactory.
 - (2) Do not employ a new procedure when it is unsatisfactory.
 - (3) Do not employ a new procedure when it is satisfactory.
 - (4) Employ a new procedure when it is satisfactory.
 - (c) H_0 : A production run of a drug is of satisfactory quality.
 - (1) Reject a run of satisfactory quality.
 - (2) Accept a run of satisfactory quality.
 - (3) Reject a run of unsatisfactory quality.
 - (4) Accept a run of unsatisfactory quality.

For each of the studies described in Exercises 40 through 55, do the following:

- (a) Perform a statistical analysis of the data (including hypothesis testing and confidence interval construction) that you think would yield useful information for the researchers.
 - (b) State all assumptions that are necessary to validate your analysis.
 - (c) Find p values for all computed test statistics.
 - (d) Describe the population(s) about which you think inferences based on your analysis would be applicable.
40. A study by Bell (A-45) investigated the hypothesis that alteration of the vitamin D–endocrine system in blacks results from reduction in serum 25-hydroxyvitamin D and that the alteration is reversed by oral treatment with 25-hydroxyvitamin D_3 . The eight subjects (three men and five women) were studied while on no treatment (control) and after having been given 25-hydroxyvitamin D_3 for 7 days

(25-OHD₃). The following are the urinary calcium (mg/d) determinations for the eight subjects under the two conditions.

Subject	Control	25-OHD ₃
A	66	98
B	115	142
C	54	78
D	88	101
E	82	134
F	115	158
G	176	219
H	46	60

Source: Data provided courtesy of Dr. Norman H. Bell.

41. Montner et al. (A-46) conducted studies to test the effects of glycerol-enhanced hyperhydration (GEH) on endurance in cycling performance. The 11 subjects, ages 22–40 years, regularly cycled at least 75 miles per week. The following are the pre-exercise urine output volumes (ml) following ingestion of glycerol and water:

Subject #	Experimental, ml (Glycerol)	Control, ml (Placebo)
1	1410	2375
2	610	1610
3	1170	1608
4	1140	1490
5	515	1475
6	580	1445
7	430	885
8	1140	1187
9	720	1445
10	275	890
11	875	1785

Source: Data provided courtesy of Dr. Paul Montner.

42. D'Alessandro et al. (A-47) wished to know if preexisting airway hyperresponsiveness (HR) predisposes subjects to a more severe outcome following exposure to chlorine. Subjects were healthy volunteers between the ages of 18 and 50 years who were classified as with and without HR. The following are the FEV₁ and specific airway resistance (Sraw) measurements taken on the subjects before and after exposure to appropriately diluted chlorine gas:

Hyperreactive Subjects				
Subject	Pre-Exposure		Post-Exposure	
	FEV ₁	Sraw	FEV ₁	Sraw
1	3.0	5.80	1.8	21.4
2	4.1	9.56	3.7	12.5
3	3.4	7.84	3.0	14.3
4	3.3	6.41	3.0	10.9
5	3.3	9.12	3.0	17.1

Normal Subjects				
Subject	Pre-Exposure		Post-Exposure	
	FEV ₁	Sraw	FEV ₁	Sraw
1	4.3	5.52	4.2	8.70
2	3.9	6.43	3.7	6.94
3	3.6	5.67	3.3	10.00
4	3.6	3.77	3.5	4.54
5	5.1	5.53	4.9	7.37

Source: Data provided courtesy of Dr. Paul Blanc.

43. Noting the paucity of information on the effect of estrogen on platelet membrane fatty acid composition, Ranganath et al. (A-48) conducted a study to examine the possibility that changes may be present in postmenopausal women and that these may be reversible with estrogen treatment. The 31 women recruited for the study had not menstruated for at least 3 months or had symptoms of the menopause. No woman was on any form of hormone replacement therapy (HRT) at the time she was recruited. The following are the platelet membrane linoleic acid values before and after a period of HRT:

Subject	Before	After	Subject	Before	After	Subject	Before	After
1	6.06	5.34	12	7.65	5.55	23	5.04	4.74
2	6.68	6.11	13	4.57	4.25	24	7.89	7.48
3	5.22	5.79	14	5.97	5.66	25	7.98	6.24
4	5.79	5.97	15	6.07	5.66	26	6.35	5.66
5	6.26	5.93	16	6.32	5.97	27	4.85	4.26
6	6.41	6.73	17	6.12	6.52	28	6.94	5.15
7	4.23	4.39	18	6.05	5.70	29	6.54	5.30
8	4.61	4.20	19	6.31	3.58	30	4.83	5.58
9	6.79	5.97	20	4.44	4.52	31	4.71	4.10
10	6.16	6.00	21	5.51	4.93			
11	6.41	5.35	22	8.48	8.80			

Source: Data provided courtesy of Dr. L. Ranganath.

44. The purpose of a study by Goran et al. (A-49) was to examine the accuracy of some widely used body-composition techniques for children through the use of the dual-energy X-ray absorptiometry (DXA) technique. Subjects were children between the ages of 4 and 10 years. The following are fat mass measurements taken on the children by three techniques—DXA, skinfold thickness (ST), and bioelectrical resistance (BR):

DXA	ST	BR	Sex
			(1 = Male, 0 = Female)
3.6483	4.5525	4.2636	1
2.9174	2.8234	6.0888	0
7.5302	3.8888	5.1175	0
6.2417	5.4915	8.0412	0
10.5891	10.4554	14.1576	0
9.5756	11.1779	12.4004	0

(Continued)

DXA	ST	BR	Sex
			(1 = Male, 0 = Female)
2.4424	3.5168	3.7389	1
3.5639	5.8266	4.3359	1
1.2270	2.2467	2.7144	1
2.2632	2.4499	2.4912	1
2.4607	3.1578	1.2400	1
4.0867	5.5272	6.8943	0
4.1850	4.0018	3.0936	1
2.7739	5.1745	*	1
4.4748	3.6897	4.2761	0
4.2329	4.6807	5.2242	0
2.9496	4.4187	4.9795	0
2.9027	3.8341	4.9630	0
5.4831	4.8781	5.4468	0
3.6152	4.1334	4.1018	1
5.3343	3.6211	4.3097	0
3.2341	2.0924	2.5711	1
5.4779	5.3890	5.8418	0
4.6087	4.1792	3.9818	0
2.8191	2.1216	1.5406	1
4.1659	4.5373	5.1724	1
3.7384	2.5182	4.6520	1
4.8984	4.8076	6.5432	1
3.9136	3.0082	3.2363	1
12.1196	13.9266	16.3243	1
15.4519	15.9078	18.0300	0
20.0434	19.5560	21.7365	0
9.5300	8.5864	4.7322	1
2.7244	2.8653	2.7251	1
3.8981	5.1352	5.2420	0
4.9271	8.0535	6.0338	0
3.5753	4.6209	5.6038	1
6.7783	6.5755	6.6942	1
3.2663	4.0034	3.2876	0
1.5457	2.4742	3.6931	0
2.1423	2.1845	2.4433	1
4.1894	3.0594	3.0203	1
1.9863	2.5045	3.2229	1
3.3916	3.1226	3.3839	1
2.3143	2.7677	3.7693	1
1.9062	3.1355	12.4938	1
3.7744	4.0693	5.9229	1
2.3502	2.7872	4.3192	0
4.6797	4.4804	6.2469	0
4.7260	5.4851	7.2809	0
4.2749	4.4954	6.6952	0
2.6462	3.2102	3.8791	0

(Continued)

DXA	ST	BR	Sex (1 = Male, 0 = Female)
2.7043	3.0178	5.6841	0
4.6148	4.0118	5.1399	0
3.0896	3.2852	4.4280	0
5.0533	5.6011	4.3556	0
6.8461	7.4328	8.6565	1
11.0554	13.0693	11.7701	1
4.4630	4.0056	7.0398	0
2.4846	3.5805	3.6149	0
7.4703	5.5016	9.5402	0
8.5020	6.3584	9.6492	0
6.6542	6.8948	9.3396	1
4.3528	4.1296	6.9323	0
3.6312	3.8990	4.2405	1
4.5863	5.1113	4.0359	1
2.2948	2.6349	3.8080	1
3.6204	3.7307	4.1255	1
2.3042	3.5027	3.4347	1
4.3425	3.7523	4.3001	1
4.0726	3.0877	5.2256	0
1.7928	2.8417	3.8734	1
4.1428	3.6814	2.9502	1
5.5146	5.2222	6.0072	0
3.2124	2.7632	3.4809	1
5.1687	5.0174	3.7219	1
3.9615	4.5117	2.7698	1
3.6698	4.9751	1.8274	1
4.3493	7.3525	4.8862	0
2.9417	3.6390	3.4951	1
5.0380	4.9351	5.6038	0
7.9095	9.5907	8.5024	0
1.7822	3.0487	3.0028	1
3.4623	3.3281	2.8628	1
11.4204	14.9164	10.7378	1
1.2216	2.2942	2.6263	1
2.9375	3.3124	3.3728	1
4.6931	5.4706	5.1432	0
8.1227	7.7552	7.7401	0
10.0142	8.9838	11.2360	0
2.5598	2.8520	4.5943	0
3.7669	3.7342	4.7384	0
4.2059	2.6356	4.0405	0
6.7340	6.6878	8.1053	0
3.5071	3.4947	4.4126	1
2.2483	2.8100	3.6705	0
7.1891	5.4414	6.6332	0
6.4390	3.9532	5.1693	0

* Missing data.

Source: Data provided courtesy of
Dr. Michael I. Goran.

45. Hartard et al. (A-50) conducted a study to determine whether a certain training regimen can counteract bone density loss in women with postmenopausal osteopenia. The following are strength measurements for five muscle groups taken on 15 subjects before (B) and after (A) 6 months of training:

Subject	Leg Press		Hip Flexor		Hip Extensor	
	(B)	(A)	(B)	(A)	(B)	(A)
1	100	180	8	15	10	20
2	155	195	10	20	12	25
3	115	150	8	13	12	19
4	130	170	10	14	12	20
5	120	150	7	12	12	15
6	60	140	5	12	8	16
7	60	100	4	6	6	9
8	140	215	12	18	14	24
9	110	150	10	13	12	19
10	95	120	6	8	8	14
11	110	130	10	12	10	14
12	150	220	10	13	15	29
13	120	140	9	20	14	25
14	100	150	9	10	15	29
15	110	130	6	9	8	12

Subject	Arm Abductor		Arm Adductor	
	(B)	(A)	(B)	(A)
1	10	12	12	19
2	7	20	10	20
3	8	14	8	14
4	8	15	6	16
5	8	13	9	13
6	5	13	6	13
7	4	8	4	8
8	12	15	14	19
9	10	14	8	14
10	6	9	6	10
11	8	11	8	12
12	8	14	13	15
13	8	19	11	18
14	4	7	10	22
15	4	8	8	12

Source: Data provided courtesy of Dr. Manfred Hartard.

46. Vitacca et al. (A-51) conducted a study to determine whether the supine position or sitting position worsens static, forced expiratory flows and measurements of lung mechanics. Subjects were aged

persons living in a nursing home who were clinically stable and without clinical evidence of cardiorespiratory diseases. Among the data collected were the following FEV₁ percent values for subjects in sitting and supine postures:

Sitting	Supine	Sitting	Supine
64	56	103	94
44	37	109	92
44	39	-99	-99
40	43	169	165
32	32	73	66
70	61	95	94
82	58	-99	-99
74	48	73	58
91	63		

Source: Data provided courtesy of Dr. M. Vitacca.

47. The purpose of an investigation by Young et al. (A-52) was to examine the efficacy and safety of a particular suburethral sling. Subjects were women experiencing stress incontinence who also met other criteria. Among the data collected were the following pre- and postoperative cystometric capacity (ml) values:

Pre	Post	Pre	Post	Pre	Post	Pre	Post
350	321	340	320	595	557	475	344
700	483	310	336	315	221	427	277
356	336	361	333	363	291	405	514
362	447	339	280	305	310	312	402
361	214	527	492	200	220	385	282
304	285	245	330	270	315	274	317
675	480	313	310	300	230	340	323
367	330	241	230	792	575	524	383
387	325	313	298	275	140	301	279
535	325	323	349	307	192	411	383
328	250	438	345	312	217	250	285
557	410	497	300	375	462	600	618
569	603	302	335	440	414	393	355
260	178	471	630	300	250	232	252
320	362	540	400	379	335	332	331
405	235	275	278	682	339	451	400
351	310	557	381				

Source: Data provided courtesy of Dr. Stephen B. Young.

48. Diamond et al. (A-53) wished to know if cognitive screening should be used to help select appropriate candidates for comprehensive inpatient rehabilitation. They studied a sample of geriatric rehabilitation patients using standardized measurement strategies. Among the data collected were the following admission and discharge scores made by the subjects on the Mini Mental State Examination (MMSE):

Admission	Discharge	Admission	Discharge
9	10	24	26
11	11	24	30
14	19	24	28
15	15	25	26
16	17	25	22
16	15	26	26
16	17	26	28
16	17	26	26
17	14	27	28
17	18	27	28
17	21	27	27
18	21	27	27
18	21	27	27
19	21	28	28
19	25	28	29
19	21	28	29
19	22	28	29
19	19	29	28
20	22	29	28
21	23	29	30
22	22	29	30
22	19	29	30
22	26	29	30
23	21	29	30
24	21	30	30
24	20		

Source: Data provided courtesy of Dr. Stephen N. Macciocchi.

49. In a study to explore the possibility of hormonal alteration in asthma, Weinstein et al. (A-54) collected data on 22 postmenopausal women with asthma and 22 age-matched, postmenopausal, women without asthma. The following are the dehydroepiandrosterone sulfate (DHEAS) values collected by the investigators:

Without Asthma	With Asthma	Without Asthma	With Asthma
20.59	87.50	15.90	166.02
37.81	111.52	49.77	129.01
76.95	143.75	25.86	31.02
77.54	25.16	55.27	47.66
19.30	68.16	33.83	171.88
35.00	136.13	56.45	241.88
146.09	89.26	19.91	235.16
166.02	96.88	24.92	25.16
96.58	144.34	76.37	78.71
24.57	97.46	6.64	111.52
53.52	82.81	115.04	54.69

Source: Data provided courtesy of Dr. Robert E. Weinstein.

50. The motivation for a study by Gruber et al. (A-55) was a desire to find a potentially useful serum marker in rheumatoid arthritis (RA) that reflects underlying pathogenic mechanisms. They measured, among other variables, the circulating levels of gelatinase B in the serum and synovial fluid (SF) of patients with RA and of control subjects. The results were as follows:

Serum		Synovial Fluid		Serum		Synovial Fluid	
RA	Control	RA	Control	RA	Control	RA	Control
26.8	23.4	71.8	3.0	36.7			
19.1	30.5	29.4	4.0	57.2			
249.6	10.3	185.0	3.9	71.3			
53.6	8.0	114.0	6.9	25.2			
66.1	7.3	69.6	9.6	46.7			
52.6	10.1	52.3	22.1	30.9			
14.5	17.3	113.1	13.4	27.5			
22.7	24.4	104.7	13.3	17.2			
43.5	19.7	60.7		10.3			
25.4	8.4	116.8		7.5			
29.8	20.4	84.9		31.6			
27.6	16.3	215.4		30.0			
106.1	16.5	33.6		42.0			
76.5	22.2	158.3		20.3			

Source: Data provided courtesy of Dr. Darius Sorbi.

51. Benini et al. (A-56) conducted a study to evaluate the severity of esophageal acidification in achalasia following successful dilatation of the cardias and to determine which factors are associated with pathological esophageal acidification in such patients. Twenty-two subjects, of whom seven were males; ranged in ages from 28 to 78 years. On the basis of established criteria they were classified as refluxers or nonrefluxers. The following are the acid clearance values (min/reflux) for the 22 subjects:

Refluxers	Nonrefluxers
8.9	2.3
30.0	0.2
23.0	0.9
6.2	8.3
11.5	0.0
	0.9
	0.4
	2.0
	0.7
	3.6
	0.5
	1.4
	0.2
	0.7
	17.9
	2.1
	0.0

Source: Data provided courtesy of Dr. Luigi Benini.

52. The objective of a study by Baker et al. (A-57) was to determine whether medical deformation alters in vitro effects of plasma from patients with preeclampsia on endothelial cell function to produce a paradigm similar to the in vivo disease state. Subjects were 24 nulliparous pregnant women before delivery, of whom 12 had preeclampsia and 12 were normal pregnant patients. Among the data collected were the following gestational ages (weeks) at delivery:

Preeclampsia	Normal Pregnant
38	40
32	41
42	38
30	40
38	40
35	39
32	39
38	41
39	41
29	40
29	40
32	40

Source: Data provided courtesy of Dr. James M. Roberts.

53. Zisselman et al. (A-58) conducted a study to assess benzodiazepine use and the treatment of depression before admission to an inpatient geriatric psychiatry unit in a sample of elderly patients. Among the data collected were the following behavior disorder scores on 27 patients treated with benzodiazepines (W) and 28 who were not (WO).

W		WO	
.00	1.00	.00	.00
.00	1.00	.00	10.00
.00	.00	.00	.00
.00	.00	.00	18.00
.00	10.00	.00	.00
.00	2.00	.00	2.00
.00		.00	5.00
.00		.00	
.00		4.00	
.00		1.00	
4.00		2.00	
3.00		.00	
2.00		6.00	
.00		.00	
10.00		.00	
2.00		1.00	
.00		2.00	
9.00		1.00	
.00		22.00	
1.00		.00	
16.00		.00	

Source: Data provided courtesy of Dr. Yochi Shmuelly.

54. The objective of a study by Reinecke et al. (A-59) was to investigate the functional activity and expression of the sarcolemmal $\text{Na}^+/\text{Ca}^{2+}$ exchange in the failing human heart. The researchers obtained left ventricular samples from failing human hearts of 11 male patients (mean age 51 years) undergoing cardiac transplantation. Nonfailing control hearts were obtained from organ donors (four females, two males, mean age 41 years) whose hearts could not be transplanted for noncardiac reasons. The following are the $\text{Na}^+/\text{Ca}^{2+}$ exchanger activity measurements for the patients with end-stage heart failure (CHF) and nonfailing controls (NF).

NF	CHF
0.075	0.221
0.073	0.231
0.167	0.145
0.085	0.112
0.110	0.170
0.083	0.207
	0.112
	0.291
	0.164
	0.195
	0.185

Source: Data provided courtesy of Dr. Hans Reinecke.

55. Reichman et al. (A-60) conducted a study with the purpose of demonstrating that negative symptoms are prominent in patients with Alzheimer's disease and are distinct from depression. The following are scores made on the Scale for the Assessment of Negative Symptoms in Alzheimer's Disease by patients with Alzheimer's disease (PT) and normal elderly, cognitively intact, comparison subjects (C).

PT	C
19	6
5	5
36	10
22	1
1	1
18	0
24	5
17	5
7	4
19	6
5	6
2	7
14	5
9	3
34	5
13	12

(Continued)

PT	C
0	0
21	5
30	1
43	2
19	3
31	19
21	3
41	5
	24
	3

Source: Data provided courtesy
of Dr. Andrew C. Coyne.

Exercises for Use with Large Data Sets Available on the Following Website:

www.wiley.com/college/daniel

1. Refer to the creatine phosphokinase data on 1005 subjects (PCKDATA). Researchers would like to know if psychologically stressful situations cause an increase in serum creatine phosphokinase (CPK) levels among apparently healthy individuals. To help the researchers reach a decision, select a simple random sample from this population, perform an appropriate analysis of the sample data, and give a narrative report of your findings and conclusions. Compare your results with those of your classmates.
2. Refer to the prothrombin time data on 1000 infants (PROTHROM). Select a simple random sample of size 16 from each of these populations and conduct an appropriate hypothesis test to determine whether one should conclude that the two populations differ with respect to mean prothrombin time. Let $\alpha = .05$. Compare your results with those of your classmates. What assumptions are necessary for the validity of the test?
3. Refer to the head circumference data of 1000 matched subjects (HEADCIRC). Select a simple random sample of size 20 from the population and perform an appropriate hypothesis test to determine if one can conclude that subjects with the sex chromosome abnormality tend to have smaller heads than normal subjects. Let $\alpha = .05$. Construct a 95 percent confidence interval for the population mean difference. What assumptions are necessary? Compare your results with those of your classmates.
4. Refer to the hemoglobin data on 500 children with iron deficiency anemia and 500 apparently healthy children (HEMOGLOB). Select a simple random sample of size 16 from population A and an independent simple random sample of size 16 from population B. Does your sample data provide sufficient evidence to indicate that the two populations differ with respect to mean Hb value? Let $\alpha = .05$. What assumptions are necessary for your procedure to be valid? Compare your results with those of your classmates.
5. Refer to the manual dexterity scores of 500 children with learning disabilities and 500 children with no known learning disabilities (MANDEXT). Select a simple random sample of size 10 from population A and an independent simple random sample of size 15 from population B. Do your samples provide sufficient evidence for you to conclude that learning-disabled children, on the average, have lower manual dexterity scores than children without a learning disability? Let $\alpha = .05$. What assumptions are necessary in order for your procedure to be valid? Compare your results with those of your classmates.

REFERENCES

Methodology References

1. M. B. BROWN and A. B. FORSYTHE, "Robust Tests for the Equality of Variances," *Journal of the American Statistical Association*, 69 (1974), 364–367.
2. H. LEVENE, "Robust Tests for Equality of Variances," in I. Olkin, ed., *Contributions to Probability and Statistics*, Stanford University Press, Palo Alto, CA, 1960, 278–292.

Applications References

- A-1. NORIMASA NAKAMURA, SHUJI HORIBE, YUKYOSHI TORITSUKA, TOMOKI MITSUOKA, HIDEKI YOSHIKAWA, and KONSEI SHINO, "Acute Grade III Medial Collateral Ligament Injury of the Knee Associated with Anterior Cruciate Ligament Tear," *American Journal of Sports Medicine*, 31 (2003), 261–267.
- A-2. DIANE KLINGLER, ROBBYA GREEN-WEIR, DAVID NERENZ, SUZANNE HAVSTAD, HOWARD S. ROSMAN, LEONARD CETNER, SAMIR SHAH, FRANCES WIMBUSH, and STEVEN BORZAK, "Perceptions of Chest Pain Differ by Race," *American Heart Journal*, 144 (2002), 51–59.
- A-3. A. ESCOBAR, J. M. QUINTANA, A. BILBAO, J. AZKÁRATE, and J. I. GÜENAGA, "Validation of the Spanish Version of the WOMAC Questionnaire for Patients with Hip or Knee Osteoarthritis," *Clinical Rheumatology*, 21 (2002), 466–471.
- A-4. PHAMORNSAK THIENPRASIDDHI, VIVIENNE C. GREENSTEIN, CANDICE S. CHEN, JEFFREY M. LIEBMANN, ROBERT RITCH, and DONALD C. HOOD, "Multifocal Visual Evoked Potential Responses in Glaucoma Patients with Unilateral Hemifield Defects," *American Journal of Ophthalmology*, 136 (2003), 34–40.
- A-5. P. F. LUGLIÈ, GUGLIELMO CAMPUS, C. DEIOLA, M. G. MELA, and D. GALLISAI, "Oral Condition, Chemistry of Saliva, and Salivary Levels of *Streptococcus Mutans* in Thalassemic Patients," *Clinical Oral Investigations*, 6 (2002), 223–226.
- A-6. ERIC W. TAM, ARTHUR F. MAK, WAI NGA LAM, JOHN H. EVANS, and YORK Y. CHOW, "Pelvic Movement and Interface Pressure Distribution During Manual Wheelchair Propulsion," *Archives of Physical Medicine and Rehabilitation*, 84 (2003), 1466–1472.
- A-7. JOHN DERNELLIS and MARIA PANARETOU, "Effects of Thyroid Replacement Therapy on Arterial Blood Pressure in Patients with Hypertension and Hypothyroidism," *American Heart Journal*, 143 (2002), 718–724.
- A-8. S. SAIRAM, B. A. BAETHGE, and T. MCNEARNEY, "Analysis of Risk Factors and Comorbid Diseases in the Development of Thrombosis in Patients with Anticardiolipin Antibodies," *Clinical Rheumatology*, 22 (2003), 24–29.
- A-9. MICHEL DABONNEVILLE, PAUL BERTHON, PHILIPPE VASLIN, and NICOLE FELLMANN, "The 5 M in Running Field Test: Test and Retest Reliability on Trained Men and Women," *European Journal of Applied Physiology*, 88 (2003), 353–360.
- A-10. B. M. INGLE and R. EASTELL, "Site-Specific Bone Measurements in Patients with Ankle Fracture," *Osteoporosis International*, 13 (2002), 342–347.
- A-11. A. HOEKEMA, B. HOVINGA, B. STEGENGA, and L. G. M. DE BONT, "Craniofacial Morphology and Obstructive Steep Apnoea: A Cephalometric Analysis," *Journal of Oral Rehabilitation*, 30 (2003), 690–696.
- A-12. GIAN PAOLO ROSSI, STEFANO TADDEI, AGOSTINO VIRDIS, MARTINA CAVALLIN, LORENZO GHIADONI, STEFANIA FAVILLA, DANIELE VERSARI, ISABELLA SUDANO, ACHILLE C. PESSINA, and ANTONIO SALVETTI, "The T⁷⁸⁶C and Glu298Asp Polymorphisms of the Endothelial Nitric Oxide Gene Affect the Forearm Blood Flow Responses of Caucasian Hypertensive Patients," *Journal of the American College of Cardiology*, 41 (2003), 938–945.
- A-13. JOSÉ GARÇÃO and JOSÉ CABRITA, "Evaluation of a Pharmaceutical Care Program for Hypertensive Patients in Rural Portugal," *Journal of the American Pharmaceutical Association*, 42 (2002), 858–864.
- A-14. JOHN M. MORTON, STEVEN P. BOWERS, TANANCHAI A. LUCKTONG, SAMER MATTAR, W. ALAN BRADSHAW, KEVIN E. BEHRNS, MARK J. KORUDA, CHARLES A. HERBST, WILLIAM MCCARTNEY, RAGHUVVEER K. HALKAR, C. DANIEL SMITH, and TIMOTHY M. FARRELL, "Gallbladder Function Before and After Fundoplication," *Journal of Gastrointestinal Surgery*, 6 (2002), 806–811.
- A-15. ELLEN DAVIS JONES, "Reminiscence Therapy for Older Women with Depression: Effects of Nursing Intervention Classification in Assisted-Living Long-Term Care," *Journal of Gerontological Nursing*, 29 (2003), 26–33.
- A-16. JOHNNY BENEY, E. BETH DEVINE, VALBY CHOW, ROBERT J. IGNOFFO, LISA MITSUNAGA, MINA SHAHKARAMI, ALEX McMILLAN, and LISA A. BERO, "Effect of Telephone Follow-Up on the Physical Well-Being Dimension of Quality of Life in Patients with Cancer," *Pharmacotherapy*, 22 (2002), 1301–1311.

- A-17. JOHN S. MORLEY, JOHN BRIDSON, TIM P. NASH, JOHN B. MILES, SARAH WHITE, and MATTHEW K. MAKIN, "Low-Dose Methadone Has an Analgesic Effect in Neuropathic Pain: A Double-Blind Randomized Controlled Crossover Trial," *Palliative Medicine*, 17 (2003), 576–587.
- A-18. W. K. WOO and K. E. MCKENNA, "Combination TL01 Ultraviolet B Phototherapy and Topical Calcipotriol for Psoriasis: A Prospective Randomized Placebo-Controlled Clinical Trial," *British Journal of Dermatology*, 149 (2003), 146–150.
- A-19. SIMONA PORCELLINI, GIULIANA VALLANTI, SILVIA NOZZA, GUIDO POLI, ADRAINO LAZZARIN, GUISEPPE TABUSSI, and ANTONIO GRASSIA, "Improved Thymopoietic Potential in Aviremic HIV-Infected Individuals with HAART by Intermittent IL-2 Administration," *AIDS*, 17 (2003), 1621–1630.
- A-20. LYNNE E. WAGENKNECHT, CARL D. LANGEFELD, ANN L. SCHERZINGER, JILL M. NORRIS, STEVEN M. HAFFNER, MOHAMMED F. SAAD, and RICHARD N. BERGMAN, "Insulin Sensitivity, Insulin Secretion, and Abdominal Fat," *Diabetes*, 52 (2003), 2490–2496.
- A-21. YVES JACQUEMYN, FATIMA AHANKOUR, and GUY MARTENS, "Flemish Obstetricians' Personal Preference Regarding Mode of Delivery and Attitude Towards Caesarean Section on Demand," *European Journal of Obstetrics and Gynecology and Reproductive Biology*, 111 (2003), 164–166.
- A-22. LI HUI and A. COLIN BELL, "Overweight and Obesity in Children from Shenzhen, People's Republic of China," *Health and Place*, 9 (2003), 371–376.
- A-23. ANNE E. BECKER, REBECCA A. BURWELL, KESAIA NAVARA, and STEPHEN E. GILMAN, "Binge Eating and Binge Eating Disorder in a Small-Scale, Indigenous Society: The View from Fiji," *International Journal of Eating Disorders*, 34 (2003), 423–431.
- A-24. JACQUELINE A. NOONAN, RENKE RAAIJMAKERS, and BRYAN D. HALL, "Adult Height in Noonan Syndrome," *American Journal of Medical Genetics*, 123A (2003), 68–71.
- A-25. SAI YIN HO, TAI HING LAM, RICHARD FIELDING, and EDWARD DENIS JANUS, "Smoking and Perceived Health in Hong Kong Chinese," *Social Science and Medicine*, 57 (2003), 1761–1770.
- A-26. MARKUS A. LANDOLT, MARGARETE VOLLRATH, KARIN RIBI, HANSPETER E. GNEHM, and FELIX H. SENNHAUSER, "Incidence and Associations of Parental and Child Posttraumatic Stress Symptoms in Pediatric Patients," *Journal of Child Psychology and Psychiatry*, 44 (2003), 1199–1207.
- A-27. MORELL MICHAEL AVRAM, DANIEL BLAUSTEIN, PAUL A. FEIN, NAVEEN GOEL, JYOTIPRAKAS CHATTOPADYAY, and NEAL MITTMAN, "Hemoglobin Predicts Long-Term Survival in Dialysis Patients: A 15-Year Single-Center Longitudinal Study and a Correlation Trend Between Prealbumin and Hemoglobin," *Kidney International*, 64 (2003), 6–11.
- A-28. JUDY P. WILKINS, OSCAR E. SUMAN, DEB A. BENJAMIN, and DAVID N. HEMDON, "Comparison of Self-Reported and Monitored Compliance of Daily Injection of Human Growth Hormone in Burned Children," *Burns*, 29 (2003), 697–701.
- A-29. B. ROBINSON, C. T. HEISE, J. W. MOORE, J. ANELLA, M. SOKOLOSKI, and E. ESHAGHPUR, "Afterload Reduction Therapy in Patients Following Intraatrial Baffle Operation for Transposition of the Great Arteries," *Pediatric Cardiology*, 23 (2003), 618–623.
- A-30. PETER BORDEN, JOHN NYLAND, DAVID N. M. CABORN, and DAVID PIENOWSKI, "Biomechanical Comparison of the FasT-FIX Meniscal Repair Suture System with Vertical Mattress and Meniscus Arrows," *American Journal of Sports Medicine*, 31 (2003), 374–378.
- A-31. CLAUDIO DORA, BEAT WÄLCHLI, ACHIM ELFERING, IMRE GAL, DOMINIK WEISHAUP, and NORBERT BOOS, "The Significance of Spinal Canal Dimensions in Discriminating Symptomatic from Asymptomatic Disc Herniations," *European Spine Journal*, 11 (2002), 575–581.
- A-32. KIMBERLY K. NAGY, ROXANNE R. ROBERTS, ROBERT F. SMITH, KIMBERLY T. JOSEPH, GARY, C. AN, FARAN BOKHARI, and JOHN BARRETT, "Trans-Mediastinal Gunshot Wounds: Are 'Stable' Patients Really Stable?" *World Journal of Surgery*, 26 (2002), 1247–1250.
- A-33. ROBERT OCHSENKÜHN, THOMAS STROWITZKI, MONIKA GURTNER, ALEXANDER STRAUSS, ANDREAS SCHULZE, HERMANN HEPP, and PETER HILLEMANN, "Pregnancy Complications, Obstetric Risks, and Neonatal Outcome in Singleton and Twin Pregnancies after GIFT and IVF," *Archives of Gynecology and Obstetrics*, 268 (2003), 256–261.
- A-34. William Tindall and the Wright State University Consulting Center (2003).
- A-35. RONALD F. VAN VOLLENHOVEN, SOFIA ERNESTAM, ANDERS HARJU, JOHAN BRATT, and LARS KLAARESKOG, "Etanercept Versus Etanercept Plus Methotrexate: A Registry-Based Study Suggesting that the Combination Is Clinically More Efficacious," *Arthritis Research & Therapy*, 5 (2003), R347–R351.
- A-36. KICKO MIYAZAKI, TOSHIYA NISHIBE, FUMIHIRO SATA, TAKAHIRO IMAI, FAVIO A. KUDO, JORGE FLORES, YASUMASA J. MIYAZAKI, and KEISHU YASUDA, "Stripping Operation with Sclerotherapy for Primary Varicose Veins Due to Greater Saphenous Vein Reflux: Three Year Results," *World Journal of Surgery*, 27 (2003), 551–553.

- A-37. CLAUDIO DORA, BEAT WÄLCHLI, ACHIM ELFERING, IMRE GAL, DOMINIK WEISHAUP, and NORBERT BOOS, "The Significance of Spinal Canal Dimensions in Discriminating Symptomatic from Asymptomatic Disc Herniations," *European Spine Journal*, 11 (2002), 575–581.
- A-38. SILVIA IANNELO, ANTONINA CAVALERI, PAOLINA MILAZZO, SANTI CANTARELLA, and FRANCESCO BELFIORE, "Low Fasting Serum Triglyceride Level as a Precocious Marker of Autoimmune Disorders," *Medscape General Medicine*, 5 (3) (2003).
- A-39. SUSAN BAZYK, PAULA MICHAUD, GLENN GOODMAN, and PAULA PAPP, "Integrating Occupational Therapy Services in a Kindergarten Curriculum: A Look at the Outcomes." Submitted to *American Journal of Occupational Therapy*, June 30, 2004.
- A-40. JOHN J. HONG, STEPHEN M. COHN, A. PETER EKEH, MARTIN NEWMAN, MOISES SALAMA, and SUZANNE D. LEBLANG, "A Prospective Randomized Study of Clinical Assessment Versus Computed Tomography for the Diagnosis of Acute Appendicitis," *Surgical Infections*, 4 (2003), 231–239.
- A-41. DENNIS DYKSTRA, AL ENRIQUEZ, and MICHAEL VALLEY, "Treatment of Overactive Bladder with Botulinum Toxin Type B: A Pilot Study," *International Urogynecological Journal*, 14 (2003), 424–426.
- A-42. NETTA HORESH, ISRAEL ORBACH, DORON GOTHELF, MEIR EFRATI, and ALAN APTER, "Comparison of the Suicidal Behavior of Adolescent Inpatients with Borderline Personality Disorder and Major Depression," *Journal of Nervous and Mental Disease*, 191 (2003), 582–588.
- A-43. LARRY B. MAUKSCH, WAYNE J. KATON, JOAN RUSSO, SUZANNE M. TUCKER, EDWARD WALKER, and JANET CAMERON, "The Content of a Low-Income, Uninsured Primary Care Population: Including the Patient Agenda," *Journal of the American Board of Family Practice*, 16 (2003), 278–289.
- A-44. DAVID HOSKING, SILVANO ADAMI, DIETER FELSEBERG, JORGE CANNATA ANDIA, MATTI VÄLIMÄKI, LAURENT BENBAMOU, JEAN-YVES REGINSTER, CAROL YACIK, ANDREA RYBAK-FEGLIN, RICHARD A. PETRUSCHKE, LUNA ZARU, and ARTHUR C. SANTORA, "Comparison of Change in Bone Resorption and Bone Mineral Density with Once-Weekly Alendronate and Daily Risedronate: A Randomised, Placebo-Controlled Study," *Current Medical Research Opinion*, 19 (2003), 383–394.
- A-45. NORMAN H. BELL, "25-Hydroxyvitamin D3[r] Reverses Alteration of the Vitamin D–Endocrine System in Blacks," *The American Journal of Medicine*, 99 (December 1995), 597–599.
- A-46. P. MONTNER, D. M. STARK, M. L. RIEDESEL, G. MURATA, R. ROBERGS, M. TIMMS, and T. W. CHICK, "Pre-exercise Glycerol Hydration Improves Cycling Endurance Time," *International Journal of Sports Medicine*, 17 (1996), 27–33.
- A-47. ALESSANDRA D'ALESSANDRO, WARE KUSHNER, HOFER WONG, HOMER A. BOUSHEY, and PAUL D. BLANC, "Exaggerated Responses to Chlorine Inhalation Among Persons with Nonspecific Airway Hyperreactivity," *Chest*, 109 (1996), 331–337.
- A-48. L. R. RANGANATH, J. A. CHRISTOFIDES, J. W. WRIGHT, and V. MARKS, "Effect of Hormone Replacement Therapy on Platelet Membrane Fatty Acid Composition," *Journal of Endocrinology*, 148 (1996), 207–212.
- A-49. MICHAEL I. GORAN, PATRICIA DRISCOLL, RACHEL JOHNSON, TIM R. NAGY, and GARY HUNTER, "Cross-Calibration of Body-Composition Techniques Against Dual-Energy X-Ray Absorptiometry in Young Children," *American Journal of Clinical Nutrition*, 63 (1996), 299–305.
- A-50. MANFRED HARTARD, PAUL HABER, DIANA ILIEVA, ELISABETH PREISINGER, GEORG SEIDL, and JOHANNES HUBER, "Systematic Strength Training as a Model of Therapeutic Intervention," *American Journal of Physical Medicine & Rehabilitation*, 75 (1996), 21–28.
- A-51. M. VITACCA, E. CLINI, W. SPASSINI, L. SCAGLIA, P. NEGRINI, and A. QUADRI, "Does the Supine Position Worsen Respiratory Function in Elderly Subjects?" *Gerontology*, 42 (1996), 46–53.
- A-52. STEPHEN B. YOUNG, PETER L. ROSENBLATT, DIANE M. PINGETON, ALLISON E. HOWARD, and STEPHEN P. BAKER, "The Mersilene Mesh Suburethral Sling: A Clinical and Urodynamic Evaluation," *American Journal of Obstetrics and Gynecology*, 173 (1995), 1719–1726.
- A-53. PAUL T. DIAMOND, GERALD FELSENTAL, STEPHEN N. MACCIOCCHI, DONNA H. BUTLER, and DENISE LALLY-CASSADY, "Effect of Cognitive Impairment on Rehabilitation Outcome," *American Journal of Physical Medicine & Rehabilitation*, 75 (1996), 40–43.
- A-54. ROBERT E. WEINSTEIN, CATHERINE A. LOBOCKI, SHARON GRAVETT, HILLMAN HUM, RONALD NEGRICH, JONATHAN HERBST, DANA GREENBERG, and DAVID R. PIEPER, "Decreased Adrenal Sex Steroid Levels in the Absence of Glucocorticoid Suppression in Postmenopausal Asthmatic Women," *Journal of Allergy and Clinical Immunology*, 97 (1996), 1–8.
- A-55. BARRY L. GRUBER, DARIUS SORBI, DEBORAH L. FRENCH, MARY J. MARCHESE, GERARD J. NUOVO, RICHARD R. KEW, and LEONARD A. ARBEIT, "Markedly Elevated Serum MMP-9 (Gelatinase B) Levels in Rheumatoid Arthritis: A Potentially Useful Laboratory Marker," *Clinical Immunology and Immunopathology*, 78 (1996), 161–171.

- A-56. LUIGI BENINI, CARLO SEMBENINI, GRETA CASTELLANI, ENRICO BARDELLI, MARIA TERESA BRENTEGANI, PIERGIORGIO GIORGETTI, and ITALO VANTINI, "Pathological Esophageal Acidification and Pneumatic Dilatation in Achalasic Patients: Too Much or Not Enough?" *Digestive Diseases and Sciences*, 41 (1996), 365–371.
- A-57. PHILIP N. BAKER, CHRISTOPHER P. STRANKO, SANDRA T. DAVIDGE, PAUL S. DAVIES, and JAMES M. ROBERTS, "Mechanical Stress Eliminates the Effects of Plasma from Patients with Preeclampsia on Endothelial Cells," *American Journal of Obstetrics and Gynecology*, 174 (1996), 703–706.
- A-58. MARC H. ZISSELMAN, BARRY W. ROVNER, and YOCHI SHMUELY, "Benzodiazepine Use in the Elderly Prior to Psychiatric Hospitalization," *Psychosomatics*, 37 (1996), 38–42.
- A-59. HANS REINECKE, ROLAND STRUDER, ROLAND VETTER, JURGEN HOLTZ, and HELMUT DREXLER, "Cardiac $\text{Na}^+/\text{Ca}^{2+}$ Exchange Activity in Patients with End-Stage Heart Failure," *Cardiovascular Research*, 31 (1996), 48–54.
- A-60. WILLIAM E. REICHMAN, ANDREW C. COYNE, SATISH AMIRNENI, BRUNO MOLINO, and SANDRA EGAN, "Negative Symptoms in Alzheimer's Disease," *American Journal of Psychiatry*, 153 (1996), 424–426.

ANALYSIS OF VARIANCE

CHAPTER OVERVIEW

This chapter introduces the first in a series of chapters devoted to linear models. The topic of this chapter, analysis of variance, provides a methodology for partitioning the total variance computed from a data set into components, each of which represents the amount of the total variance that can be attributed to a specific source of variation. The results of this partitioning can then be used to estimate and test hypotheses about population variances and means. In this chapter we focus our attention on hypothesis testing of means. Specifically, we discuss the testing of differences among means when there is interest in more than two populations or two or more variables. The techniques discussed in this chapter are widely used in the health sciences.

TOPICS

- 8.1 INTRODUCTION
- 8.2 THE COMPLETELY RANDOMIZED DESIGN
- 8.3 THE RANDOMIZED COMPLETE BLOCK DESIGN
- 8.4 THE REPEATED MEASURES DESIGN
- 8.5 THE FACTORIAL EXPERIMENT
- 8.6 SUMMARY

LEARNING OUTCOMES

After studying this chapter, the student will

1. understand the basic statistical concepts related to linear models.
2. understand how the total variation in a data set can be partitioned into different components.
3. be able to compare the means of more than two samples simultaneously.
4. understand multiple comparison tests and when their use is appropriate.
5. understand commonly used experimental designs.

8.1 INTRODUCTION

In the preceding chapters the basic concepts of statistics have been examined, and they provide a foundation for this and the next several chapters. In this chapter and the three that follow, we provide an overview of two of the most commonly employed analytical tools used by applied statisticians, analysis of variance and linear regression. The conceptual foundations of these analytical tools are statistical models that provide useful representations of the relationships among several variables simultaneously.

Linear Models A statistical model is a mathematical representation of the relationships among variables. More specifically for the purposes of this book, a statistical model is most often used to describe how random variables are related to one another in a context in which the value of one *outcome variable*, often referred to with the letter “y,” can be modeled as a function of one or more *explanatory variables*, often referred to with the letter “x.” In this way, we are interested in determining how much variability in outcomes can be explained by random variables that were measured or controlled as part of an experiment. The linear model can be expanded easily to the more generalized form, in which we include multiple outcome variables simultaneously. These models are referred to as General Linear Models, and can be found in more advanced statistics books.

DEFINITION

An *outcome variable* is represented by the set of measured values that result from an experiment or some other statistical process. An *explanatory variable*, on the other hand, is a variable that is useful for predicting the value of the outcome variable.

A linear model is any model that is linear in the parameters that define the model. We can represent such models generically in the form:

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \dots + \beta_k X_{kj} + \epsilon_j \quad (8.1.1)$$

In this equation, β_j represents the coefficients in the model and ϵ_j represents random error. Therefore, any model that can be represented in this form, where the coefficients are constants and the algebraic order of the model is one, is considered a linear model. Though at first glance this equation may seem daunting, it actually is generally easy to find values for the parameters using basic algebra or calculus, as we shall see as the chapter progresses.

We will see many representations of linear models in this and other forms in the next several chapters. In particular, we will focus on the use of linear models for analyzing data using the analysis of variance for testing differences among means, regression for making predictions, and correlation for understanding associations among variables. In the context of analysis of variance, the predictor variables are *classification variables* used to define factors of interest (e.g., differentiating between a control group and a treatment group), and in the context of correlation and linear regression the predictor variables are most often continuous variables, or at least variables at a higher level than nominal classes. Though the underlying purposes of these tasks may seem quite different, studying these techniques and

the structure of the models used to represent them will prove to be valuable for understanding some of the most commonly used inferential statistics.

Analysis of Variance This chapter is concerned with *analysis of variance*, which may be defined as *a technique whereby the total variation present in a set of data is partitioned into two or more components. Associated with each of these components is a specific source of variation, so that in the analysis it is possible to ascertain the magnitude of the contributions of each of these sources to the total variation.*

The development of analysis of variance (ANOVA) is due mainly to the work of R. A. Fisher (1), whose contributions to statistics, spanning the years 1912 to 1962, have had a tremendous influence on modern statistical thought (2,3).

Applications Analysis of variance finds its widest application in the analysis of data derived from experiments. The principles of the design of experiments are well covered in many books, including those by Hinkelmann and Kempthorne (4), Montgomery (5), and Myers and Well (6). We do not study this topic in detail, since to do it justice would require a minimum of an additional chapter. Some of the important concepts in experimental design, however, will become apparent as we discuss analysis of variance.

Analysis of variance is used for two different purposes: (1) to estimate and test hypotheses about population variances, and (2) to estimate and test hypotheses about population means. We are concerned here with the latter use. However, as we will see, our conclusions regarding the means will depend on the magnitudes of the observed variances.

The concepts and techniques that we cover under the heading of analysis of variance are extensions of the concepts and techniques covered in Chapter 7. In Chapter 7 we learned to test the null hypothesis that two means are equal. In this chapter we learn to test the null hypothesis that three or more means are equal. Whereas, for example, what we learned in Chapter 7 enables us to determine if we can conclude that two treatments differ in effectiveness, what we learn in this chapter enables us to determine if we can conclude that three or more treatments differ in effectiveness. The following example illustrates some basic ideas involved in the application of analysis of variance. These will be extended and elaborated on later in this chapter.

EXAMPLE 8.1.1

Suppose we wish to know if three drugs differ in their effectiveness in lowering serum cholesterol in human subjects. Some subjects receive drug A, some drug B, and some drug C. After a specified period of time, measurements are taken to determine the extent to which serum cholesterol was reduced in each subject. We find that the amount by which serum cholesterol was lowered is not the same in all subjects. In other words, there is *variability among* the measurements. Why, we ask ourselves, are the measurements not all the same? Presumably, one reason they are not the same is that the subjects received different drugs. We now look at the measurements of those subjects who received drug A. We find that the amount by which serum cholesterol was lowered is not the same among these subjects. We find this to be the case when we look at the measurements for subjects

who received drug B and those subjects who received drug C. We see that there is *variability* among the measurements *within* the treatment groups. Why, we ask ourselves again, are these measurements not the same? Among the reasons that come to mind are differences in the genetic makeup of the subjects and differences in their diets. Through an analysis of the *variability* that we have observed, we will be able to reach a conclusion regarding the equivalence of the effectiveness of the three drugs. To do this we employ the techniques and concepts of analysis of variance. ■

Variables In our example we allude to three kinds of variables. We find these variables to be present in all situations in which the use of analysis of variance is appropriate. First, we have the *treatment variable*, which in our example was “drug.” We had three “values” of this variable, drug A, drug B, and drug C. The second kind of variable we refer to is the *response variable*. In the example it is change in serum cholesterol. The response variable is the variable that we expect to exhibit different values when different “values” of the treatment variable are employed. Finally, we have the other variables that we mention—genetic composition and diet. These are called *extraneous variables*. These variables may have an effect on the response variable, but they are not the focus of our attention in the experiment. The treatment variable is the variable of primary concern, and the question to be answered is: Do the different “values” of the treatment variable result in differences, on the average, in the response variable?

Assumptions Underlying the valid use of analysis of variance as a tool of statistical inference is a set of fundamental assumptions. Although an experimenter must not expect to find all the assumptions met to perfection, it is important that the user of analysis of variance techniques be aware of the underlying assumptions and be able to recognize when they are substantially unsatisfied. Because experiments in which all the assumptions are perfectly met are rare, analysis of variance results should be considered as approximate rather than exact. These assumptions are pointed out at appropriate points in the following sections.

We discuss analysis of variance as it is used to analyze the results of two different experimental designs, the completely randomized and the randomized complete block designs. In addition to these, the concept of a factorial experiment is given through its use in a completely randomized design. These do not exhaust the possibilities. A discussion of additional designs may be found in the references (4–6).

The ANOVA Procedure In our presentation of the analysis of variance for the different designs, we follow the ten-step procedure presented in Chapter 7. The following is a restatement of the steps of the procedure, including some new concepts necessary for its adaptation to analysis of variance.

1. **Description of data.** In addition to describing the data in the usual way, we display the sample data in tabular form.
2. **Assumptions.** Along with the assumptions underlying the analysis, we present the model for each design we discuss. The model consists of a symbolic representation of a typical value from the data being analyzed.
3. **Hypotheses.**

4. **Test statistic.**
5. **Distribution of test statistic.**
6. **Decision rule.**
7. **Calculation of test statistic.** The results of the arithmetic calculations will be summarized in a table called the analysis of variance (ANOVA) table. The entries in the table make it easy to evaluate the results of the analysis.
8. **Statistical decision.**
9. **Conclusion.**
10. **Determination of p value.**

We discuss these steps in greater detail in Section 8.2.

The Use of Computers The calculations required by analysis of variance are lengthier and more complicated than those we have encountered in preceding chapters. For this reason the computer assumes an important role in analysis of variance. All the exercises appearing in this chapter are suitable for computer analysis and may be solved with the statistical packages mentioned in Chapter 1. The output of the statistical packages may vary slightly from that presented in this chapter, but this should pose no major problem to those who use a computer to analyze the data of the exercises. The basic concepts of analysis of variance that we present here should provide the necessary background for understanding the description of the programs and their output in any of the statistical packages.

8.2 THE COMPLETELY RANDOMIZED DESIGN

We saw in Chapter 7 how it is possible to test the null hypothesis of no difference between two population means. It is not unusual for the investigator to be interested in testing the null hypothesis of no difference among several population means. The student first encountering this problem might be inclined to suggest that all possible pairs of sample means be tested separately by means of the Student t test. Suppose there are five populations involved. The number of possible pairs of sample means is ${}_5C_2 = 10$. As the amount of work involved in carrying out this many t tests is substantial, it would be worthwhile if a more efficient alternative for analysis were available. A more important consequence of performing all possible t tests, however, is that it is very likely to lead to a false conclusion.

Suppose we draw five samples from populations having equal means. As we have seen, there would be 10 tests if we were to do each of the possible tests separately. If we select a significance level of $\alpha = .05$ for each test, the probability of failing to reject a hypothesis of no difference in each case would be .95. By the multiplication rule of probability, if the tests were independent of one another, the probability of failing to reject a hypothesis of no difference in all 10 cases would be $(.95)^{10} = .5987$. The probability of rejecting at least one hypothesis of no difference, then, would be $1 - .5987 = .4013$. Since we know that the null hypothesis is true in every case in this illustrative example, rejecting the null hypothesis constitutes the committing of a type I error. In the long run, then, in

testing all possible pairs of means from five samples, we would commit a type I error 40 percent of the time. The problem becomes even more complicated in practice, since three or more t tests based on the same data would not be independent of one another.

It becomes clear, then, that some other method for testing for a significant difference among several means is needed. Analysis of variance provides such a method.

One-Way ANOVA The simplest type of analysis of variance is that known as *one-way analysis of variance*, in which only one source of variation, or *factor*, is investigated. It is an extension to three or more samples of the t test procedure (discussed in Chapter 7) for use with two independent samples. Stated another way, we can say that the t test for use with two independent samples is a special case of one-way analysis of variance.

In a typical situation we want to use one-way analysis of variance to test the null hypothesis that three or more treatments are equally effective. The necessary experiment is designed in such a way that the treatments of interest are assigned completely at random to the subjects or objects on which the measurements to determine treatment effectiveness are to be made. For this reason the design is called the *completely randomized experimental design*.

We may randomly allocate subjects to treatments as follows. Suppose we have 16 subjects available to participate in an experiment in which we wish to compare four drugs. We number the subjects from 01 through 16. We then go to a table of random numbers and select 16 consecutive, unduplicated numbers between 01 and 16. To illustrate, let us use Appendix Table A and a random starting point that, say, is at the intersection of Row 4 and Columns 11 and 12. The two-digit number at this intersection is 98. The succeeding (moving downward) 16 consecutive two-digit numbers between 01 and 16 are 16, 09, 06, 15, 14, 11, 02, 04, 10, 07, 05, 13, 03, 12, 01, and 08. We allocate subjects 16, 09, 06, and 15 to drug A; subjects 14, 11, 02, and 04 to drug B; subjects 10, 07, 05, and 13 to drug C; and subjects 03, 12, 01, and 08 to drug D. We emphasize that the number of subjects in each treatment group does not have to be the same. Figure 8.2.1 illustrates the scheme of random allocation.

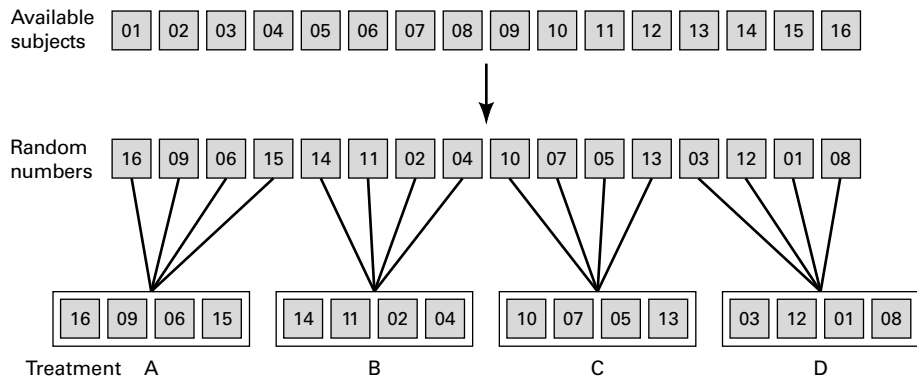


FIGURE 8.2.1 Allocation of subjects to treatments, completely randomized design.

TABLE 8.2.1 Table of Sample Values for the Completely Randomized Design

	Treatment					
	1	2	3	...	k	
	X_{11}	X_{12}	X_{13}	...	X_{1k}	
	X_{21}	X_{22}	X_{23}	...	X_{2k}	
	X_{31}	X_{32}	X_{33}	...	X_{3k}	
	\vdots	\vdots	\vdots	\vdots	\vdots	
	$X_{n_1,1}$	$X_{n_2,2}$	$X_{n_3,3}$...	$X_{n_k,k}$	
Total	$T_{.1}$	$T_{.2}$	$T_{.3}$...	$T_{.k}$	$T_{..}$
Mean	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.3}$...	$\bar{x}_{.k}$	$\bar{x}_{..}$

Hypothesis Testing Steps Once we decide that the completely randomized design is the appropriate design, we may proceed with the hypothesis testing steps. We discuss these in detail first, and follow with an example.

- 1. Description of data.** The measurements (or observations) resulting from a completely randomized experimental design, along with the means and totals that can be computed from them, may be displayed for convenience as in Table 8.2.1. The symbols used in Table 8.2.1 are defined as follows:

x_{ij} = the i th observation resulting from the j th treatment
(there are a total of k treatments)

$i = 1, 2, \dots, n_j, \quad j = 1, 2, \dots, k$

$T_{.j} = \sum_{i=1}^{n_j} x_{ij}$ = total of the j th treatment

$\bar{x}_{.j} = \frac{T_{.j}}{n_j}$ = mean of the j th treatment

$T_{..} = \sum_{j=1}^k T_{.j} = \sum_{j=1}^k \sum_{i=1}^{n_j} x_{ij}$ = total of all observations

$\bar{x}_{..} = \frac{T_{..}}{N}, \quad N = \sum_{j=1}^k n_j$

- 2. Assumptions.** Before stating the assumptions, let us specify the model for the experiment described here.

The Model As already noted, a model is a symbolic representation of a typical value of a data set. To write down the model for the completely randomized experimental design, let us begin by identifying a typical value from the set of data represented by the sample displayed in Table 8.2.1. We use the symbol x_{ij} to represent this typical value.

The one-way analysis of variance model may be written as follows:

$$x_{ij} = \mu + \tau_j + \epsilon_{ij}; \quad i = 1, 2, \dots, n_j, \quad j = 1, 2, \dots, k \quad (8.2.1)$$

The terms in this model are defined as follows:

1. μ represents the mean of all k population means and is called the *grand mean*.
2. τ_j represents the difference between the mean of the j th population and the grand mean and is called the *treatment effect*.
3. ϵ_{ij} represents the amount by which an individual measurement differs from the mean of the population to which it belongs and is called the *error term*.

Components of the Model By looking at our model we can see that a typical observation from the total set of data under study is composed of (1) the grand mean, (2) a treatment effect, and (3) an error term representing the deviation of the observation from its group mean.

In most situations we are interested only in the k treatments represented in our experiment. Any inferences that we make apply only to these treatments. We do not wish to extend our inference to any larger collection of treatments. When we place such a restriction on our inference goals, we refer to our model as the *fixed-effects model*, or *model 1*. The discussion in this book is limited to this model.

Assumptions of the Model The assumptions for the fixed-effects model are as follows:

- (a) The k sets of observed data constitute k independent random samples from the respective populations.
 - (b) Each of the populations from which the samples come is normally distributed with mean μ_j and variance σ_j^2 .
 - (c) Each of the populations has the same variance. That is, $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2$ the common variance.
 - (d) The τ_j are unknown constants and $\sum \tau_j = 0$ since the sum of all deviations of the μ_j from their mean, μ , is zero.
 - (e) The ϵ_{ij} have a mean of 0, since the mean of x_{ij} is μ_j .
 - (f) The ϵ_{ij} have a variance equal to the variance of the x_{ij} , since the ϵ_{ij} and x_{ij} differ only by a constant; that is, the error variance is equal to σ^2 , the common variance specified in assumption c.
 - (g) The ϵ_{ij} are normally (and independently) distributed.
3. **Hypotheses.** We test the null hypothesis that all population or treatment means are equal against the alternative that the members of at least one pair are not equal. We may state the hypotheses formally as follows:

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

$$H_A : \text{not all } \mu_j \text{ are equal}$$

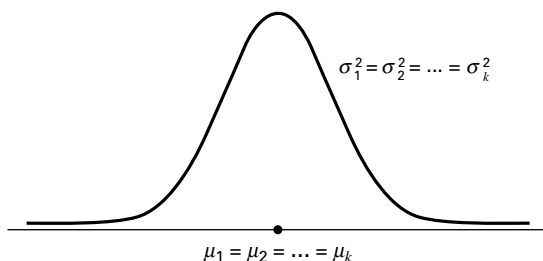


FIGURE 8.2.2 Picture of the populations represented in a completely randomized design when H_0 is true and the assumptions are met.

If the population means are equal, each treatment effect is equal to zero, so that, alternatively, the hypotheses may be stated as

$$\begin{aligned} H_0 : \tau_j &= 0, & j &= 1, 2, \dots, k \\ H_A : &\text{not all } \tau_j = 0 \end{aligned}$$

If H_0 is true and the assumptions of equal variances and normally distributed populations are met, a picture of the populations will look like Figure 8.2.2. When H_0 is true the population means are all equal, and the populations are centered at the same point (the common mean) on the horizontal axis. If the populations are all normally distributed with equal variances the distributions will be identical, so that in drawing their pictures each is superimposed on each of the others, and a single picture sufficiently represents them all.

When H_0 is false it may be false because one of the population means is different from the others, which are all equal. Or, perhaps, all the population means are different. These are only two of the possibilities when H_0 is false. There are many other possible combinations of equal and unequal means. Figure 8.2.3 shows a picture of the populations when the assumptions are met, but H_0 is false because no two population means are equal.

4. Test statistic. The test statistic for one-way analysis of variance is a computed variance ratio, which we designate by V.R. as we did in Chapter 7. The two

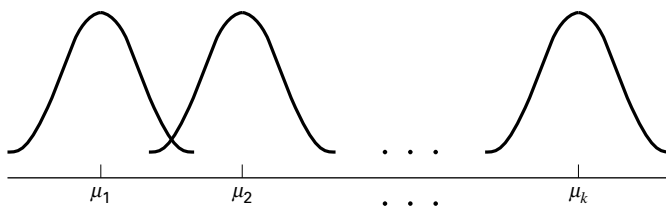


FIGURE 8.2.3 Picture of the populations represented in a completely randomized design when the assumptions of equal variances and normally distributed populations are met, but H_0 is false because none of the population means are equal.

variances from which V.R. is calculated are themselves computed from the sample data. The methods by which they are calculated will be given in the discussion that follows.

5. **Distribution of test statistic.** As discussed in Section 7.8, V.R. is distributed as the F distribution when H_0 is true and the assumptions are met.
6. **Decision rule.** In general, the decision rule is: reject the null hypothesis if the computed value of V.R. is equal to or greater than the critical value of F for the chosen α level.
7. **Calculation of test statistic.** We have defined analysis of variance as a process whereby the total variation present in a set of data is partitioned into components that are attributable to different sources. The term *variation* used in this context refers to the *sum of squared deviations of observations from their mean*, or *sum of squares* for short.

The initial computations performed in one-way ANOVA consist of the partitioning of the total variation present in the observed data into its basic components, each of which is attributable to an identifiable source.

Those who use a computer for calculations may wish to skip the following discussion of the computations involved in obtaining the test statistic.

The Total Sum of Squares Before we can do any partitioning, we must first obtain the total sum of squares. The total sum of squares is the sum of the squares of the deviations of individual observations from the mean of all the observations taken together. This *total sum of squares* is defined as

$$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{..})^2 \quad (8.2.2)$$

where $\sum_{i=1}^{n_j}$ tells us to sum the squared deviations for each treatment group, and $\sum_{j=1}^k$ tells us to add the k group totals obtained by applying $\sum_{i=1}^{n_j}$. The reader will recognize Equation 8.2.2 as the numerator of the variance that may be computed from the complete set of observations taken together.

The Within Groups Sum of Squares Now let us show how to compute the first of the two components of the total sum of squares.

The first step in the computation calls for performing certain calculations *within* each group. These calculations involve computing within each group the sum of the squared deviations of the individual observations from their mean. When these calculations have been performed within each group, we obtain the sum of the individual group results. This component of variation is called the *within groups sum of squares* and may be designated SSW . This quantity is sometimes referred to as the *residual* or *error* sum of squares. The expression for these calculations is written as follows:

$$SSW = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \quad (8.2.3)$$

The Among Groups Sum of Squares To obtain the second component of the total sum of squares, we compute for each group the squared deviation of the group mean from the grand mean and multiply the result by the size of the group. Finally, we add these results over all groups. This quantity is a measure of the variation among groups and is referred to as the *sum of squares among groups* or *SSA*. The formula for calculating this quantity is as follows:

$$SSA = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x}_{..})^2 \quad (8.2.4)$$

In summary, then, we have found that the total sum of squares is equal to the sum of the among and the within sum of squares. We express this relationship as follows:

$$SST = SSA + SSW$$

From the sums of squares that we have now learned to compute, it is possible to obtain two estimates of the common population variance, σ^2 . It can be shown that when the assumptions are met and the population means are all equal, both the among sum of squares and the within sum of squares, when divided by their respective degrees of freedom, yield independent and unbiased estimates of σ^2 .

The First Estimate of σ^2 Within any sample,

$$\frac{\sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{n_j - 1}$$

provides an unbiased estimate of the true variance of the population from which the sample came. Under the assumption that the population variances are all equal, we may pool the k estimates to obtain

$$MSW = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2}{\sum_{j=1}^k (n_j - 1)} \quad (8.2.5)$$

This is our first estimate of σ^2 and may be called the *within groups variance*, since it is the within groups sum of squares of Equation 8.2.3 divided by the appropriate degrees of freedom. The student will recognize this as an extension to k samples of the pooling of variances procedure encountered in Chapters 6 and 7 when the variances from two samples were pooled in order to use the t distribution. The quantity in Equation 8.2.5 is customarily referred to as the within groups *mean square* rather than the within groups variance.

The within groups mean square is a valid estimate of σ^2 only if the population variances are equal. It is not necessary, however, for H_0 to be true in order for the within groups mean square to be a valid estimate of σ^2 ; that is, the within groups mean square estimates σ^2 regardless of whether H_0 is true or false, as long as the population variances are equal.

The Second Estimate of σ^2 The second estimate of σ^2 may be obtained from the familiar formula for the variance of sample means, $\sigma_{\bar{x}}^2 = \sigma^2/n$. If we solve this equation for σ^2 , the variance of the population from which the samples were drawn, we have

$$\sigma^2 = n\sigma_{\bar{x}}^2 \quad (8.2.6)$$

An unbiased estimate of $\sigma_{\bar{x}}^2$ computed from sample data is provided by

$$\frac{\sum_{j=1}^k (\bar{x}_j - \bar{x}_{..})^2}{k-1}$$

If we substitute this quantity into Equation 8.2.6, we obtain the desired estimate of σ^2 ,

$$\text{MSA} = \frac{n \sum_{j=1}^k (\bar{x}_j - \bar{x}_{..})^2}{k-1} \quad (8.2.7)$$

The reader will recognize the numerator of Equation 8.2.7 as the among groups sum of squares for the special case when all sample sizes are equal. This sum of squares when divided by the associated degrees of freedom $k-1$ is referred to as the *among groups mean square*.

When the sample sizes are not all equal, an estimate of σ^2 based on the variability among sample means is provided by

$$\text{MSA} = \frac{\sum_{j=1}^k n_j (\bar{x}_j - \bar{x}_{..})^2}{k-1} \quad (8.2.8)$$

If, indeed, the null hypothesis is true we would expect these two estimates of σ^2 to be fairly close in magnitude. If the null hypothesis is false, that is, if all population means are not equal, we would expect the among groups mean square, which is computed by using the squared deviations of the sample means from the overall mean, to be larger than the within groups mean square.

In order to understand analysis of variance we must realize that the among groups mean square provides a valid estimate of σ^2 when the assumption of equal population

variances is met *and when* H_0 is true. Both conditions, a true null hypothesis and equal population variances, must be met in order for the among groups mean square to be a valid estimate of σ^2 .

The Variance Ratio What we need to do now is to compare these two estimates of σ^2 , and we do this by computing the following variance ratio, which is the desired test statistic:

$$\text{V.R.} = \frac{\text{among groups mean square}}{\text{within groups mean square}} = \frac{\text{MSA}}{\text{MSW}}$$

If the two estimates are about equal, V.R. will be close to 1. A ratio close to 1 tends to support the hypothesis of equal population means. If, on the other hand, the among groups mean square is considerably larger than the within groups mean square, V.R. will be considerably greater than 1. A value of V.R. sufficiently greater than 1 will cast doubt on the hypothesis of equal population means.

We know that because of the vagaries of sampling, even when the null hypothesis is true, it is unlikely that the among and within groups mean squares will be equal. We must decide, then, how big the observed difference must be before we can conclude that the difference is due to something other than sampling fluctuation. In other words, how large a value of V.R. is required for us to be willing to conclude that the observed difference between our two estimates of σ^2 is not the result of chance alone?

The F Test To answer the question just posed, we must consider the sampling distribution of the ratio of two sample variances. In Chapter 6 we learned that the quantity $(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$ follows a distribution known as the F distribution when the sample variances are computed from random and independently drawn samples from normal populations. The F distribution, introduced by R. A. Fisher in the early 1920s, has become one of the most widely used distributions in modern statistics. We have already become acquainted with its use in constructing confidence intervals for, and testing hypotheses about, population variances. In this chapter, we will see that it is the distribution fundamental to analysis of variance. For this reason the ratio that we designate V.R. is frequently referred to as F , and the testing procedure is frequently called the F test. It is of interest to note that the F distribution is the ratio of two Chi-square distributions.

In Chapter 7 we learned that when the population variances are the same, they cancel in the expression $(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$, leaving s_1^2/s_2^2 , which is itself distributed as F . The F distribution is really a family of distributions, and the particular F distribution we use in a given situation depends on the number of degrees of freedom associated with the sample variance in the numerator (*numerator degrees of freedom*) and the number of degrees of freedom associated with the sample variance in the denominator (*denominator degrees of freedom*).

Once the appropriate F distribution has been determined, the size of the observed V.R. that will cause rejection of the hypothesis of equal population variances depends on the significance level chosen. The significance level chosen determines the critical value of F , the value that separates the nonrejection region from the rejection region.

TABLE 8.2.2 Analysis of Variance Table for the Completely Randomized Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Variance Ratio
Among samples	$SSA = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x}_{..})^2$	$k - 1$	$MSA = SSA/(k - 1)$	$V.R. = \frac{MSA}{MSW}$
Within samples	$SSW = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$	$N - k$	$MSW = SSW/(N - k)$	
Total	$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{..})^2$	$N - 1$		

As we have seen, we compute V.R. in situations of this type by placing the among groups mean square in the numerator and the within groups mean square in the denominator, so that the numerator degrees of freedom is equal to $(k - 1)$, the number of groups minus 1, and the denominator degrees of freedom value is equal to

$$\sum_{j=1}^k (n_j - 1) = \left(\sum_{j=1}^k n_j \right) - k = N - k$$

The ANOVA Table The calculations that we perform may be summarized and displayed in a table such as Table 8.2.2, which is called the ANOVA table.

8. Statistical decision. To reach a decision we must compare our computed V.R. with the critical value of F , which we obtain by entering Appendix Table G with $k - 1$ numerator degrees of freedom and $N - k$ denominator degrees of freedom.

If the computed V.R. is equal to or greater than the critical value of F , we reject the null hypothesis. If the computed value of V.R. is smaller than the critical value of F , we do not reject the null hypothesis.

Explaining a Rejected Null Hypothesis There are two possible explanations for a rejected null hypothesis. If the null hypothesis is true, that is, if the two sample variances are estimates of a common variance, we know that the probability of getting a value of V.R. as large as or larger than the critical F is equal to our chosen level of significance. When we reject H_0 we may, if we wish, conclude that the null hypothesis is true and assume that because of chance we got a set of data that gave rise to a rare event. On the other hand, we may prefer to take the position that our large computed V.R. value does not represent a rare event brought about by chance but, instead, reflects the fact that something other than chance is operative. We then conclude that we have a false null hypothesis.

It is this latter explanation that we usually give for computed values of V.R. that exceed the critical value of F . In other words, if the computed value of V.R. is greater than the critical value of F , we reject the null hypothesis.

It will be recalled that the original hypothesis we set out to test was

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

Does rejection of the hypothesis about variances imply a rejection of the hypothesis of equal population means? The answer is yes. A large value of V.R. resulted from the fact that the among groups mean square was considerably larger than the within groups mean square. Since the among groups mean square is based on the dispersion of the sample means about their mean (called the grand mean), this quantity will be large when there is a large discrepancy among the sizes of the sample means. Because of this, then, a significant value of V.R. tells us to reject the null hypothesis that all population means are equal.

9. Conclusion. When we reject H_0 , we conclude that not all population means are equal. When we fail to reject H_0 , we conclude that the population means are not significantly different from each other.

10. Determination of p value.

EXAMPLE 8.2.1

Game meats, including those from white-tailed deer and eastern gray squirrels, are used as food by families, hunters, and other individuals for health, cultural, or personal reasons. A study by David Holben (A-1) assessed the selenium content of meat from free-roaming white-tailed deer (venison) and gray squirrel (squirrel) obtained from a low selenium region of the United States. These selenium content values were also compared to those of beef produced within and outside the same region. We want to know if the selenium levels are different among the four meat groups.

Solution:

1. Description of data. Selenium content of raw venison (VEN), squirrel meat (SQU), region-raised beef (RRB), and nonregion-raised beef (NRB), in $\mu\text{g}/100\text{ g}$ of dry weight, are shown in Table 8.2.3. A graph of the data in the form of a *dotplot* is shown in Figure 8.2.4. Such a graph highlights the main features of the data and brings into clear focus differences in selenium levels among the different meats.

TABLE 8.2.3 Selenium Content, in $\mu\text{g}/100\text{ g}$, of Four Different Meat Types

Meat Type						
VEN	SQU	RRB	NRB			
26.72	14.86	37.42	37.57	11.23	15.82	44.33
28.58	16.47	56.46	25.71	29.63	27.74	76.86
29.71	25.19	51.91	23.97	20.42	22.35	4.45
26.95	37.45	62.73	13.82	10.12	34.78	55.01
10.97	45.08	4.55	42.21	39.91	35.09	58.21
21.97	25.22	39.17	35.88	32.66	32.60	74.72

(Continued)

Meat Type						
VEN	SQU	RRB	NRB	VEN	SQU	NRB
14.35	22.11	38.44	10.54	38.38	37.03	11.84
32.21	33.01	40.92	27.97	36.21	27.00	139.09
19.19	31.20	58.93	41.89	16.39	44.20	69.01
30.92	26.50	61.88	23.94	27.44	13.09	94.61
10.42	32.77	49.54	49.81	17.29	33.03	48.35
35.49	8.70	64.35	30.71	56.20	9.69	37.65
36.84	25.90	82.49	50.00	28.94	32.45	66.36
25.03	29.80	38.54	87.50	20.11	37.38	72.48
33.59	37.63	39.53	68.99	25.35	34.91	87.09
33.74	21.69			21.77	27.99	26.34
18.02	21.49			31.62	22.36	71.24
22.27	18.11			32.63	22.68	90.38
26.10	31.50			30.31	26.52	50.86
20.89	27.36			46.16	46.01	
29.44	21.33			56.61	38.04	
				24.47	30.88	
				29.39	30.04	
				40.71	25.91	
				18.52	18.54	
				27.80	25.51	
				19.49		

Source: Data provided courtesy of David H. Holben, Ph.D.

2. Assumptions. We assume that the four sets of data constitute independent simple random samples from the four indicated populations. We assume that the four populations of measurements are normally distributed with equal variances.

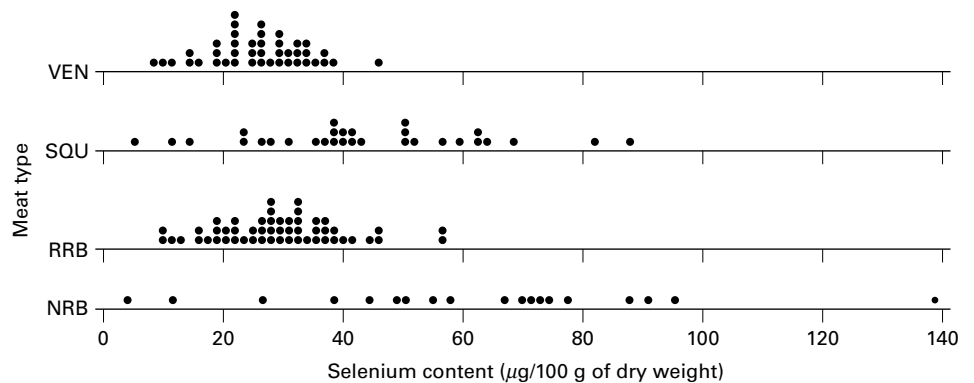


FIGURE 8.2.4 Selenium content of four meat types. VEN = venison, SQU = squirrel, RRB = region-raised beef, and NRB = nonregion-raised beef.

TABLE 8.2.4 ANOVA Table for Example 8.2.1

Source	SS	df	MS	F
Among samples	21261.82886	3	7087.27629	27.00
Within samples	36747.22674	140	262.48019	
Total	58009.05560	143		

3. **Hypotheses.** $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ (On average the four meats have the same selenium content.)

H_A : Not all μ 's are equal (At least one meat yields an average selenium content different from the average selenium content of at least one other meat.)

4. **Test statistic.** The test statistic is V.R. = MSA/MSW .
5. **Distribution of test statistic.** If H_0 is true and the assumptions are met, the V.R. follows the F distribution with $4 - 1 = 3$ numerator degrees of freedom and $144 - 4 = 140$ denominator degrees of freedom.
6. **Decision rule.** Suppose we let $\alpha = .01$. The critical value of F from Appendix Table G is < 3.95 . The decision rule, then, is reject H_0 if the computed V.R. statistic is equal to or greater than 3.95.
7. **Calculation of test statistic.** By Equation 8.2.2 we compute

$$SST = 58009.05560$$

By Equation 8.2.4 we compute

$$SSA = 21261.82886$$

$$SSW = 58009.05560 - 21261.82886 = 36747.22674$$

The results of our calculations are displayed in Table 8.2.4.

8. **Statistical decision.** Since our computed F of 27.00 is greater than 3.95 we reject H_0 .
9. **Conclusion.** Since we reject H_0 , we conclude that the alternative hypothesis is true. That is, we conclude that the four meat types do not all have the same average selenium content.
10. **p value.** Since $27.00 > 3.95$, $p < .01$ for this test. ■

A Word of Caution The completely randomized design is simple and, therefore, widely used. It should be used, however, only when the units receiving the treatments are homogeneous. If the experimental units are not homogeneous, the researcher should consider an alternative design such as one of those to be discussed later in this chapter.

In our illustrative example the treatments are treatments in the usual sense of the word. This is not always the case, however, as the term “treatment” as used in experimental design is quite general. We might, for example, wish to study the response to the same

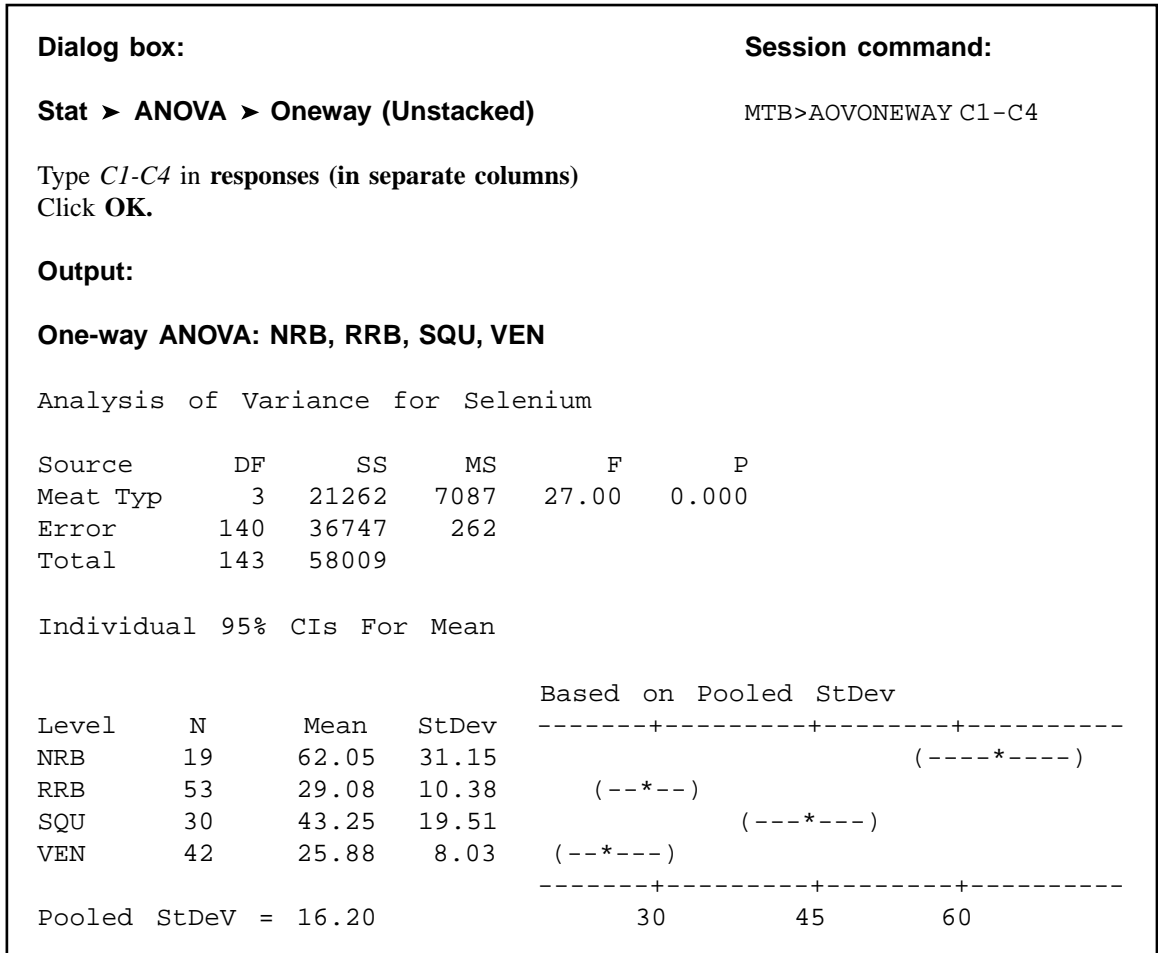


FIGURE 8.2.5 MINITAB procedure and output for Example 8.2.1.

treatment (in the usual sense of the word) of several breeds of animals. We would, however, refer to the breed of animal as the “treatment.”

We must also point out that, although the techniques of analysis of variance are more often applied to data resulting from controlled experiments, the techniques also may be used to analyze data collected by a survey, provided that the underlying assumptions are reasonably well met.

Computer Analysis Figure 8.2.5 shows the computer procedure and output for Example 8.2.1 provided by a one-way analysis of variance program found in the MINITAB package. The data were entered into Columns 1 through 4. When you compare the ANOVA table on this printout with the one given in Table 8.2.4, you see that the printout uses the label “factor” instead of “among samples.” The different treatments are referred to on the printout as levels. Thus level 1 = treatment 1, level 2 = treatment 2, and so on. The printout gives the four sample means and standard deviations as well as the pooled

The SAS System**Analysis of Variance Procedure**

Dependent Variable: selen

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	21261.82886	7087.27629	27.00	<.0001
Error	140	36747.22674	262.48019		
Corrected Total	143	58009.05560			
R-Square	Coeff Var	Root MSE	selen Mean		
0.366526	45.70507	16.20124	35.44736		

FIGURE 8.2.6 Partial SAS® printout for Example 8.2.1.

standard deviation. This last quantity is equal to the square root of the error mean square shown in the ANOVA table. Finally, the computer output gives graphic representations of the 95% confidence intervals for the mean of each of the four populations represented by the sample data.

Figure 8.2.6 contains a partial SAS® printout resulting from analysis of the data of Example 8.2.1 through use of the SAS® statement PROC ANOVA. SAS® computes some additional quantities as shown in the output. $R\text{-Square} = SSA/SST$. This quantity tells us what proportion of the total variability present in the observations is accounted for by differences in response to the treatments. $C.V. = 100(\text{root MSE}/\text{selen mean})$. Root MSE is the square root of MSW, and selen mean is the mean of all observations.

Note that the test statistic V.R. is labeled differently by different statistical software programs. MINITAB, for example, uses F rather than V.R. SAS® uses the label F Value.

A useful device for displaying important characteristics of a set of data analyzed by one-way analysis of variance is a graph consisting of side-by-side boxplots. For each sample a boxplot is constructed using the method described in Chapter 2. Figure 8.2.7 shows the side-by-side boxplots for Example 8.2.1. Note that in Figure 8.2.7 the variable of interest is represented by the vertical axis rather than the horizontal axis.

Alternatives If the data available for analysis do not meet the assumptions for one-way analysis of variance as discussed here, one may wish to consider the use of the Kruskal-Wallis procedure, a nonparametric technique discussed in Chapter 13.

Testing for Significant Differences Between Individual Pairs of Means When the analysis of variance leads to a rejection of the null hypothesis of no difference among population means, the question naturally arises regarding just which pairs of means are different. In fact, the desire, more often than not, is to carry out a significance test on each and every pair of treatment means. For instance, in

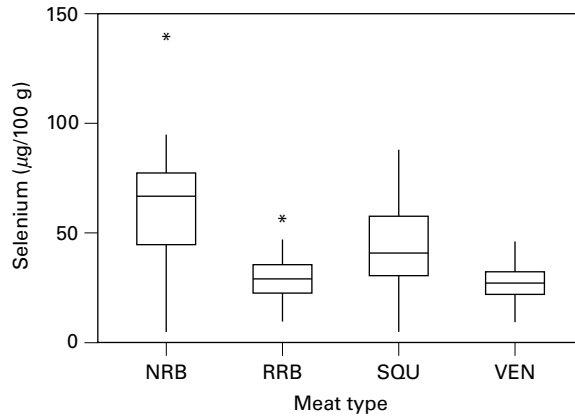


FIGURE 8.2.7 Side-by-side boxplots for Example 8.2.1.

Example 8.2.1, where there are four treatments, we may wish to know, after rejecting $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$, which of the six possible individual hypotheses should be rejected. The experimenter, however, must exercise caution in testing for significant differences between individual means and must always make certain that the procedure is valid. The critical issue in the procedure is the level of significance. Although the probability, α , of rejecting a true null hypothesis for the test as a whole is made small, the probability of rejecting at least one true hypothesis when several pairs of means are tested is, as we have seen, greater than α . There are several multiple comparison procedures commonly used in practice. Below we illustrate two popular procedures, namely Tukey's HSD test and Bonferroni's method. The interested student is referred to the books by Hsu (7) and Westfall et al. (8) for additional techniques.

Tukey's HSD Test Over the years several procedures for making multiple comparisons have been suggested. A multiple comparison procedure developed by Tukey (9) is frequently used for testing the null hypothesis that all possible pairs of treatment means are equal when the samples are all of the same size. When this test is employed we select an overall significance level of α . The probability is α , then, that one or more of the null hypotheses is false.

Tukey's test, which is usually referred to as the HSD (*honestly significant difference*) test, makes use of a single value against which all differences are compared. This value, called the HSD, is given by

$$\text{HSD} = q_{\alpha, k, N-k} \sqrt{\frac{\text{MSE}}{n}} \quad (8.2.9)$$

where α is the chosen level of significance, k is the number of means in the experiment, N is the total number of observations in the experiment, n is the number of observations in a treatment, MSE is the error or within mean square from the ANOVA table, and q is obtained by entering Appendix Table H with α , k , and $N - k$.

The statistic q , tabulated in Appendix Table H, is known as the studentized range statistic. It is defined as the difference between the largest and smallest treatment means from an ANOVA (that is, it is the range of the treatment means) divided by the error mean square over n , the number of observations in a treatment. The studentized range is discussed in detail by Winer (10).

All possible differences between pairs of means are computed, and any difference that yields an absolute value that exceeds HSD is declared significant.

Tukey's Test for Unequal Sample Sizes When the samples are not all the same size, as is the case in Example 8.2.1, Tukey's HSD test given by Equation 8.2.9 is not applicable. Tukey himself (9) and Kramer (11), however, have extended the Tukey procedure to the case where the sample sizes are different. Their procedure, which is sometimes called the Tukey-Kramer method, consists of replacing MSE/n in Equation 8.2.9 with $(MSE/2)(1/n_i + 1/n_j)$, where n_i and n_j are the sample sizes of the two groups to be compared. If we designate the new quantity by HSD^* , we have as the new test criterion

$$HSD^* = q_{\alpha, k, N-k} \sqrt{\frac{MSE}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad (8.2.10)$$

Any absolute value of the difference between two sample means that exceeds HSD^* is declared significant.

Bonferroni's Method Another very commonly used multiple comparison test is based on a method developed by C. E. Bonferroni. As with Tukey's method, we desire to maintain an overall significance level of α for the total of all pair-wise tests. In the Bonferroni method, we simply divide the desired significance level by the number of individual pairs that we are testing. That is, instead of testing at a significance level of α , we test at a significance level of α/k , where k is the number of paired comparisons. The sum of all α/k terms cannot, then, possibly exceed our stated level of α . For example, if one has three samples, A, B, and C, then there are $k = 3$ pair-wise comparisons. These are $\mu_A = \mu_B$, $\mu_A = \mu_C$, and $\mu_B = \mu_C$. If we choose a significance level of $\alpha = .05$, then we would proceed with the comparisons and use a Bonferroni-corrected significance level of $\alpha/3 = .017$. Therefore, our p value must be no greater than .017 in order to reject the null hypothesis and conclude that two means differ.

Most computer packages compute values using the Bonferroni method and produce an output similar to the Tukey's HSD or other multiple comparison procedures. In general, these outputs report the actual corrected p value using the Bonferroni method. Given the basic relationship that $p = \alpha/k$, then algebraically we can multiply both sides of the equation by k to obtain $\alpha = pk$. In other words, the total α is simply the sum of all of the pk values, and the actual corrected p value is simply the calculated p value multiplied by the number of tests that were performed.

EXAMPLE 8.2.2

Let us illustrate the use of the HSD test with the data from Example 8.2.1.

Solution: The first step is to prepare a table of all possible (ordered) differences between means. The results of this step for the present example are displayed in Table 8.2.5.

Suppose we let $\alpha = .05$. Entering Table H with $\alpha = .05$, $k = 4$, and $N - k = 140$, we find that $q < 3.68$. The actual value is $q = 3.667$, which can be obtained from SAS[®]. In Table 8.2.4 we have $MSE = 262.4802$.

The hypotheses that can be tested, the value of HSD^* , and the statistical decision for each test are shown in Table 8.2.6.

SAS[®] uses Tukey's procedure to test the hypothesis of no difference between population means for all possible pairs of sample means. The output also contains

TABLE 8.2.5 Differences Between Sample Means (Absolute Value) for Example 8.2.2

	VEN	RRB	SQR	NRB
VEN	–	3.208	17.37	36.171
RRB		–	14.163	32.963
SOU			–	18.801
NRB				–

TABLE 8.2.6 Multiple Comparison Tests Using Data of Example 8.2.1 and HSD^*

Hypotheses	HSD^*	Statistical Decision
$H_0: \mu_{VEN} = \mu_{RRB}$	$HSD^* = 3.677 \sqrt{\frac{262.4802}{2} \left(\frac{1}{42} + \frac{1}{53} \right)} = 8.68$	Do not reject H_0 since $3.208 < 8.68$
$H_0: \mu_{VEN} = \mu_{SQU}$	$HSD^* = 3.677 \sqrt{\frac{262.4802}{2} \left(\frac{1}{42} + \frac{1}{30} \right)} = 10.04$	Reject H_0 since $17.37 > 10.04$
$H_0: \mu_{VEN} = \mu_{NRB}$	$HSD^* = 3.677 \sqrt{\frac{262.4802}{2} \left(\frac{1}{42} + \frac{1}{19} \right)} = 11.61$	Reject H_0 since $36.171 > 11.61$
$H_0: \mu_{RRB} = \mu_{SQU}$	$HSD^* = 3.677 \sqrt{\frac{262.4802}{2} \left(\frac{1}{53} + \frac{1}{30} \right)} = 9.60$	Reject H_0 since $14.163 > 9.60$
$H_0: \mu_{RRB} = \mu_{NRB}$	$HSD^* = 3.677 \sqrt{\frac{262.4802}{2} \left(\frac{1}{53} + \frac{1}{19} \right)} = 11.23$	Reject H_0 since $32.963 > 11.23$
$H_0: \mu_{SQU} = \mu_{NRB}$	$HSD^* = 3.677 \sqrt{\frac{262.4802}{2} \left(\frac{1}{30} + \frac{1}{19} \right)} = 12.32$	Reject H_0 since $18.801 > 12.32$

The SAS System**Analysis of Variance Procedure**

Tukey's Studentized Range (HSD) Test for selen

NOTE: This test controls the Type I experimentwise error rate.

Alpha	0.05
Error Degrees of Freedom	140
Error Mean Square	262.4802
Critical Value of Studentized Range	3.67719

Comparisons significant at the 0.05 level are indicated by ***.

type Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
NRB - SQU	18.801	6.449	31.152	***
NRB - RRB	32.963	21.699	44.228	***
NRB - VEN	36.171	24.524	47.818	***
SQU - NRB	-18.801	-31.152	-6.449	***
SQU - RRB	14.163	4.538	23.787	***
SQU - VEN	17.370	7.300	27.440	***
RRB - NRB	-32.963	-44.228	-21.699	***
RRB - SQU	-14.163	-23.787	-4.538	***
RRB - VEN	3.208	-5.495	11.910	
VEN - NRB	-36.171	-47.818	-24.524	***
VEN - SQU	-17.370	-27.440	-7.300	***
VEN - RRB	-3.208	-11.910	5.495	

FIGURE 8.2.8 SAS® multiple comparisons for Example 8.2.1.

confidence intervals for the difference between all possible pairs of population means. This SAS output for Example 8.2.1 is displayed in Figure 8.2.8.

One may also use SPSS to perform multiple comparisons by a variety of methods, including Tukey's. The SPSS outputs for Tukey's HSD and Bonferroni's method for the data for Example 8.2.1 are shown in Figures 8.2.9 and 8.2.10, respectively. The outputs contain an exhaustive comparison of sample means, along with the associated standard errors, p values, and 95% confidence intervals. ■

Multiple Comparisons						
Dependent Variable: Selenium						
Tukey HSD						
(I) Meat_type	(J) Meat_type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
VEN	SQU	-17.370190*	3.872837210	.000	-27.44017302	-7.30020793
	RRB	-3.2075427	3.346936628	.773	-11.91010145	5.49501609
	NRB	-36.170840*	4.479316382	.000	-47.81776286	-24.52391634
SQU	VEN	17.370190*	3.872837210	.000	7.30020793	27.44017302
	RRB	14.162648*	3.701593729	.001	4.53792509	23.78737051
	NRB	-18.800649*	4.750167007	.001	-31.15182638	-6.44947187
RRB	VEN	3.2075427	3.346936628	.773	-5.49501609	11.91010145
	SQU	-14.162648*	3.701593729	.001	-23.78737051	-4.53792509
	NRB	-32.963297*	4.332113033	.000	-44.22746845	-21.69912540
NRB	VEN	36.170840*	4.479316382	.000	24.52391634	47.81776286
	SQU	18.800649*	4.750167007	.001	6.44947187	31.15182638
	RRB	32.963297*	4.332113033	.000	21.69912540	44.22746845

*The mean difference is significant at the .05 level.

FIGURE 8.2.9 SPSS output for Tukey's HSD using data from Example 8.2.1.

Multiple Comparisons						
Dependent Variable: Selenium						
Bonferroni						
(I) Meat_type	(J) Meat_type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
VEN	RRB	-3.20754	3.34694	1.000	-12.1648	5.7497
	SQU	-17.37019*	3.87284	.000	-27.7349	-7.0055
	NRB	-36.17084*	4.47932	.000	-48.1587	-24.1830
RRB	VEN	3.20754	3.34694	1.000	-5.7497	12.1648
	SQU	-14.16265*	3.70159	.001	-24.0691	-4.2562
	NRB	-32.96330*	4.33211	.000	-44.5572	-21.3694
SQU	VEN	17.37019*	3.87284	.000	7.0055	27.7349
	RRB	14.16265*	3.70159	.001	4.2562	24.0691
	NRB	-18.80065*	4.75017	.001	-31.5134	-6.0879
NRB	VEN	36.17084*	4.47932	.000	24.1830	48.1587
	RRB	32.96330*	4.33211	.000	21.3694	44.5572
	SQU	18.80065*	4.75017	.001	6.0879	31.5134

*The mean difference is significant at the .05 level.

FIGURE 8.2.10 SPSS output for Bonferroni's method using data from Example 8.2.1.

EXERCISES

In Exercise 8.2.1 to 8.2.7, go through the ten steps of analysis of variance hypothesis testing to see if you can conclude that there is a difference among population means. Let $\alpha = .05$ for each test. Use Tukey's HSD procedure to test for significant differences among individual pairs of means (if appropriate). Use the same α value for the F test. Construct a dot plot and side-by-side boxplots of the data.

- 8.2.1.** Researchers at Case Western Reserve University (A-2) wanted to develop and implement a transducer, manageable in a clinical setting, for quantifying isometric moments produced at the elbow joint by individuals with tetraplegia (paralysis or paresis of all four limbs). The apparatus, called an elbow moment transducer (EMT), measures the force the elbow can exert when flexing. The output variable is voltage. The machine was tested at four different elbow extension angles, 30, 60, 90, and 120 degrees, on a mock elbow consisting of two hinged aluminum beams. The data are shown in the following table.

Elbow Angle (Degrees)							
30		60		90		120	
-0.003	1.094	0.000	-0.001	0.000	-0.007	0.558	0.003
0.050	1.061	0.053	0.010	0.006	0.012	0.529	0.062
0.272	1.040	0.269	0.028	0.026	-0.039	0.524	0.287
0.552	1.097	0.555	0.055	0.053	-0.080	0.555	0.555
1.116	1.080	1.103	0.105	0.108	-0.118	0.539	1.118
2.733	1.051	2.727	0.272	0.278	-0.291	0.536	2.763
0.000	1.094	-0.002	0.553	0.555	-0.602	0.557	0.006
0.056	1.075	0.052	0.840	0.834	-0.884	0.544	0.050
0.275	1.035	0.271	1.100	1.106	-1.176	0.539	0.277
0.556	1.096	0.550	1.647	1.650	-1.725	1.109	0.557
1.100	1.100	1.097	2.728	2.729	0.003	1.085	1.113
2.723	1.096	2.725	-0.001	0.005	0.003	1.070	2.759
-0.003	1.108	0.003	0.014	-0.023	-0.011	1.110	0.010
0.055	1.099	0.052	0.027	-0.037	-0.060	1.069	0.060
0.273	1.089	0.270	0.057	-0.046	-0.097	1.045	0.286
0.553	1.107	0.553	0.111	-0.134	-0.320	1.110	0.564
1.100	1.094	1.100	0.276	-0.297	-0.593	1.066	1.104
2.713	1.092	2.727	0.555	-0.589	-0.840	1.037	2.760
0.007	1.092	0.022	0.832	-0.876	-1.168	2.728	-0.003
-0.066	1.104	-0.075	1.099	-1.157	-1.760	2.694	-0.060
-0.258	1.121	-0.298	1.651	-1.755	0.004	2.663	-0.289
-0.581	1.106	-0.585	2.736	-2.862	0.566	2.724	-0.585
-1.162	1.135	-1.168	0.564	0.000	1.116	2.693	-1.180
0.008	1.143	0.017	0.556	0.245	2.762	2.670	0.000
-0.045	1.106	-0.052	0.555	0.497	0.563	2.720	-0.034
-0.274	1.135	-0.258	0.567	0.001	0.551	2.688	-0.295
-0.604	1.156	-0.548	0.559	0.248	0.551	2.660	-0.579
-1.143	1.112	-1.187	0.551	0.498	0.561	0.556	-1.165
-0.004	1.104	0.019	1.107	0.001	0.555	0.560	-0.019

(Continued)

Elbow Angle (Degrees)							
30		60		90		120	
-0.050	1.107	-0.044	1.104	0.246	0.558	0.557	-0.056
-0.290	1.107	-0.292	1.102	0.491	0.551	0.551	-0.270
-0.607	1.104	-0.542	1.112	0.001	0.566	0.564	-0.579
-1.164	1.117	-1.189	1.103	0.262	0.560	0.555	-1.162
1.105	1.101		1.104	0.527	1.107	0.551	
	1.103		1.114	0.001	1.104	0.563	
			1.095	0.260	1.109	0.559	
			1.100	0.523	1.108	1.113	
			2.739	-0.005	1.106	1.114	
			2.721	0.261	1.102	1.101	
			2.687	0.523	1.111	1.113	
			2.732	2.696	1.102	1.113	
			2.702	2.664	1.107	1.097	
			2.660	2.722	2.735	1.116	
			2.743	2.686	2.733	1.112	
			2.687	2.661	2.659	1.098	
			2.656	0.548	2.727	2.732	
			2.733	2.739	0.542	2.722	
			2.731	2.742	0.556	2.734	
					2.728	2.747	

Source: Data provided courtesy of S. A. Snyder, M.S.

- 8.2.2.** Patients suffering from rheumatic diseases or osteoporosis often suffer critical losses in bone mineral density (BMD). Alendronate is one medication prescribed to build or prevent further loss of BMD. Holcomb and Rothenberg (A-3) looked at 96 women taking alendronate to determine if a difference existed in the mean percent change in BMD among five different primary diagnosis classifications. Group 1 patients were diagnosed with rheumatoid arthritis (RA). Group 2 patients were a mixed collection of patients with diseases including lupus, Wegener's granulomatosis and polyarteritis, and other vasculitic diseases (LUPUS). Group 3 patients had polymyalgia rheumatica or temporal arthritis (PMRTA). Group 4 patients had osteoarthritis (OA) and group 5 patients had osteoporosis (O) with no other rheumatic diseases identified in the medical record. Changes in BMD are shown in the following table.

Diagnosis					
RA	LUPUS	PMRTA	OA	O	
11.091	7.412	2.961	-3.669	11.146	2.937
24.414	5.559	0.293	-7.816	-0.838	15.968
10.025	4.761	8.394	4.563	4.082	5.349
-3.156	-3.527	2.832	-0.093	6.645	1.719
6.835	4.839	-1.369	-0.185	4.329	6.445
3.321	1.850	11.288	1.302	1.234	20.243
1.493	-3.933	3.997	5.299	-2.817	3.290

(Continued)

Diagnosis					
	RA	LUPUS	PMRTA	OA	O
-1.864	9.669	7.260	10.734	3.544	8.992
5.386	4.659	5.546	1.399	4.160	6.120
3.868	1.137		0.497	1.160	25.655
6.209	7.521		0.592	-0.247	
-5.640	0.073		3.950	5.372	
3.514	-8.684		0.674	6.721	
-2.308	-0.372		9.354	9.950	
15.981	21.311		2.610	10.820	
-9.646	10.831		5.682	7.280	
5.188	3.351			6.605	
-1.892	9.557			7.507	
16.553				5.075	
				0.163	
				12.767	
				3.481	
				0.917	
				15.853	

Source: Data provided courtesy of John P. Holcomb, Ph.D. and Ralph J. Rothenberg, M.D.

- 8.2.3. Ilich-Ernst et al. (A-4) investigated dietary intake of calcium among a cross section of 113 healthy women ages 20–88. The researchers formed four age groupings as follows: Group A, 20.0–45.9 years; group B, 46.0–55.9 years; group C, 56.0–65.9 years; and group D, over 66 years. Calcium from food intake was measured in mg/day. The data below are consistent with summary statistics given in the paper.

Age Groups (Years)				Age Groups (Years)			
A	B	C	D	A	B	C	D
1820	191	724	1652			1020	775
2588	1098	613	1309			805	1393
2670	644	918	1002			631	533
1022	136	949	966			641	734
1555	1605	877	788			760	485
222	1247	1368	472				449
1197	1529	1692	471				236
1249	1422	697	771				831
1520	445	849	869				698
489	990	1199	513				167
2575	489	429	731				824
1426	2408	798	1130				448
1846	1064	631	1034				991
1088	629	1016	1261				590
912		1025	42				994
1383		948	767				1781

(Continued)

Age Groups (Years)				Age Groups (Years)			
A	B	C	D	A	B	C	D
1483		1085	752				937
1723		775	804				1022
727		1307	1182				1073
1463		344	1243				948
1777		961	985				222
1129		239	1295				721
		944	1676				375
		1096	754				1187

- 8.2.4.** Gold et al. (A-5) investigated the effectiveness on smoking cessation of a nicotine patch, bupropion SR, or both, when co-administered with cognitive-behavioral therapy. Consecutive consenting patients ($n = 164$) assigned themselves to one of three treatments according to personal preference: nicotine patch (NTP, $n = 13$), bupropion SR (B; $n = 92$), and bupropion SR plus nicotine patch (BNTP, $n = 59$). At their first smoking cessation class, patients estimated the number of packs of cigarettes they currently smoked per day and the numbers of years they smoked. The “pack years” is the average number of packs the subject smoked per day multiplied by the number of years the subject had smoked. The results are shown in the following table.

	Pack Years				
	NTP	B		BNTP	
15	8	60	90	8	80
17	10	60	90	15	80
18	15	60	90	25	82
20	20	60	95	25	86
20	22	60	96	25	87
20	24	60	98	26	90
30	25	60	98	30	90
37	26	66	99	34	90
43	27	66	100	35	90
48	29	67	100	36	90
60	30	68	100	40	95
100	30	68	100	45	99
100	35	70	100	45	100
	35	70	100	45	102
	39	70	105	45	105
	40	75	110	48	105
	40	75	110	48	105
	40	75	120	49	111
	40	75	120	52	113
	40	76	123	60	120
	40	80	125	60	120
	45	80	125	60	125
	45	80	126	64	125

(Continued)

Pack Years					
NTP	B			BNTP	
	45	80	130	64	129
	50	80	130	70	130
	51	80	132	70	133
	52	80	132	70	135
	55	84	142	75	140
	58	84	157	75	154
	60	84	180	76	
	60	90			

Source: Data provided courtesy of Paul B. Gold, Ph.D.

- 8.2.5.** In a study by Wang et al. (A-6), researchers examined bone strength. They collected 10 cadaveric femurs from subjects in three age groups: young (19–49 years), middle-aged (50–69 years), and elderly (70 years or older) [Note: one value was missing in the middle-aged group]. One of the outcome measures (W) was the force in Newtons required to fracture the bone. The following table shows the data for the three age groups.

Young (Y)	Middle-aged (MA)	Elderly (E)
193.6	125.4	59.0
137.5	126.5	87.2
122.0	115.9	84.4
145.4	98.8	78.1
117.0	94.3	51.9
105.4	99.9	57.1
99.9	83.3	54.7
74.0	72.8	78.6
74.4	83.5	53.7
112.8		96.0

Source: Data provided courtesy of Xiaodu Wang, Ph.D.

- 8.2.6.** In a study of 90 patients on renal dialysis, Farhad Atassi (A-7) assessed oral home care practices. He collected data from 30 subjects who were in (1) dialysis for less than 1 year, (2) dialysis for 1 to 3 years, and (3) dialysis for more than 3 years. The following table shows plaque index scores for these subjects. A higher score indicates a greater amount of plaque.

Group 1		Group 2		Group 3	
2.00	2.67	2.83	2.83	1.83	1.83
1.00	2.17	2.00	1.83	2.00	2.67
2.00	1.00	2.67	2.00	1.83	1.33
1.50	2.00	2.00	1.83	1.83	2.17
2.00	2.00	2.83	2.00	2.83	3.00
1.00	2.00	2.17	2.17	2.17	2.33

(Continued)

Group 1		Group 2		Group 3	
1.00	2.33	2.17	1.67	2.83	2.50
1.00	1.50	2.00	2.33	2.50	2.83
1.00	1.00	2.00	2.00	2.17	2.83
1.67	2.00	1.67	2.00	1.67	2.33
1.83	.83	2.33	2.17	2.17	2.33
2.17	.50	2.00	3.00	1.83	2.67
1.00	2.17	1.83	2.50	2.83	2.00
2.17	2.33	1.67	2.17	2.33	2.00
2.83	2.83	2.17	2.00	2.00	2.00

Source: Data provided courtesy of Farhad Atassi, DDS, MSC, FICOI.

- 8.2.7.** Thrombocytopenia is a condition of abnormally low platelets that often occurs during necrotizing enterocolitis (NEC)—a serious illness in infants that can cause tissue damage to the intestines. Ragazzi et al. (A-8) investigated differences in the \log_{10} of platelet counts in 178 infants with NEC. Patients were grouped into four categories of NEC status. Group 0 referred to infants with no gangrene, group 1 referred to subjects in whom gangrene was limited to a single intestinal segment, group 2 referred to patients with two or more intestinal segments of gangrene, and group 3 referred to patients with the majority of small and large bowel involved. The following table gives the \log_{10} platelet counts for these subjects.

Gangrene Grouping								
0			1		2		3	
1.97	2.33	2.48	1.38	2.45	1.87	2.37	1.77	
0.85	2.60	2.23	1.86	2.60	1.90	1.75	1.68	
1.79	1.88	2.51	2.26	1.83	2.43	2.57	1.46	
2.30	2.33	2.38	1.99	2.47	1.32	1.51	1.53	
1.71	2.48	2.31	1.32	1.92	2.06	1.08	1.36	
2.66	2.15	2.08	2.11	2.51	1.04	2.36	1.65	
2.49	1.41	2.49	2.54	1.79	1.99	1.58	2.12	
2.37	2.03	2.21	2.06	2.17	1.52	1.83	1.73	
1.81	2.59	2.45	2.41	2.18	1.99	2.55	1.91	
2.51	2.23	1.96	2.23	2.53	2.52	1.80	1.57	
2.38	1.61	2.29	2.00	1.98	1.93	2.44	2.27	
2.58	1.86	2.54	2.74	1.93	2.29	2.81	1.00	
2.58	2.33	2.23	2.00	2.42	1.75	2.17	1.81	
2.84	2.34	2.78	2.51	0.79	2.16	2.72	2.27	
2.55	1.38	2.36	2.08	1.38	1.81	2.44	2.43	
1.90	2.52	1.89			2.46	1.98	1.74	
2.28	2.35	2.26			1.66	1.57	1.60	
2.33	2.63	1.79			2.51	2.05	2.08	
1.77	2.03	1.87			1.76	2.30	2.34	
1.83	1.08	2.51			1.72	1.36	1.89	
1.67	2.40	2.29			2.57	2.48	1.75	
2.67	1.77	2.38			2.30	1.40	1.69	

(Continued)

Gangrene Grouping				
0	1	2		3
1.80		0.70	1.75	2.49
		2.16	2.67	1.75
		2.17	2.37	1.86
		2.12	1.46	1.26
		2.27	1.91	2.36

Source: Data provided courtesy of Simon Eaton, M.D.

- 8.2.8.** The objective of a study by Romita et al. (A-9) was to determine whether there is a different response to different calcium channel blockers. Two hundred and fifty patients with mild-to-moderate hypertension were randomly assigned to 4 weeks of treatment with once-daily doses of (1) lercanidipine, (2) felodipine, or (3) nifedipine. Prior to treatment and at the end of 4 weeks, each of the subjects had his or her systolic blood pressure measured. Researchers then calculated the change in systolic blood pressure. What is the treatment variable in this study? The response variable? What extraneous variables can you think of whose effects would be included in the error term? What are the “values” of the treatment variable? Construct an analysis of variance table in which you specify for this study the sources of variation and the degrees of freedom.
- 8.2.9.** Kosmiski et al. (A-10) conducted a study to examine body fat distributions of men infected and not infected with HIV, taking and not taking protease inhibitors (PI), and having been diagnosed and not diagnosed with lipodystrophy. Lipodystrophy is a syndrome associated with HIV/PI treatment that remains controversial. Generally, it refers to fat accumulation in the abdomen or viscera accompanied by insulin resistance, glucose intolerance, and dyslipidemia. In the study, 14 subjects were taking protease inhibitors and were diagnosed with lipodystrophy, 12 were taking protease inhibitors, but were not diagnosed with lipodystrophy, five were HIV positive, not taking protease inhibitors, nor had diagnosed lipodystrophy, and 43 subjects were HIV negative and not diagnosed with lipodystrophy. Each of the subjects underwent body composition and fat distribution analyses by dual-energy X-ray absorptiometry and computed tomography. Researchers were able to then examine the percent of body fat in the trunk. What is the treatment variable? The response variable? What are the “values” of the treatment variable? Who are the subjects? What extraneous variables can you think of whose effects would be included in the error term? What was the purpose of including HIV-negative men in the study? Construct an ANOVA table in which you specify the sources of variation and the degrees of freedom for each. The authors reported a computed V.R. of 11.79. What is the p value for the test?

8.3 THE RANDOMIZED COMPLETE BLOCK DESIGN

The *randomized complete block* design was developed about 1925 by R. A. Fisher, who was seeking methods of improving agricultural field experiments. The randomized complete block design is a design in which the units (called *experimental units*) to which the treatments are applied are subdivided into homogeneous groups called *blocks*, so that the number of experimental units in a block is equal to the number (or some multiple of the number) of treatments being studied. The treatments are then assigned at random to the experimental units within each block. It should be emphasized that each treatment appears in every block, and each block receives every treatment.

Objective The objective in using the randomized complete block design is to isolate and remove from the error term the variation attributable to the blocks, while assuring that treatment means will be free of block effects. The effectiveness of the design depends on the ability to achieve homogeneous blocks of experimental units. The ability to form homogeneous blocks depends on the researcher's knowledge of the experimental material. When blocking is used effectively, the error mean square in the ANOVA table will be reduced, the V.R. will be increased, and the chance of rejecting the null hypothesis will be improved.

In animal experiments, the breed of animal may be used as a blocking factor. Litters may also be used as blocks, in which case an animal from each litter receives a treatment. In experiments involving human beings, if it is desired that differences resulting from age be eliminated, then subjects may be grouped according to age so that one person of each age receives each treatment. The randomized complete block design also may be employed effectively when an experiment must be carried out in more than one laboratory (block) or when several days (blocks) are required for completion.

The random allocation of treatments to subjects is restricted in the randomized complete block design. That is, each treatment must be represented an equal number of times (one or more times) within each blocking unit. In practice this is generally accomplished by assigning a random permutation of the order of treatments to subjects within each block. For example, if there are four treatments representing three drugs and a placebo (drug A, drug B, drug C, and placebo [P]), then there are $4! = 24$ possible permutations of the four treatments: (A, B, C, P) or (A, C, B, P) or (C, A, P, B), and so on. One permutation is then randomly assigned to each block.

Advantages One of the advantages of the randomized complete block design is that it is easily understood. Furthermore, certain complications that may arise in the course of an experiment are easily handled when this design is employed.

It is instructive here to point out that the paired comparisons analysis presented in Chapter 7 is a special case of the randomized complete block design. Example 7.4.1, for example, may be treated as a randomized complete block design in which the two points in time (Pre-op and Post-op) are the treatments and the individuals on whom the measurements were taken are the blocks.

Data Display In general, the data from an experiment utilizing the randomized complete block design may be displayed in a table such as Table 8.3.1. The following new notation in this table should be observed:

$$\text{total of the } i\text{th block} = T_i = \sum_{j=1}^k x_{ij}$$

$$\text{mean of the } i\text{th block} = \bar{x}_i = \frac{\sum_{j=1}^k x_{ij}}{k} = \frac{T_i}{k}$$

$$\text{grand total} = T_{..} = \sum_{j=1}^k T_{.j} = \sum_{i=1}^n T_i$$

TABLE 8.3.1 Table of Sample Values for the Randomized Complete Block Design

Blocks	Treatments					Total	Mean
	1	2	3	...	k		
1	X_{11}	X_{12}	X_{13}	...	X_{1k}	$T_{1.}$	$\bar{X}_{1.}$
2	X_{21}	X_{22}	X_{23}	...	X_{2k}	$T_{2.}$	$\bar{X}_{2.}$
3	X_{31}	X_{32}	X_{33}	...	X_{3k}	$T_{3.}$	$\bar{X}_{3.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	X_{n1}	X_{n2}	X_{n3}	...	X_{nk}	$T_{n.}$	$\bar{X}_{n.}$
Total	$T_{.1}$	$T_{.2}$	$T_{.3}$...	$T_{.k}$	$T_{..}$	
Mean	$\bar{X}_{.1}$	$\bar{X}_{.2}$	$\bar{X}_{.3}$...	$\bar{X}_{.k}$		$\bar{X}_{..}$

indicating that the grand total may be obtained either by adding row totals or by adding column totals.

Two-Way ANOVA The technique for analyzing the data from a randomized complete block design is called *two-way analysis of variance* since an observation is categorized on the basis of two criteria—the block to which it belongs as well as the treatment group to which it belongs.

The steps for hypothesis testing when the randomized complete block design is used are as follows:

- 1. Data.** After identifying the treatments, the blocks, and the experimental units, the data, for convenience, may be displayed as in Table 8.3.1.
- 2. Assumptions.** The model for the randomized complete block design and its underlying assumptions are as follows:

The Model

$$x_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij} \quad (8.3.1)$$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, k$$

In this model

x_{ij} is a typical value from the overall population.

μ is an unknown constant.

β_i represents a block effect reflecting the fact that the experimental unit fell in the i th block.

τ_j represents a treatment effect, reflecting the fact that the experimental unit received the j th treatment.

ϵ_{ij} is a residual component representing all sources of variation other than treatments and blocks.

Assumptions of the Model

- (a) Each x_{ij} that is observed constitutes a random independent sample of size 1 from one of the kn populations represented.
- (b) Each of these kn populations is normally distributed with mean μ_{ij} and the same variance σ^2 . This implies that the ϵ_{ij} are independently and normally distributed with mean 0 and variance σ^2 .
- (c) The block and treatment effects are additive. This assumption may be interpreted to mean that there is no *interaction* between treatments and blocks. In other words, a particular block-treatment combination does not produce an effect that is greater or less than the sum of their individual effects. It can be shown that when this assumption is met,

$$\sum_{j=1}^k \tau_j = \sum_{i=1}^n \beta_i = 0$$

The consequences of a violation of this assumption are misleading results. One need not become concerned with the violation of the additivity assumption unless the largest mean is more than 50 percent greater than the smallest.

When these assumptions hold true, the τ_j and β_i are a set of fixed constants, and we have a situation that fits the fixed-effects model.

3. Hypotheses. We may test

$$H_0 : \tau_j = 0, \quad j = 1, 2, \dots, k$$

against the alternative

$$H_A : \text{not all } \tau_j = 0$$

A hypothesis test regarding block effects is not usually carried out under the assumptions of the fixed-effects model for two reasons. First, the primary interest is in treatment effects, the usual purpose of the blocks being to provide a means of eliminating an extraneous source of variation. Second, although the experimental units are randomly assigned to the treatments, the blocks are obtained in a nonrandom manner.

- 4. Test statistic.** The test statistic is V.R.
- 5. Distribution of test statistic.** When H_0 is true and the assumptions are met, V.R. follows an F distribution.
- 6. Decision rule.** Reject the null hypothesis if the computed value of the test statistic V.R. is equal to or greater than the critical value of F .
- 7. Calculation of test statistic.** It can be shown that the total sum of squares for the randomized complete block design can be partitioned into three components, one each attributable to blocks (SSBl), treatments (SSTr), and error (SSE). That is,

$$SST = SSBl + SSTr + SSE \quad (8.3.2)$$

The formulas for the quantities in Equation 8.3.2 are as follows:

$$SST = \sum_{j=1}^k \sum_{i=1}^n (x_{ij} - \bar{x}_{..})^2 \tag{8.3.3}$$

$$SSBl = \sum_{j=1}^k \sum_{i=1}^n (\bar{x}_{.j} - \bar{x}_{..})^2 \tag{8.3.4}$$

$$SSTr = \sum_{j=1}^k \sum_{i=1}^n (\bar{x}_{.j} - \bar{x}_{..})^2 \tag{8.3.5}$$

$$SSE = SST - SSBl - SSTr \tag{8.3.6}$$

The appropriate degrees of freedom for each component of Equation 8.3.2 are

$$\begin{matrix} \text{total} & \text{blocks} & \text{treatments} & \text{residual(error)} \\ kn - 1 & = (n - 1) & + (k - 1) & + (n - 1)(k - 1) \end{matrix}$$

The residual degrees of freedom, like the residual sum of squares, may be obtained by subtraction as follows:

$$\begin{aligned} (kn - 1) - (n - 1) - (k - 1) &= kn - 1 - n + 1 - k + 1 \\ &= n(k - 1) - 1(k - 1) = (n - 1)(k - 1) \end{aligned}$$

The ANOVA Table The results of the calculations for the randomized complete block design may be displayed in an ANOVA table such as Table 8.3.2.

8. Statistical decision. It can be shown that when the fixed-effects model applies and the null hypothesis of no treatment effects (all $\tau_i = 0$) is true, both the error, or residual, mean square and the treatments mean square are estimates of the common variance σ^2 . When the null hypothesis is true, therefore, the quantity

$$MSTr/MSE$$

is distributed as F with $k - 1$ numerator degrees of freedom and $(n - 1) \times (k - 1)$ denominator degrees of freedom. The computed variance ratio, therefore, is compared with the critical value of F .

TABLE 8.3.2 ANOVA Table for the Randomized Complete Block Design

Source	SS	d.f.	MS	V.R.
Treatments	$SSTr$	$(k - 1)$	$MSTr = SSTr / (k - 1)$	$MSTr / MSE$
Blocks	$SSBl$	$(n - 1)$	$MSBl = SSBl / (n - 1)$	
Residual	SSE	$(n - 1)(k - 1)$	$MSE = SSE / (n - 1)(k - 1)$	
Total	SST	$kn - 1$		

9. Conclusion. If we reject H_0 , we conclude that the alternative hypothesis is true. If we fail to reject H_0 , we conclude that H_0 may be true.

10. p value.

The following example illustrates the use of the randomized complete block design.

EXAMPLE 8.3.1

A physical therapist wished to compare three methods for teaching patients to use a certain prosthetic device. He felt that the rate of learning would be different for patients of different ages and wished to design an experiment in which the influence of age could be taken into account.

Solution: The randomized complete block design is the appropriate design for this physical therapist.

1. Data. Three patients in each of five age groups were selected to participate in the experiment, and one patient in each age group was randomly assigned to each of the teaching methods. The methods of instruction constitute our three treatments, and the five age groups are the blocks. The data shown in Table 8.3.3 were obtained.

2. Assumptions. We assume that each of the 15 observations constitutes a simple random sample of size 1 from one of the 15 populations defined by a block-treatment combination. For example, we assume that the number 7 in the table constitutes a randomly selected response from a population of responses that would result if a population of subjects under the age of 20 received teaching method A. We assume that the responses in the 15 represented populations are normally distributed with equal variances.

TABLE 8.3.3 Time (in Days) Required to Learn the Use of a Certain Prosthetic Device

Age Group	Teaching Method			Total	Mean
	A	B	C		
Under 20	7	9	10	26	8.67
20 to 29	8	9	10	27	9.00
30 to 39	9	9	12	30	10.00
40 to 49	10	9	12	31	10.33
50 and over	11	12	14	37	12.33
Total	45	48	58	151	
Mean	9.0	9.6	11.6		10.07

3. Hypotheses.

$$H_0 : \tau_j = 0 \quad j = 1, 2, 3$$

$$H_A : \text{not all } \tau_j = 0$$

4. **Test statistic.** The test statistic is $V.R. = MSTr/MSE$.
5. **Distribution of test statistic.** When H_0 is true and the assumptions are met, V.R. follows an F distribution with 2 and 8 degrees of freedom.
6. **Decision rule.** Let $\alpha = .05$. Reject the null hypothesis if the computed V.R. is equal to or greater than the critical F , which we find in Appendix Table G to be 4.46.
7. **Calculation of test statistic.** We compute the following sums of squares:

$$SST = (7 - 10.07)^2 + (8 - 10.07)^2 + \cdots + (14 - 10.07)^2 = 46.9335$$

$$SSBI = 3[(8.67 - 10.07)^2 + (9.00 - 10.07)^2 + \cdots + (12.33 - 10.07)^2] = 24.855$$

$$SSTr = 5[(9 - 10.07)^2 + (9.6 - 10.07)^2 + (11.6 - 10.07)^2] = 18.5335$$

$$SSE = 46.9335 - 24.855 - 18.5335 = 3.545$$

The degrees of freedom are total = $(3)(5) - 1 = 14$, blocks = $5 - 1 = 4$, treatments = $3 - 1 = 2$, and residual = $(5 - 1)(3 - 1) = 8$. The results of the calculations may be displayed in an ANOVA table as in Table 8.3.4

8. **Statistical decision.** Since our computed variance ratio, 20.91, is greater than 4.46, we reject the null hypothesis of no treatment effects on the assumption that such a large V.R. reflects the fact that the two sample mean squares are not estimating the same quantity. The only other explanation for this large V.R. would be that the null hypothesis is really true, and we have just observed an unusual set of results. We rule out the second explanation in favor of the first.

TABLE 8.3.4 ANOVA Table for Example 8.3.1

Source	SS	d.f.	MS	V.R.
Treatments	18.5335	2	9.26675	20.91
Blocks	24.855	4	6.21375	
Residual	3.545	8	.443125	
Total	46.9335	14		

- 9. Conclusion.** We conclude that not all treatment effects are equal to zero, or equivalently, that not all treatment means are equal.
- 10. p value.** For this test $p < .005$. ■

Computer Analysis Most statistical software packages will analyze data from a randomized complete block design. We illustrate the input and output for MINITAB. We use the data from the experiment to set up a MINITAB worksheet consisting of three columns. Column 1 contains the observations, Column 2 contains numbers that identify the block to which each observation belongs, and Column 3 contains numbers that identify the treatment to which each observation belongs. Figure 8.3.1 shows the MINITAB worksheet for Example 8.3.1. Figure 8.3.2 contains the MINITAB dialog box that initiates the analysis and the resulting ANOVA table.

The ANOVA table from the SAS[®] output for the analysis of Example 8.3.1 is shown in Figure 8.3.3. Note that in this output the model SS is equal to the sum of SSB and $SSTr$.

Alternatives When the data available for analysis do not meet the assumptions of the randomized complete block design as discussed here, the Friedman procedure discussed in Chapter 13 may prove to be a suitable nonparametric alternative.

ROW	C1	C2	C3
1	7	1	1
2	9	1	2
3	10	1	3
4	8	2	1
5	9	2	2
6	10	2	3
7	9	3	1
8	9	3	2
9	12	3	3
10	10	4	1
11	9	4	2
12	12	4	3
13	11	5	1
14	12	5	2
15	14	5	3

FIGURE 8.3.1 MINITAB worksheet for the data in Figure 8.3.2.

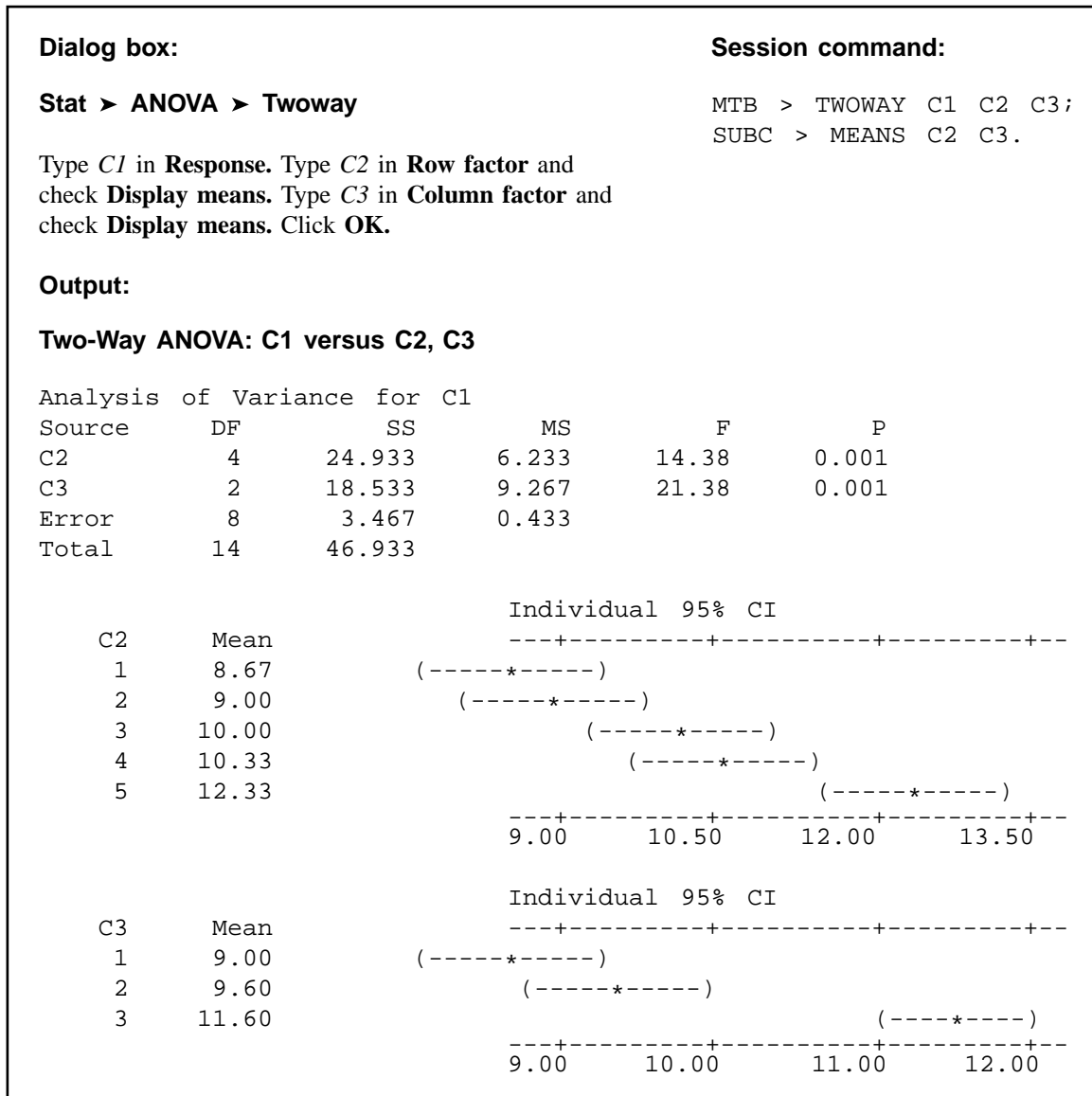


FIGURE 8.3.2 MINITAB dialog box and output for two-way analysis of variance, Example 8.3.1.

EXERCISES

For Exercise 8.3.1 to 8.3.5 perform the ten-step hypothesis testing procedure for analysis of variance.

- 8.3.1.** The objective of a study by Brooks et al. (A-11) was to evaluate the efficacy of using a virtual kitchen for vocational training of people with learning disabilities. Twenty-four students participated

The SAS System					
Analysis of Variance Procedure					
Dependent Variable: DAYS					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	43.46666667	7.244444444	16.72	0.0004
Error	8	3.46666667	0.43333333		
Corrected Total	14	46.93333333			
	R-Square	C.V.	Root MSE	DAYS Mean	
	0.926136	6.539211	0.65828059	10.06666667	
Source	DF	Anova SS	Mean Square	F Value	Pr > F
GROUP	2	18.53333333	9.26666667	21.38	0.0006
AGE	4	24.93333333	6.23333333	14.38	0.0010

FIGURE 8.3.3 Partial SAS® output for analysis of Example 8.3.1.

in the study. Each participant performed four food preparation tasks and they were scored on the quality of the preparation. Then each participant received regular vocational training in food preparation (real training), virtual training using a TV and computer screen of a typical kitchen, workbook training with specialized reading materials, and no training (to serve as a control). After each of these trainings, the subjects were tested on food preparation. Improvement scores for each of the four training methods are shown in the following table.

Subject No.	Real Training	Virtual Training	Workbook Training	No Training
1	2	10	2	-4
2	4	3	2	1
3	4	13	0	1
4	6	11	2	1
5	5	13	5	1
6	2	0	1	4
7	10	17	2	6
8	5	5	2	2
9	10	4	5	2
10	3	6	9	3
11	11	9	8	7
12	10	9	6	10
13	5	8	4	1

(Continued)

Subject No.	Real Training	Virtual Training	Workbook Training	No Training
14	8	11	1	1
15	4	8	5	2
16	11	8	10	2
17	6	11	1	3
18	2	5	1	2
19	3	1	0	-3
20	7	5	0	-6
21	7	10	4	4
22	8	7	-2	8
23	4	9	3	0
24	9	6	3	5

Source: Data provided courtesy of B. M. Brooks, Ph.D.

After eliminating subject effects, can we conclude that the improvement scores differ among methods of training? Let $\alpha = .05$.

- 8.3.2.** McConville et al. (A-12) report the effects of chewing one piece of nicotine gum (containing 2 mg nicotine) on tic frequency in patients whose Tourette's disorder was inadequately controlled by haloperidol. The following are the tic frequencies under four conditions:

Patient	Number of Tics During 30-Minute Period			
	Baseline	Gum Chewing	After End of Chewing	
			0-30 Minutes	30-60 Minutes
1	249	108	93	59
2	1095	593	600	861
3	83	27	32	61
4	569	363	342	312
5	368	141	167	180
6	326	134	144	158
7	324	126	312	260
8	95	41	63	71
9	413	365	282	321
10	332	293	525	455

Source: Data provided courtesy of Brian J. McConville, M. Harold Fogelson, Andrew B. Norman, William M. Klykylo, Pat Z. Manderscheid, Karen W. Parker, and Paul R. Sanberg. "Nicotine Potentiation of Haloperidol in Reducing Tic Frequency in Tourette's Disorder," *American Journal of Psychiatry*, 148 (1991), 793-794. Copyright © 1991, American Psychiatric Association.

After eliminating patient effects, can we conclude that the mean number of tics differs among the four conditions? Let $\alpha = .01$.

- 8.3.3.** A remotivation team in a psychiatric hospital conducted an experiment to compare five methods for remotivating patients. Patients were grouped according to level of initial motivation. Patients in each

group were randomly assigned to the five methods. At the end of the experimental period the patients were evaluated by a team composed of a psychiatrist, a psychologist, a nurse, and a social worker, none of whom was aware of the method to which patients had been assigned. The team assigned each patient a composite score as a measure of his or her level of motivation. The results were as follows:

Level of Initial Motivation	Remotivation Method				
	A	B	C	D	E
Nil	58	68	60	68	64
Very low	62	70	65	80	69
Low	67	78	68	81	70
Average	70	81	70	89	74

Do these data provide sufficient evidence to indicate a difference in mean scores among methods? Let $\alpha = .05$.

- 8.3.4. The nursing supervisor in a local health department wished to study the influence of time of day on length of home visits by the nursing staff. It was thought that individual differences among nurses might be large, so the nurse was used as a blocking factor. The nursing supervisor collected the following data:

Nurse	Length of Home Visit by Time of Day			
	Early Morning	Late Morning	Early Afternoon	Late Afternoon
A	27	28	30	23
B	31	30	27	20
C	35	38	34	30
D	20	18	20	14

Do these data provide sufficient evidence to indicate a difference in length of home visit among the different times of day? Let $\alpha = .05$.

- 8.3.5. Four subjects participated in an experiment to compare three methods of relieving stress. Each subject was placed in a stressful situation on three different occasions. Each time a different method for reducing stress was used with the subject. The response variable is the amount of decrease in stress level as measured before and after treatment application. The results were as follows:

Subject	Treatment		
	A	B	C
1	16	26	22
2	16	20	23
3	17	21	22
4	28	29	36

Can we conclude from these data that the three methods differ in effectiveness? Let $\alpha = .05$.

- 8.3.6.** In a study by Valencia et al. (A-13), the effects of environmental temperature and humidity on 24-hour energy expenditure were measured using whole-body indirect calorimetry in eight normal-weight young men who wore standardized light clothing and followed a controlled activity regimen. Temperature effects were assessed by measurements at 20, 23, 26, and 30 degrees Celsius at ambient humidity and at 20 and 30 degrees Celsius with high humidity. What is the blocking variable? The treatment variable? How many blocks are there? How many treatments? Construct an ANOVA table in which you specify the sources of variability and the degrees of freedom for each. What are the experimental units? What extraneous variables can you think of whose effects would be included in the error term?
- 8.3.7.** Hodgson et al. (A-14) conducted a study in which they induced gastric dilatation in six anesthetized dogs maintained with constant-dose isoflurane in oxygen. Cardiopulmonary measurements prior to stomach distension (baseline) were compared with measurements taken during .1, .5, 1.0, 1.5, 2.5, and 3.5 hours of stomach distension by analyzing the change from baseline. After distending the stomach, cardiac index increased from 1.5 to 3.5 hours. Stroke volume did not change. During inflation, increases were observed in systemic arterial, pulmonary arterial, and right atrial pressure. Respiratory frequency was unchanged. PaO₂ tended to decrease during gastric dilatation. What are the experimental units? The blocks? Treatment variable? Response variable(s)? Can you think of any extraneous variable whose effect would contribute to the error term? Construct an ANOVA table for this study in which you identify the sources of variability and specify the degrees of freedom.

8.4 THE REPEATED MEASURES DESIGN

One of the most frequently used experimental designs in the health sciences field is the repeated measures design.

DEFINITION

A repeated measures design is one in which measurements of the same variable are made on each subject on two or more different occasions.

The different occasions during which measurements are taken may be either points in time or different conditions such as different treatments.

When to Use Repeated Measures The usual motivation for using a repeated measures design is a desire to control for variability among subjects. In such a design each subject serves as its own control. When measurements are taken on only two occasions, we have the paired comparisons design that we discussed in Chapter 7. One of the most frequently encountered situations in which the repeated measures design is used is the situation in which the investigator is concerned with responses over time.

Advantages The major advantage of the repeated measures design is, as previously mentioned, its ability to control for extraneous variation among subjects. An additional advantage is the fact that fewer subjects are needed for the repeated measures design than

for a design in which different subjects are used for each occasion on which measurements are made. Suppose, for example, that we have four treatments (in the usual sense) or four points in time on each of which we would like to have 10 measurements. If a different sample of subjects is used for each of the four treatments or points in time, 40 subjects would be required. If we are able to take measurements on the same subject for each treatment or point in time—that is, if we can use a repeated measures design—only 10 subjects would be required. This can be a very attractive advantage if subjects are scarce or expensive to recruit.

Disadvantages A major potential problem to be on the alert for is what is known as the *carry-over effect*. When two or more treatments are being evaluated, the investigator should make sure that a subject's response to one treatment does not reflect a residual effect from previous treatments. This problem can frequently be solved by allowing a sufficient length of time between treatments.

Another possible problem is the *position effect*. A subject's response to a treatment experienced last in a sequence may be different from the response that would have occurred if the treatment had been first in the sequence. In certain studies, such as those involving physical participation on the part of the subjects, enthusiasm that is high at the beginning of the study may give way to boredom toward the end. A way around this problem is to randomize the sequence of treatments independently for each subject.

Single-Factor Repeated Measures Design The simplest repeated measures design is the one in which, in addition to the treatment variable, one additional variable is considered. The reason for introducing this additional variable is to measure and isolate its contribution to the total variability among the observations. We refer to this additional variable as a *factor*.

DEFINITION

The repeated measures design in which one additional factor is introduced into the experiment is called a *single-factor repeated measures design*.

We refer to the additional factor as *subjects*. In the single-factor repeated measures design, each subject receives each of the treatments. The order in which the subjects are exposed to the treatments, when possible, is random, and the randomization is carried out independently for each subject.

Assumptions The following are the assumptions of the single-factor repeated measures design that we consider in this text. A design in which these assumptions are met is called a *fixed-effects additive* design.

1. The subjects under study constitute a simple random sample from a population of similar subjects.
2. Each observation is an independent simple random sample of size 1 from each of kn populations, where n is the number of subjects and k is the number of treatments to which each subject is exposed.

3. The kn populations have potentially different means, but they all have the same variance.
4. The k treatments are fixed; that is, they are the only treatments about which we have an interest in the current situation. We do not wish to make inferences to some larger collection of treatments.
5. There is no interaction between treatments and subjects; that is, the treatment and subject effects are additive.

Experimenters may find frequently that their data do not conform to the assumptions of fixed treatments and/or additive treatment and subject effects. For such cases the references at the end of this chapter may be consulted for guidance.

In addition to the assumptions just listed, it should be noted that in a repeated-measures experiment there is a presumption that correlations should exist among the repeated measures. That is, measurements at time 1 and 2 are likely correlated, as are measurements at time 1 and 3, 2 and 3, and so on. This is expected because the measurements are taken on the same individuals through time.

An underlying assumption of the repeated-measures ANOVA design is that all of these correlations are the same, a condition referred to as *compound symmetry*. This assumption, coupled with assumption 3 concerning equal variances, is referred to as *sphericity*. Violations of the sphericity assumption can result in an inflated type I error. Most computer programs provide a formal test for the sphericity assumption along with alternative estimation methods if the sphericity assumption is violated.

The Model The model for the fixed-effects additive single-factor repeated measures design is

$$x_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij} \quad (8.4.1)$$

$$i = 1, 2, \dots, n; \quad j = 1, 2, \dots, k$$

The reader will recognize this model as the model for the randomized complete block design discussed in Section 8.3. The subjects are the blocks. Consequently, the notation, data display, and hypothesis testing procedure are the same as for the randomized complete block design as presented earlier. The following is an example of a repeated measures design.

EXAMPLE 8.4.1

Licciardone et al. (A-15) examined subjects with chronic, nonspecific low back pain. In this study, 18 of the subjects completed a survey questionnaire assessing physical functioning at baseline, and after 1, 3, and 6 months. Table 8.4.1 shows the data for these subjects who received a sham treatment that appeared to be genuine osteopathic manipulation. Higher values indicate better physical functioning. The goal of the experiment was to determine if subjects would report improvement over time even though the treatment they received would provide minimal improvement. We wish to know if there is a difference in the mean survey values among the four points in time.

TABLE 8.4.1 SF-36 Health Scores at Four Different Points in Time

Subject	Baseline	Month 1	Month 3	Month 6
1	80	60	95	100
2	95	90	95	95
3	65	55	50	45
4	50	45	70	70
5	60	75	80	85
6	70	70	75	70
7	80	80	85	80
8	70	60	75	65
9	80	80	70	65
10	65	30	45	60
11	60	70	95	80
12	50	50	70	60
13	50	65	80	65
14	85	45	85	80
15	50	65	90	70
16	15	30	20	25
17	10	15	55	75
18	80	85	90	70

Source: Data provided courtesy of John C. Licciardone.

Solution:

- Data.** See Table 8.4.1.
- Assumptions.** We assume that the assumptions for the fixed-effects, additive single-factor repeated measures design are met.
- Hypotheses.**

$$H_0: \mu_B = \mu_{M1} = \mu_{M3} = \mu_{M6}$$

$$H_A: \text{not all } \mu\text{'s are equal}$$

- Test statistic.** V.R. = treatment MS/error MS.
- Distribution of test statistic.** F with $4 - 1 = 3$ numerator degrees of freedom and $71 - 3 - 17 = 51$ denominator degrees of freedom.
- Decision rule.** Let $\alpha = .05$. The critical value of F is 2.80 (obtained by interpolation). Reject H_0 if computed V.R. is equal to or greater than 2.80.
- Calculation of test statistic.** We use MINITAB to perform the calculations. We first enter the measurements in Column 1, the row (subject) codes in Column 2, the treatment (time period) codes in Column 3, and proceed as shown in Figure 8.4.1.

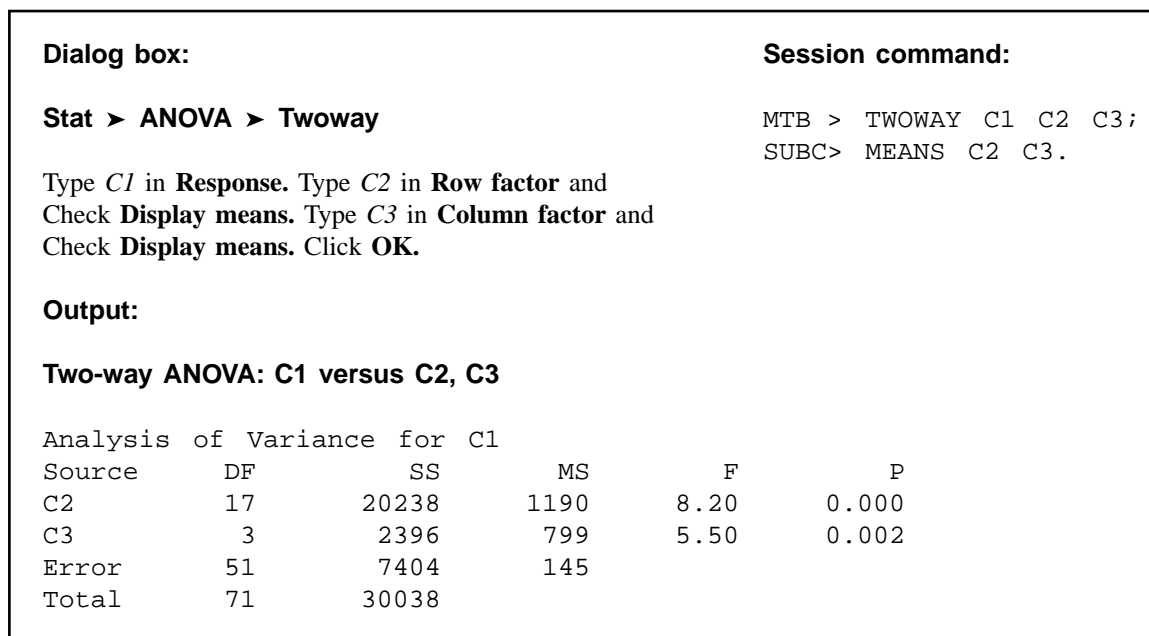


FIGURE 8.4.1 MINITAB procedure and output (ANOVA table) for Example 8.4.1.

8. **Statistical decision.** Since $V.R. = 5.50$ is greater than 2.80, we are able to reject the null hypothesis.
9. **Conclusion.** We conclude that there is a difference in the four population means.
10. **p value.** Since 5.50 is greater than 4.98, the F value for $\alpha = .005$ and $df = 40$, the p value is less than .005.

Figure 8.4.2, shows the SAS[®] output for the analysis of Example 8.4.1 and Figure 8.4.3 shows the SPSS output for the same example. Note that SPSS provides four potential tests. The first test is used under an assumption of sphericity and matches the outputs in Figures 8.4.1 and 8.4.2. The next three tests are modifications if the assumption of sphericity is violated. Note that SPSS modifies the degrees of freedom for these three tests, which changes the mean squares and the p values, but not the $V.R.$ Note that the assumption of sphericity was violated for these data, but that the decision rule did not change, since all of the p values were less than $\alpha = .05$. ■

Two-Factor Repeated Measures Design Repeated measures ANOVA is not useful just for testing means among different observation times. The analyses are easily expanded to include testing for differences among times for different treatment groups. As an example, a clinic may wish to test a placebo treatment against a new medication treatment. Researchers will randomly assign patients to one of the two treatment groups and will obtain measurements through time for each subject. In the end they are interested

The ANOVA Procedure

Dependent Variable: sf36

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	20	22633.33333	1131.66667	7.79	<.0001
Error	51	7404.16667	145.17974		
Corrected Total	71	30037.50000			

R-Square	Coeff Var	Root MSE	sf36 Mean
0.753503	18.18725	12.04906	66.25000

Source	DF	Anova SS	Mean Square	F Value	Pr > F
subj	17	20237.50000	1190.44118	8.20	<.0001
time	3	2395.83333	798.61111	5.50	0.0024

FIGURE 8.4.2 SAS[®] output for analysis of Example 8.4.1.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
factor 1	Sphericity Assumed	2395.833	3	798.611	5.501	.002
	Greenhouse-Geisser	2395.833	2.216	1080.998	5.501	.006
	Huynh-Feldt	2395.833	2.563	934.701	5.501	.004
	Lower-bound	2395.833	1.000	2395.833	5.501	.031
Error (factor 1)	Sphericity Assumed	7404.167	51	145.180		
	Greenhouse-Geisser	7404.167	37.677	196.515		
	Huynh-Feldt	7404.167	43.575	169.919		
	Lower-bound	7404.167	17.000	435.539		

FIGURE 8.4.3 SPSS output for the analysis of Example 8.4.1.

in knowing if there were differences between the two treatments on subjects that were measured multiple times.

Assumptions The assumptions of the two-factor repeated measures design are the same as the single-factor repeated measures design. However, it is not uncommon for there to be interactions among the treatments in this design, a potential violation of Assumption 5, above. Interaction effects can be interesting to examine, but are complex to calculate. For this reason, and at the level of the intended audience using this text, we will assume that interaction effects, when present, are mathematically handled using a statistical software package that provides correct calculations for this issue.

The Model The model for the two-factor repeated measures design must represent the fact that there are two factors, A and B , and they have a potential interaction. These features, along with the block effect and error, must be accounted for in the model, which is given by

$$x_{ijk} = \mu + \rho_{ij} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \quad (8.4.2)$$

$$i = 1, 2, \dots, a; \quad j = 1, 2, \dots, b; \quad k = 1, 2, \dots, n$$

In this model

x_{ijk} is a typical individual from the overall population

μ an unknown constant

ρ_{ij} represents a block effect

α_j represents the main effect of factor A

β_k represents the main effect of factor B

$(\alpha\beta)_{jk}$ represents the interaction effect of factor A and factor B

ϵ_{ijk} is a residual component representing all sources of variation other than treatments and blocks.

This model is very similar to the two-factor ANOVA model presented in Section 8.5.

EXAMPLE 8.4.2

The Mid-Michigan Medical Center (A-16) examined 25 subjects with neck cancer and measured as one of the outcome variables an oral health condition score. Patients were randomly divided into two treatment groups. These were a placebo treatment (treatment 1) and an aloe juice group (treatment 2). Cancer health was measured at baseline and at the end of 2, 4, and 6 weeks of treatment. The goal was to discern if there was any change in oral health condition over the course of the experiment and to see if there were any differences between the two treatment conditions.

Solution:

- 1. Data.** See Table 8.4.2.
- 2. Assumptions.** We assume that the assumptions for the two-factor repeated measures experiment are met.

TABLE 8.4.2 Oral Health Condition Scores at Four Different Points in Time Under Two Treatment Conditions

Subject	Treatment		TotalC1	TotalC2	TotalC3	TotalC4
	1 = placebo	2 = aloe juice				
1	1		6	6	6	7
2	1		9	6	10	9
3	1		7	9	17	19
4	1		6	7	9	3
5	1		6	7	16	13
6	1		6	6	6	11
7	1		6	11	11	10
8	1		6	11	15	15
9	1		6	9	6	8
10	1		6	4	8	7
11	1		7	8	11	11
12	1		6	6	9	6
13	1		8	8	9	10
14	1		7	16	9	10
15	2		6	10	11	9
16	2		4	6	8	7
17	2		6	11	11	14
18	2		6	7	6	6
19	2		12	11	12	9
20	2		5	7	13	12
21	2		6	7	7	7
22	2		8	11	16	16
23	2		5	7	7	7
24	2		6	8	16	16
25	2		7	8	10	8

Source: Mid-Michigan Medical Center, Midland, Michigan, 1999: A study of oral condition of cancer patients. Available in the public domain at: http://calcnet.mth.cmich.edu/org/spss/Prj_cancer_data.htm.

3. Hypotheses.

a. $H_0: \alpha_i = 0 \quad i = 1, 2, \dots, a$

H_a : not all $\alpha_i = 0$

b. $H_0: \beta_j = 0 \quad j = 1, 2, \dots, b$

H_a : not all $\beta_j = 0$

c. $H_0: (\alpha\beta)_{ij} = 0 \quad i = 1, 2, \dots, a; j = 1, 2, \dots, b$

H_a : not all $(\alpha\beta)_{ij} = 0$

4. **Test statistic.** The test statistic for each hypothesis set is V.R.
5. **Distribution of test statistics.** When H_0 is true and the assumptions are met, each of the test statistics is distributed as F . If all assumptions are met for the within-subjects effects, we will have F with $4 - 1 = 3$ numerator degrees of freedom for the time factor, $(4 - 1)(2 - 1) = 3$ numerator degrees of freedom for the interaction factor, and $(4 - 1)(25 - 2) = 69$ denominator degrees of freedom for both tests; interpolation from Table G provides a critical F value of 2.74. Further, for the between-subjects factor, we will have $(2 - 1) = 1$ numerator degrees of freedom and $25 - 2 = 23$ denominator degrees of freedom; Table G gives the critical F value to be 4.28. If we do not meet the assumptions, specifically of sphericity, then the computer program will alter the degrees of freedom and hence the critical value for comparisons.
6. **Decision rule.** Let $\alpha = .05$. Reject H_0 if the computed p value is less than α .
7. **Calculation of test statistic.** We use SPSS to perform the calculations. We enter the data just as it is shown in Table 8.4.2, though we do not need to enter the "Subject" number. The SPSS code and pertinent output are shown in Figure 8.4.4.
8. **Statistical decision.** SPSS provides a formal test for sphericity called "Mauchly's Test of Sphericity". Since we reject the null for this test according to the output in Figure 8.4.2, we will use the "Greenhouse-Geisser" test statistic. Since V.R. is greater than the critical value for TotalC, we reject the null hypothesis for this variable. However, both the critical values for the interaction effect and the between-subjects factor are quite small and less than the necessary critical value, and we therefore fail to reject these two null hypotheses.
9. **Conclusion.** We conclude that there is no statistical difference between treatments, but that subjects did have a change in oral condition through time regardless of the treatment they received.
10. **p value.** As seen in Figure 8.4.4, all p values are provided for each test. To summarize: since $p < .001$, we reject the null hypothesis concerning changes through time. Since $p = .931$, we fail to reject the null hypothesis concerning the interaction of time and treatment. Since $p = .815$, we fail to reject the null hypothesis concerning differences between treatments.

Though the output provided in Figure 8.4.2 can be valuable for statistical interpretation, it is often useful to examine plots to obtain a visual interpretation of the results. Figure 8.4.5 shows a plot of marginal means against time, with lines representing each of the treatments. It is evident that changes in oral condition did occur through time, but that the two treatments were very similar, as can be seen by the close proximity of the two curves. Further, it is evident that interaction between time and treatment occurred, as evidenced by the crossing of the plotted lines. ■

SPSS Code

```
GLM TOTALCIN T0TALCW2 TOTALCW4 T0TALCW6 BY TRT
```

```
/WSFACTOR=TotalC 4 Polynomial
```

```
/METHOD=SSTYPE(3)
```

```
/PLOT=PROFILE(TRT*TotalC*TRT)
```

```
/EMMEANS=TABLES(TRT)
```

```
/EMMEANS=TABLES(TotalC)
```

```
/EMMEANS=TABLES(TRT*TotalC)
```

```
/PRINT=DESCRIPTIVE
```

```
/CRITERIA=ALPHA(.05)
```

```
/WSDESIGN=TotalC
```

```
/DESIGN=TRT.
```

Partial SPSS Outout

Mauchly's Test of Sphericity^b

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.
TotalC	.487	15.620	5	.008

Tests of Within-Subjects Effects

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
TotalC	Sphericity Assumed	233.391	3	77.797	13.926	.000
	Greenhouse-Geisser	233.391	2.025	115.261	13.926	.000
	Huynh-Feldt	233.391	2.318	100.682	13.926	.000
	Lower-bound	233.391	1.000	233.391	13.926	.001
TotalC * TRT	Sphericity Assumed	1.231	3	.410	.073	.974
	Greenhouse-Geisser	1.231	2.025	.608	.073	.931
	Huynh-Feldt	1.231	2.318	.531	.073	.949
	Lower-bound	1.231	1.000	1.231	.073	.789
Error(TotalC)	Sphericity Assumed	385.469	69	5.587		
	Greenhouse-Geisser	385.469	46.572	8.277		
	Huynh-Feldt	385.469	53.316	7.230		
	Lower-bound	385.469	23.000	16.760		

Tests of Between-Subjects Effects

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	7637.274	1	7637.274	382.508	.000
TRT	1.114	1	1.114	.056	.815
Error	459.226	23	19.966		

FIGURE 8.4.4 SPSS code and partial output for Example 8.4.2.

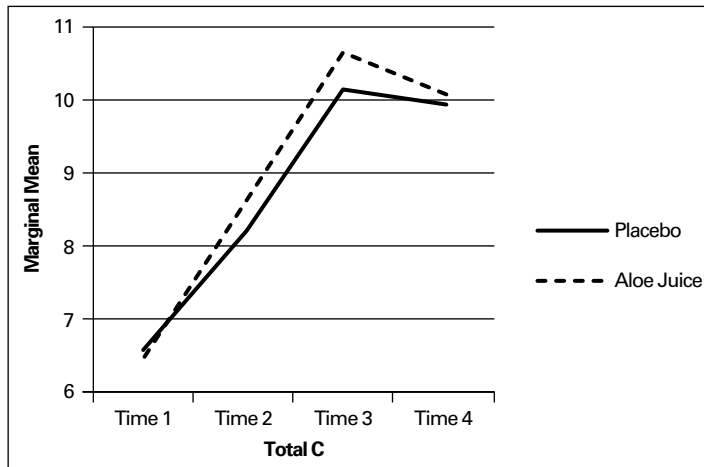


FIGURE 8.4.5 Excel plot of marginal means against total oral health score for the data of Example 8.4.2.

EXERCISES

For Exercises 8.4.1 to 8.4.3 perform the ten-step hypothesis testing procedure. Let $\alpha = .05$.

- 8.4.1.** One of the purposes of a study by Liu et al. (A-17) was to determine the effects of MRZ 2/579 on neurological deficit in Sprague-Dawley rats. In this study, 10 rats were measured at four time periods following occlusion of the middle carotid artery and subsequent treatment with the uncompetitive *N*-methyl-D-aspartate antagonist MRZ 2/579, which previous studies had suggested provides neuroprotective activity. The outcome variable was a neurological function variable measured on a scale of 0–12. A higher number indicates a higher degree of neurological impairment.

Rat	60 Minutes	24 Hours	48 Hours	72 Hours
1	11	9	8	4
2	11	7	5	3
3	11	10	8	6
4	11	4	3	2
5	11	10	9	9
6	11	6	5	5
7	11	6	6	6
8	11	7	6	5
9	11	7	5	5
10	11	9	7	7

Source: Data provided courtesy of Ludmila Belayev, M.D.

- 8.4.2.** Starch et al. (A-18) wanted to show the effectiveness of a central four-quadrant sleeve and screw in anterior cruciate ligament reconstruction. The researchers performed a series of reconstructions on eight cadaveric knees. The following table shows the loads (in newtons) required to achieve different graft laxities (mm) for seven specimens (data not available for one specimen) using five different load weights. Graft laxity is the separation (in mm) of the femur and the tibia at the points of graft fixation.

Is there sufficient evidence to conclude that different loads are required to produce different levels of graft laxity? Let $\alpha = .05$.

Specimen	Graft Laxity (mm)				
	1	2	3	4	5
1	297.1	297.1	297.1	297.1	297.1
2	264.4	304.6	336.4	358.2	379.3
3	188.8	188.8	188.8	188.8	188.8
4	159.3	194.7	211.4	222.4	228.1
5	228.2	282.1	282.1	334.8	334.8
6	100.3	105.0	106.3	107.7	108.7
7	116.9	140.6	182.4	209.7	215.4

Source: David W. Starch, Jerry W. Alexander, Philip C. Noble, Suraj Reddy, and David M. Lintner, "Multi-stranded Hamstring Tendon Graft Fixation with a Central Four-Quadrant or a Standard Tibial Interference Screw for Anterior Cruciate Ligament Reconstruction," *American Journal of Sports Medicine*, 31 (2003), 338–344.

- 8.4.3.** Holben et al. (A-19) designed a study to evaluate selenium intake in young women in the years of puberty. The researchers studied a cohort of 16 women for three consecutive summers. One of the outcome variables was the selenium intake per day. The researchers examined dietary journals of the subjects over the course of 2 weeks and then computed the average daily selenium intake. The following table shows the average daily selenium intake values (in $\mu\text{g}/\text{d}$) for the 16 women in years 1, 2, and 3 of the study.

Subject	Year 1	Year 2	Year 3	Subject	Year 1	Year 2	Year 3
1	112.51	121.28	94.99	9	95.05	93.89	73.26
2	106.20	121.14	145.69	10	112.65	100.47	145.69
3	102.00	121.14	130.37	11	103.74	121.14	123.97
4	103.74	90.21	135.91	12	103.74	121.14	135.91
5	103.17	121.14	145.69	13	112.67	104.66	136.87
6	112.65	98.11	145.69	14	106.20	121.14	126.42
7	106.20	121.14	136.43	15	103.74	121.14	136.43
8	83.57	102.87	144.35	16	106.20	100.47	135.91

Source: Data provided courtesy of David H. Holben, Ph.D. and John P. Holcomb, Ph.D.

- 8.4.4.** Linke et al. (A-20) studied seven male mongrel dogs. They induced diabetes by injecting the animals with alloxan monohydrate. The researchers measured the arterial glucose (mg/dl), arterial lactate (mmol/L), arterial free fatty acid concentration, and arterial β -hydroxybutyric acid concentration prior to the alloxan injection, and again in weeks 1, 2, 3, and 4 post-injection. What is the response variable(s)? Comment on carryover effect and position effect as they may or may not be of concern in this study. Construct an ANOVA table for this study in which you identify the sources of variability and specify the degrees of freedom for each.
- 8.4.5.** Werther et al. (A-21) examined the vascular endothelial growth factor (VEGF) concentration in blood from colon cancer patients. Research suggests that inhibiting VEGF may disrupt tumor growth. The researchers measured VEGF concentration (ng/L) for 10 subjects and found an upward trend in VEGF concentrations during the clotting time measured at baseline, and hours 1 and 2. What is the response variable? What is the treatment variable? Construct an ANOVA table for this study in which you identify the sources of variability and specify the degrees of freedom for each.

- 8.4.6. Yucha et al. (A-22) conducted a study to determine if nursing students who were assigned to a home hospital (HH) experience differed from those traditionally placed (TP) in hospitals throughout their nursing training. A small subset of data is provided in the table below. In this data set, hospital placement is the between-subjects variable. Anxiety, as measured by Spielberger's State Anxiety Scale (where higher scores suggest higher levels of anxiety), is the within-subjects variable and is provided at four points in time during nursing training. Is there evidence that anxiety level changed through time for these nursing students? Is there a difference in anxiety between those in a home hospital placement versus traditional placement? Is there significant interaction between placement type and anxiety? Let $\alpha = .05$.

Subject	Hospital Placement	Anxiety 1	Anxiety 2	Anxiety 3	Anxiety 4
1	HH	51	33	12	31
2	HH	50	51	50	44
3	HH	65	58	45	37
4	HH	43	40	31	51
5	HH	67	56	50	42
6	HH	46	69	62	46
7	HH	29	28	28	43
8	HH	76	69	62	60
9	HH	66	39	47	38
10	HH	56	46	34	31
11	TP	44	48	51	59
12	TP	44	50	54	40
13	TP	54	49	35	46
14	TP	38	38	32	37
15	TP	25	27	25	24
16	TP	61	60	55	66
17	TP	42	51	42	34
18	TP	36	49	49	51
19	TP	52	63	50	64
20	TP	41	55	56	34

Source: Data provided Courtesy of Carolyn B. Yucha, RN, PhD, FAAN.

8.5 THE FACTORIAL EXPERIMENT

In the experimental designs that we have considered up to this point, we have been interested in the effects of only one variable—the treatments. Frequently, however, we may be interested in studying, simultaneously, the effects of two or more variables. We refer to the variables in which we are interested as *factors*. The experiment in which two or more factors are investigated simultaneously is called a *factorial experiment*.

The different designated categories of the factors are called *levels*. Suppose, for example, that we are studying the effect on reaction time of three dosages of some drug. The drug factor, then, is said to occur at three levels. Suppose the second factor of interest in the study is age, and it is thought that two age groups, under 65 years and 65 years and older, should be included. We then have two levels of the age factor. In general, we say that factor *A* occurs at *a* levels and factor *B* occurs at *b* levels.

In a factorial experiment we may study not only the effects of individual factors but also, if the experiment is properly conducted, the interaction between factors. To illustrate the concept of interaction let us consider the following example.

EXAMPLE 8.5.1

Suppose, in terms of effect on reaction time, that the true relationship between three dosage levels of some drug and the age of human subjects taking the drug is known. Suppose further that age occurs at two levels—“young” (under 65) and “old” (65 and older). If the true relationship between the two factors is known, we will know, for the three dosage levels, the mean effect on reaction time of subjects in the two age groups. Let us assume that effect is measured in terms of reduction in reaction time to some stimulus. Suppose these means are as shown in Table 8.5.1.

The following important features of the data in Table 8.5.1 should be noted.

1. For both levels of factor *A* the difference between the means for any two levels of factor *B* is the same. That is, for both levels of factor *A*, the difference between means for levels 1 and 2 is 5, for levels 2 and 3 the difference is 10, and for levels 1 and 3 the difference is 15.
2. For all levels of factor *B* the difference between means for the two levels of factor *A* is the same. In the present case the difference is 5 at all three levels of factor *B*.
3. A third characteristic is revealed when the data are plotted as in Figure 8.5.1. We note that the curves corresponding to the different levels of a factor are all parallel.

When population data possess the three characteristics listed above, we say that there is no interaction present.

TABLE 8.5.1 Mean Reduction in Reaction Time (milliseconds) of Subjects in Two Age Groups at Three Drug Dosage Levels

Factor A—Age	Factor B—Drug Dosage		
	<i>j</i> = 1	<i>j</i> = 2	<i>j</i> = 3
Young (<i>i</i> = 1)	$\mu_{11} = 5$	$\mu_{12} = 10$	$\mu_{13} = 20$
Old (<i>i</i> = 2)	$\mu_{21} = 10$	$\mu_{22} = 15$	$\mu_{23} = 25$

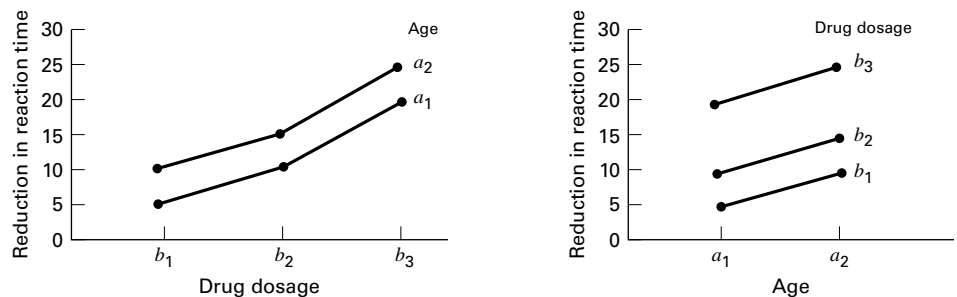
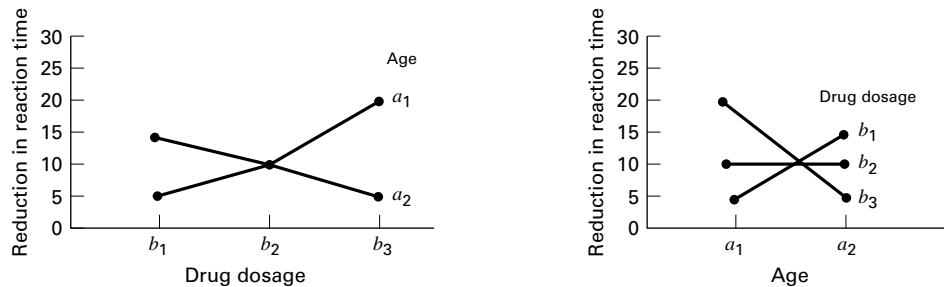


FIGURE 8.5.1 Age and drug effects, no interaction present.

TABLE 8.5.2 Data of Table 8.5.1 Altered to Show the Effect of One Type of Interaction

Factor A—Age	Factor B—Drug Dosage		
	$j = 1$	$j = 2$	$j = 3$
Young ($i = 1$)	$\mu_{11} = 5$	$\mu_{12} = 10$	$\mu_{13} = 20$
Old ($i = 2$)	$\mu_{21} = 15$	$\mu_{22} = 10$	$\mu_{23} = 5$

**FIGURE 8.5.2** Age and drug effects, interaction present.

The presence of interaction between two factors can affect the characteristics of the data in a variety of ways depending on the nature of the interaction. We illustrate the effect of one type of interaction by altering the data of Table 8.5.1 as shown in Table 8.5.2.

The important characteristics of the data in Table 8.5.2 are as follows.

1. The difference between means for any two levels of factor B is not the same for both levels of factor A . We note in Table 8.5.2, for example, that the difference between levels 1 and 2 of factor B is -5 for the young age group and $+5$ for the old age group.
2. The difference between means for both levels of factor A is not the same at all levels of factor B . The differences between factor A means are -10 , 0 , and 15 for levels 1, 2, and 3, respectively, of factor B .
3. The factor level curves are not parallel, as shown in Figure 8.5.2.

When population data exhibit the characteristics illustrated in Table 8.5.2 and Figure 8.5.2, we say that there is interaction between the two factors. We emphasize that the kind of interaction illustrated by the present example is only one of many types of interaction that may occur between two factors. ■

In summary, then, we can say that *there is interaction between two factors if a change in one of the factors produces a change in response at one level of the other factor different from that produced at other levels of this factor.*

Advantages The advantages of the factorial experiment include the following.

1. The interaction of the factors may be studied.
2. There is a saving of time and effort.

In the factorial experiment all the observations may be used to study the effects of each of the factors under investigation. The alternative, when two factors are being investigated, would be to conduct two different experiments, one to study each of the two factors. If this were done, some of the observations would yield information only on one of the factors, and the remainder would yield information only on the other factor. To achieve the level of accuracy of the factorial experiment, more experimental units would be needed if the factors were studied through two experiments. It is seen, then, that 1 two-factor experiment is more economical than 2 one-factor experiments.

3. Because the various factors are combined in one experiment, the results have a wider range of application.

The Two-Factor Completely Randomized Design A factorial arrangement may be studied with either of the designs that have been discussed. We illustrate the analysis of a factorial experiment by means of a two-factor completely randomized design.

1. **Data.** The results from a two-factor completely randomized design may be presented in tabular form as shown in Table 8.5.3.

Here we have a levels of factor A , b levels of factor B , and n observations for each combination of levels. Each of the ab combinations of levels of factor A with levels of factor B is a treatment. In addition to the totals and means shown in Table 8.5.3, we note that the total and mean of the ij th cell are

$$T_{ij\cdot} = \sum_{k=1}^n x_{ijk} \quad \text{and} \quad \bar{x}_{ij\cdot} = T_{ij\cdot}/n$$

TABLE 8.5.3 Table of Sample Data from a Two-Factor Completely Randomized Experiment

Factor A	Factor B				Totals	Means
	1	2	...	b		
1	x_{111}	x_{121}	...	x_{1b1}	$T_{1\cdot}$	$\bar{x}_{1\cdot}$
	\vdots	\vdots	\vdots	\vdots		
	x_{11n}	x_{12n}	...	x_{1bn}	$T_{2\cdot}$	$\bar{x}_{2\cdot}$
2	x_{211}	x_{221}	...	x_{2b1}		
	\vdots	\vdots	\vdots	\vdots	$T_{2\cdot}$	$\bar{x}_{2\cdot}$
	x_{21n}	x_{22n}	...	x_{2bn}		
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
a	x_{a11}	x_{a21}	...	x_{ab1}	$T_{a\cdot}$	$\bar{x}_{a\cdot}$
	\vdots	\vdots	\vdots	\vdots		
	x_{a1n}	x_{a2n}	...	x_{abn}	$T_{a\cdot}$	$\bar{x}_{a\cdot}$
Totals	$T_{\cdot 1}$	$T_{\cdot 2}$...	$T_{\cdot b}$		
Means	$\bar{x}_{\cdot 1}$	$\bar{x}_{\cdot 2}$...	$\bar{x}_{\cdot b}$		\bar{x}_{\dots}

respectively. The subscript i runs from 1 to a and j runs from 1 to b . The total number of observations is nab .

To show that Table 8.5.3 represents data from a completely randomized design, we consider that each combination of factor levels is a treatment and that we have n observations for each treatment. An alternative arrangement of the data would be obtained by listing the observations of each treatment in a separate column. Table 8.5.3 may also be used to display data from a two-factor randomized block design if we consider the first observation in each cell as belonging to block 1, the second observation in each cell as belonging to block 2, and so on to the n th observation in each cell, which may be considered as belonging to block n .

Note the similarity of the data display for the factorial experiment as shown in Table 8.5.3 to the randomized complete block data display of Table 8.3.1. The factorial experiment, in order that the experimenter may test for interaction, requires at least two observations per cell, whereas the randomized complete block design requires only one observation per cell. We use two-way analysis of variance to analyze the data from a factorial experiment of the type presented here.

- 2. Assumptions.** We assume a fixed-effects model and a two-factor completely randomized design. For a discussion of other designs, consult the references at the end of this chapter.

The Model The fixed-effects model for the two-factor completely randomized design may be written as

$$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \quad (8.5.1)$$

$$i = 1, 2, \dots, a; \quad j = 1, 2, \dots, b; \quad k = 1, 2, \dots, n$$

where x_{ijk} is a typical observation, μ is a constant, α_i represents an effect due to factor A , β_j represents an effect due to factor B , $(\alpha\beta)_{ij}$ represents an effect due to the interaction of factors A and B , and ϵ_{ijk} represents the experimental error.

Assumptions of the Model

- a. The observations in each of the ab cells constitute a random independent sample of size n drawn from the population defined by the particular combination of the levels of the two factors.
 - b. Each of the ab populations is normally distributed.
 - c. The populations all have the same variance.
- 3. Hypotheses.** The following hypotheses may be tested:
- a. $H_0: \alpha_i = 0 \quad i = 1, 2, \dots, a$
 $H_A: \text{not all } \alpha_i = 0$
 - b. $H_0: \beta_j = 0 \quad j = 1, 2, \dots, b$
 $H_A: \text{not all } \beta_j = 0$
 - c. $H_0: (\alpha\beta)_{ij} = 0 \quad i = 1, 2, \dots, a; j = 1, 2, \dots, b$
 $H_A: \text{not all } (\alpha\beta)_{ij} = 0$

Before collecting data, the researchers may decide to test only one of the possible hypotheses. In this case they select the hypothesis they wish to test, choose a significance level α , and proceed in the familiar, straightforward fashion. This procedure is free of the complications that arise if the researchers wish to test all three hypotheses.

When all three hypotheses are tested, the situation is complicated by the fact that the three tests are not independent in the probabilistic sense. If we let α be the significance level associated with the test as a whole, and α' , α'' , and α''' the significance levels associated with hypotheses 1, 2, and 3, respectively, we find

$$\alpha < 1 - (1 - \alpha')(1 - \alpha'')(1 - \alpha''') \quad (8.5.2)$$

If $\alpha' = \alpha'' = \alpha''' = .05$, then $\alpha < 1 - (.95)^3$, or $\alpha < .143$. This means that the probability of rejecting one or more of the three hypotheses is less than .143 when a significance level of .05 has been chosen for the hypotheses and all are true. To demonstrate the hypothesis testing procedure for each case, we perform all three tests. The reader, however, should be aware of the problem involved in interpreting the results.

4. **Test statistic.** The test statistic for each hypothesis set is V.R.
5. **Distribution of test statistic.** When H_0 is true and the assumptions are met, each of the test statistics is distributed as F .
6. **Decision rule.** Reject H_0 if the computed value of the test statistic is equal to or greater than the critical value of F .
7. **Calculation of test statistic.** By an adaptation of the procedure used in partitioning the total sum of squares for the completely randomized design, it can be shown that the total sum of squares under the present model can be partitioned into two parts as follows:

$$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{ij.} - \bar{x}_{...})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij.})^2 \quad (8.5.3)$$

or

$$SST = SSTr + SSE \quad (8.5.4)$$

The sum of squares for treatments can be partitioned into three parts as follows:

$$\begin{aligned} \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{ij.} - \bar{x}_{...})^2 &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{i..} - \bar{x}_{...})^2 \\ &+ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{j.} - \bar{x}_{...})^2 \\ &+ \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{j.} + \bar{x}_{...})^2 \end{aligned} \quad (8.5.5)$$

TABLE 8.5.4 Analysis of Variance Table for a Two-Factor Completely Randomized Experiment (Fixed-Effects Model)

Source	SS	d.f.	MS	V.R.
A	SSA	$a - 1$	$MSA = SSA/(a - 1)$	MSA/MSE
B	SSB	$b - 1$	$MSB = SSB/(b - 1)$	MSB/MSE
AB	SSAB	$(a - 1)(b - 1)$	$MSAB = SSAB/(a - 1)(b - 1)$	$MSAB/MSE$
Treatments	SSTr	$ab - 1$		
Residual	SSE	$ab(n - 1)$	$MSE = SSE/ab(n - 1)$	
Total	SST	$abn - 1$		

or

$$SSTr = SSA + SSB + SSAB$$

The ANOVA Table The results of the calculations for the fixed-effects model for a two-factor completely randomized experiment may, in general, be displayed as shown in Table 8.5.4.

8. **Statistical decision.** If the assumptions stated earlier hold true, and if each hypothesis is true, it can be shown that each of the variance ratios shown in Table 8.5.4 follows an F distribution with the indicated degrees of freedom. We reject H_0 if the computed V.R. values are equal to or greater than the corresponding critical values as determined by the degrees of freedom and the chosen significance levels.
9. **Conclusion.** If we reject H_0 , we conclude that H_A is true. If we fail to reject H_0 , we conclude that H_0 may be true.
10. **p value.**

EXAMPLE 8.5.2

In a study of length of time spent on individual home visits by public health nurses, data were reported on length of home visit, in minutes, by a sample of 80 nurses. A record was made also of each nurse's age and the type of illness of each patient visited. The researchers wished to obtain from their investigation answers to the following questions:

1. Does the mean length of home visit differ among different age groups of nurses?
2. Does the type of patient affect the mean length of home visit?
3. Is there interaction between nurse's age and type of patient?

Solution:

1. **Data.** The data on length of home visit that were obtained during the study are shown in Table 8.5.5.

TABLE 8.5.5 Length of Home Visit in Minutes by Public Health Nurses by Nurse's Age Group and Type of Patient

Factor A (Type of Patient) Levels	Factor B (Nurse's Age Group) Levels			
	1 (20 to 29)	2 (30 to 39)	3 (40 to 49)	4 (50 and Over)
1 (Cardiac)	20	25	24	28
	25	30	28	31
	22	29	24	26
	27	28	25	29
	21	30	30	32
2 (Cancer)	30	30	39	40
	45	29	42	45
	30	31	36	50
	35	30	42	45
	36	30	40	60
3 (C.V.A.)	31	32	41	42
	30	35	45	50
	40	30	40	40
	35	40	40	55
	30	30	35	45
4 (Tuberculosis)	20	23	24	29
	21	25	25	30
	20	28	30	28
	20	30	26	27
	19	31	23	30

2. Assumptions. To analyze these data, we assume a fixed-effects model and a two-factor completely randomized design.

3. Hypotheses. For our illustrative example we may test the following hypotheses subject to the conditions mentioned above.

- a. $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$ $H_A: \text{not all } \alpha_i = 0$
b. $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ $H_A: \text{not all } \beta_j = 0$
c. $H_0: \text{all } (\alpha\beta)_{ij} = 0$ $H_A: \text{not all } (\alpha\beta)_{ij} = 0$

Let $\alpha = .05$

4. Test statistic. The test statistic for each hypothesis set is V.R.

5. Distribution of test statistic. When H_0 is true and the assumptions are met, each of the test statistics is distributed as F .

6. **Decision rule.** Reject H_0 if the computed value of the test statistic is equal to or greater than the critical value of F . The critical values of F for testing the three hypotheses of our illustrative example are 2.76, 2.76, and 2.04, respectively. Since denominator degrees of freedom equal to 64 are not shown in Appendix Table G, 60 was used as the denominator degrees of freedom.
7. **Calculation of test statistic.** We use MINITAB to perform the calculations. We put the measurements in Column 1, the row (factor A) codes in Column 2, and the column (factor B) codes in Column 3. The resulting column contents are shown in Table 8.5.6. The MINITAB output is shown in Figure 8.5.3.

TABLE 8.5.6 Column Contents for MINITAB Calculations, Example 8.5.2

Row	C1	C2	C3	Row	C1	C2	C3
1	20	1	1	41	31	3	1
2	25	1	1	42	30	3	1
3	22	1	1	43	40	3	1
4	27	1	1	44	35	3	1
5	21	1	1	45	30	3	1
6	25	1	2	46	32	3	2
7	30	1	2	47	35	3	2
8	29	1	2	48	30	3	2
9	28	1	2	49	40	3	2
10	30	1	2	50	30	3	2
11	24	1	3	51	41	3	3
12	28	1	3	52	45	3	3
13	24	1	3	53	40	3	3
14	25	1	3	54	40	3	3
15	30	1	3	55	35	3	3
16	28	1	4	56	42	3	4
17	31	1	4	57	50	3	4
18	26	1	4	58	40	3	4
19	29	1	4	59	55	3	4
20	32	1	4	60	45	3	4
21	30	2	1	61	20	4	1
22	45	2	1	62	21	4	1
23	30	2	1	63	20	4	1
24	35	2	1	64	20	4	1
25	36	2	1	65	19	4	1
26	30	2	2	66	23	4	2
27	29	2	2	67	25	4	2
28	31	2	2	68	28	4	2

(Continued)

Row	C1	C2	C3	Row	C1	C2	C3
29	30	2	2	69	30	4	2
30	30	2	2	70	31	4	2
31	39	2	3	71	24	4	3
32	42	2	3	72	25	4	3
33	36	2	3	73	30	4	3
34	42	2	3	74	26	4	3
35	40	2	3	75	23	4	3
36	40	2	4	76	29	4	4
37	45	2	4	77	30	4	4
38	50	2	4	78	28	4	4
39	45	2	4	79	27	4	4
40	60	2	4	80	30	4	4

8. Statistical decision. The variance ratios are $V.R. (A) = 997.5/14.7 = 67.86$, $V.R. (B) = 400.4/14.7 = 27.24$, and $V.R. (AB) = 67.6/14.7 = 4.60$. Since the three computed values of VR. are all greater than the corresponding critical values, we reject all three null hypotheses.

9. Conclusion. When $H_0: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ is rejected, we conclude that there are differences among the levels of A , that is, differences in the average amount of time spent in home visits with different types of patients. Similarly, when $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$ is rejected, we conclude that there are differences among the levels of B , or differences in the average amount of time spent on home visits among the different nurses when grouped by age. When $H_0: (\alpha\beta)_{ij} = 0$ is rejected, we conclude that factors A and B interact; that is, different combinations of levels of the two factors produce different effects.

10. p value. Since 67.86, 27.24, and 4.60 are all greater than the critical values of $F_{.995}$ for the appropriate degrees of freedom, the p value for each of the tests is less than .005. When the hypothesis of no interaction is rejected, interest in the levels of factors A and B usually become subordinate to interest in the interaction effects. In other words, we are more interested in learning what combinations of levels are significantly different.

Figure 8.5.4 shows the SAS[®] output for the analysis of Example 8.5.2. ■

We have treated only the case where the number of observations in each cell is the same. When the number of observations per cell is not the same for every cell, the analysis becomes more complex.

In such cases the design is said to be unbalanced. To analyze these designs with MINITAB we use the general linear (GLM) procedure. Other software packages such as SAS[®] also will accommodate unequal cell sizes.

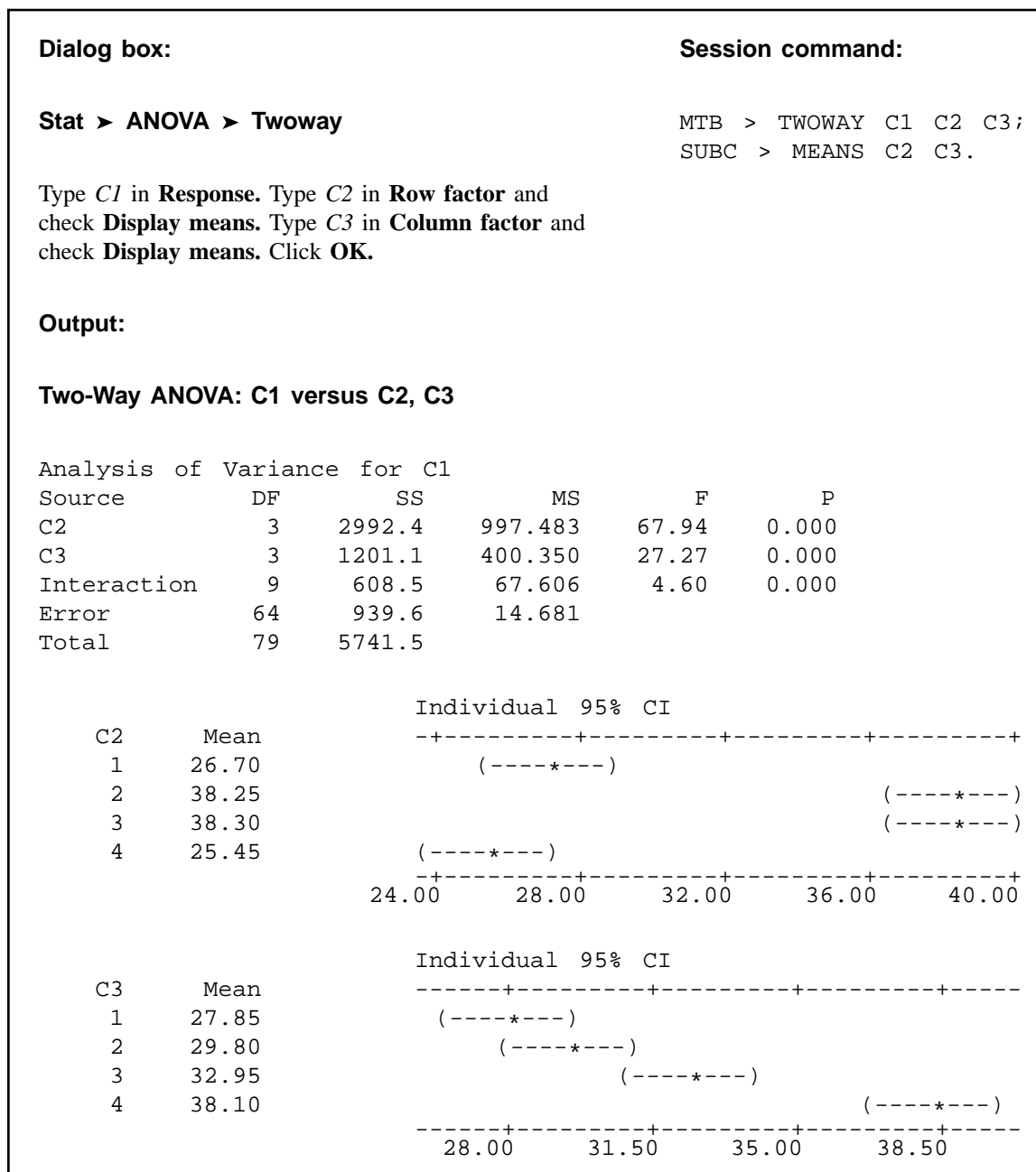


FIGURE 8.5.3 MINITAB procedure and ANOVA table for Example 8.5.2.

The SAS System					
Analysis of Variance Procedure					
Dependent Variable: TIME					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	4801.95000000	320.13000000	21.81	0.0001
Error	64	939.60000000	14.68125000		
Corrected Total	79	5741.55000000			
	R-Square	C.V.	Root MSE		TIME Mean
	0.836351	11.90866	3.83161193		32.17500000
Source	DF	Anova SS	Mean Square	F Value	Pr > F
FACTORB	3	1201.05000000	400.35000000	27.27	0.0001
FACTORA	3	2992.45000000	997.48333333	67.94	0.0001
FACTORB*FACTORA	9	608.45000000	67.60555556	4.60	0.0001

FIGURE 8.5.4 SAS[®] output for analysis of Example 8.5.2.

EXERCISES

For Exercises 8.5.1 to 8.5.4, perform the analysis of variance, test appropriate hypotheses at the .05 level of significance, and determine the p value associated with each test.

- 8.5.1.** Uryu et al. (A-23) studied the effect of three different doses of troglitazone (μM) on neuro cell death. Cell death caused by stroke partially results from the accumulation of high concentrations of glutamate. The researchers wanted to determine if different doses of troglitazone (1.3, 4.5, and $13.5\mu\text{M}$) and different ion forms (– and +) of LY294002, a PI3-kinase inhibitor, would give different levels of neuroprotection. Four rats were studied at each dose and ion level, and the measured variable is the percent of cell death as compared to glutamate. Therefore, a higher value implies less neuroprotection. The results are displayed in the table below.

Percent Compared to Glutamate	–LY294002 vs + LY294002	Troglitazone Dose (μM)
73.61	Negative	1.3
130.69	Negative	1.3
118.01	Negative	1.3
140.20	Negative	1.3

(Continued)

Percent Compared to Glutamate	-LY294002 vs + LY294002	Troglitazone Dose (μM)
97.11	Positive	1.3
114.26	Positive	1.3
120.26	Positive	1.3
92.39	Positive	1.3
26.95	Negative	4.5
53.23	Negative	4.5
59.57	Negative	4.5
53.23	Negative	4.5
28.51	Positive	4.5
30.65	Positive	4.5
44.37	Positive	4.5
36.23	Positive	4.5
-8.83	Negative	13.5
25.14	Negative	13.5
20.16	Negative	13.5
34.65	Negative	13.5
-35.80	Positive	13.5
-7.93	Positive	13.5
-19.08	Positive	13.5
5.36	Positive	13.5

Source: Data provided courtesy of Shigeo Uryu.

- 8.5.2.** Researchers at a trauma center wished to develop a program to help brain-damaged trauma victims regain an acceptable level of independence. An experiment involving 72 subjects with the same degree of brain damage was conducted. The objective was to compare different combinations of psychiatric treatment and physical therapy. Each subject was assigned to one of 24 different combinations of four types of psychiatric treatment and six physical therapy programs. There were three subjects in each combination. The response variable is the number of months elapsing between initiation of therapy and time at which the patient was able to function independently. The results were as follows:

Physical Therapy Program	Psychiatric Treatment			
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
I	11.0	9.4	12.5	13.2
	9.6	9.6	11.5	13.2
	10.8	9.6	10.5	13.5
II	10.5	10.8	10.5	15.0
	11.5	10.5	11.8	14.6
	12.0	10.5	11.5	14.0
III	12.0	11.5	11.8	12.8
	11.5	11.5	11.8	13.7
	11.8	12.3	12.3	13.1

(Continued)

Physical Therapy Program	Psychiatric Treatment			
	A	B	C	D
IV	11.5	9.4	13.7	14.0
	11.8	9.1	13.5	15.0
	10.5	10.8	12.5	14.0
V	11.0	11.2	14.4	13.0
	11.2	11.8	14.2	14.2
	10.0	10.2	13.5	13.7
VI	11.2	10.8	11.5	11.8
	10.8	11.5	10.2	12.8
	11.8	10.2	11.5	12.0

Can one conclude on the basis of these data that the different psychiatric treatment programs have different effects? Can one conclude that the physical therapy programs differ in effectiveness? Can one conclude that there is interaction between psychiatric treatment programs and physical therapy programs? Let $\alpha = .05$ for each test.

Exercises 8.5.3 and 8.5.4 are optional since they have unequal cell sizes. It is recommended that the data for these be analyzed using SAS[®] or some other software package that will accept unequal cell sizes.

- 8.5.3.** Main et al. (A-24) state, “Primary headache is a very common condition and one that nurses encounter in many different care settings. Yet, there is a lack of evidence as to whether advice given to sufferers is effective and what improvements may be expected in the conditions.” The researchers assessed frequency of headaches at the beginning and end of the study for 19 subjects in an intervention group (treatment 1) and 25 subjects in a control group (treatment 2). Subjects in the intervention group received health education from a nurse, while the control group did not receive education. In the 6 months between pre- and post-evaluation, the subjects kept a headache diary. The following table gives as the response variable the difference (pre – post) in frequency of headaches over the 6 months for two factors: (1) treatment with two levels (intervention and control), and (2) migraine status with two levels (migraine sufferer and nonmigraine sufferer).

Change in Frequency of Headaches	Migraine Sufferer (1 = No, 2 = Yes)		Change in Frequency of Headaches	Migraine Sufferer (1 = No, 2 = Yes)	
	Treatment	Treatment		Treatment	Treatment
-2	1	1	-3	2	2
2	2	1	-6	2	2
33	1	1	11	1	2
-6	2	1	64	1	2
6	2	1	65	1	2
98	1	1	14	1	2
2	2	1	8	1	2
6	2	1	6	2	2

(Continued)

Change in Frequency of Headaches	Migraine Sufferer (1 = No, 2 = Yes)	Treatment	Change in Frequency of Headaches	Migraine Sufferer (1 = No, 2 = Yes)	Treatment
33	1	1	14	1	2
-7	2	1	-11	2	2
-1	2	1	53	1	2
-12	2	1	26	2	2
12	1	1	3	1	2
64	1	1	15	1	2
36	2	1	3	1	2
6	2	1	41	1	2
4	2	1	16	1	2
11	2	1	-4	2	2
0	2	1	-6	2	2
			9	1	2
			9	2	2
			-3	2	2
			9	2	2
			3	1	2
			4	2	2

Source: Data provided courtesy of A. Main, H. Abu-Saad, R. Salt, I. Vlachonikolis, and A. Dowson, "Management by Nurses of Primary Headache: A Pilot Study," *Current Medical Research Opinion*, 18 (2002), 471-478.

Can one conclude on the basis of these data that there is a difference in the reduction of headache frequency between the control and treatment groups? Can one conclude that there is a difference in the reduction of headache frequency between migraine and non-migraine sufferers? Can one conclude that there is interaction between treatments and migraine status? Let $\alpha = .05$ for each test.

- 8.5.4.** The purpose of a study by Porcellini et al. (A-25) was to study the difference in CD4 cell response in patients taking highly active antiretroviral therapy (HAART, treatment 1) and patients taking HAART plus intermittent interleukin (IL-2, treatment 2). Another factor of interest was the HIV-RNA plasma count at baseline of study. Subjects were classified as having fewer than 50 copies/ml (plasma 1) or having 50 or more copies/ml (plasma 2). The outcome variable is the percent change in CD4 T cell count from baseline to 12 months of treatment. Can one conclude that there is a difference in the percent change in CD4 T cell count between the two treatments? The results are shown in the following table. Can one conclude that there is a difference in the percent change in CD4 T cell count between those who have fewer than 50/ml plasma copies of HIV-RNA and those who do not? Can one conclude that there is interaction between treatments and plasma levels? Let $\alpha = .05$ for each test.

Percent Change in CD4 T Cell	Treatment	Plasma
-12.60	1	1
-14.60	2	1
28.10	2	1

(Continued)

Percent Change in CD4 T Cell	Treatment	Plasma
77.30	1	1
-0.44	1	1
50.20	1	1
48.60	2	2
86.20	2	2
205.80	1	2
100.00	1	2
34.30	1	2
82.40	1	2
118.30	1	2

Source: Data provided courtesy of Simona Porcellini, Guiliana Vallanti, Silvia Nozza, Guido Poli, Adriano Lazzarin, Guisepppe Tabussi, and Antonio Grassia, "Improved Thymopoietic Potential in Aviremic HIV Infected Individuals with HAART by Intermittent IL-2 Administration," *AIDS*, 17 (2003), 1621-1630.

- 8.5.5.** A study by Górecka et al. (A-26) assessed the manner in which among middle-aged smokers the diagnosis of airflow limitation (AL) combined with advice to stop smoking influences the smoking cessation rate. Their concerns were whether having AL, whether the subject successfully quit smoking, and whether interaction between AL and smoking status were significant factors in regard to baseline variables and lung capacity variables at the end of the study. Some of the variables of interest were previous years of smoking (pack years), age at which subject first began smoking, forced expiratory volume in one second (FEV_1), and forced vital capacity (FVC). There were 368 subjects in the study. What are the factors in this study? At how many levels does each occur? Who are the subjects? What is (are) the response variable(s)? Can you think of any extraneous variables whose effects are included in the error term?
- 8.5.6.** A study by Meltzer et al. (A-27) examined the response to 5 mg desloratadine, an H1-receptor antagonist, in patients with seasonal allergies. During the fall allergy season, 172 subjects were randomly assigned to receive treatments of desloratadine and 172 were randomly assigned to receive a placebo. Subjects took the medication for 2 weeks after which changes in the nasal symptom score were calculated. A significant reduction was noticed in the treatment group compared to the placebo group, but gender was not a significant factor. What are the factors in the study? At how many levels does each occur? What is the response variable?

8.6 SUMMARY

The purpose of this chapter is to introduce the student to the basic ideas and techniques of analysis of variance. Two experimental designs, the completely randomized and the randomized complete block, are discussed in considerable detail. In addition, the concept of repeated measures designs and a factorial experiment as used with the completely randomized design are introduced. Individuals who wish to pursue further any aspect of analysis of variance will find the methodology references at the end of the chapter most helpful.

SUMMARY OF FORMULAS FOR CHAPTER 8

Formula Number	Name	Formula
8.2.1	One-way ANOVA model	$x_{ij} = \mu + \tau_j + \epsilon_{ij}$
8.2.2	Total sum-of-squares	$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{..})^2$
8.2.3	Within-group sum-of-squares	$SSW = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2$
8.2.4	Among-group sum-of-squares	$SSA = \sum_{j=1}^k n_j (\bar{x}_{.j} - \bar{x}_{..})^2$
8.2.5	Within-group variance	$MSW = \frac{\sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2}{\sum_{j=1}^k (n_j - 1)}$
8.2.6	Among-group variance I	$\sigma^2 = n\sigma_{\bar{x}}^2$
8.2.7	Among-group variance II (equal sample sizes)	$MSA = \frac{n \sum_{j=1}^k (\bar{x}_{.j} - \bar{x}_{..})^2}{k - 1}$
8.2.8	Among-group variance III (unequal sample sizes)	$MSA = \frac{\sum_{j=1}^k n_j (\bar{x}_{.j} - \bar{x}_{..})^2}{k - 1}$
8.2.9	Tukey's HSD (equal sample sizes)	$HSD = q_{\alpha, k, N-k} \sqrt{\frac{MSE}{n}}$
8.2.10	Tukey's HSD (unequal sample sizes)	$HSD^* = q_{\alpha, k, N-k} \sqrt{\frac{MSE}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$
8.3.1	Two-way ANOVA model	$x_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}$
8.3.2	Sum-of-squares representation	$SST = SSBl + SSTr + SSE$

(Continued)

8.3.3	Sum-of-squares total	$SST = \sum_{j=1}^k \sum_{i=1}^n (x_{ij} - \bar{x}_{..})^2$
8.3.4	Sum-of-squares block	$SSBl = \sum_{j=1}^k \sum_{i=1}^n (\bar{x}_{.j} - \bar{x}_{..})^2$
8.3.5	Sum-of-squares treatments	$SSTr = \sum_{j=1}^k \sum_{i=1}^n (\bar{x}_{.j} - \bar{x}_{..})^2$
8.3.6	Sum-of-squares error	$SSE = SST - SSBl - SSTr$
8.4.1	Fixed-effects, additive single-factor, repeated-measures ANOVA model	$x_{ij} = \mu + \beta_i + \tau_j + \epsilon_{ij}$
8.4.2	Two-factor repeated measures model	$x_{ijk} = \mu + \rho_{ij} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
8.5.1	Two-factor completely randomized fixed-effects factorial model	$x_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
8.5.2	Probabilistic representation of α	$\alpha < 1 - (1 - \alpha')(1 - \alpha'')(1 - \alpha''')$
8.5.3	Sum-of-squares total I	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (x_{ijk} - \bar{x}_{...})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{ij.} - \bar{x}_{...})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij.})^2$
8.5.4	Sum-of-squares total II	$SST = SSTr + SSE$
8.5.5	Sum-of-squares treatment partition	$\sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{ij.} - \bar{x}_{...})^2 = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{i..} - \bar{x}_{...})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{.j.} - \bar{x}_{...})^2 + \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...})^2$
Symbol Key	<ul style="list-style-type: none"> • α = Probability of Type I error • α_i = treatment A effect • β_j = treatment B effect • β_i = block effect • $(\alpha\beta)_{ij}$ = interaction effect • ϵ_{ij} = error term • HSD = honestly significant difference • k = number of treatments 	

- | | |
|--|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | <ul style="list-style-type: none"> • μ = mean of population (or the grand mean) • n = number of blocks • n_x = sample size • ρ_{ij} = block effect for two-factor repeated measures around • σ^2 = variance • SSX = sum of squares (where X : A = among, Bl = block, T = total, Tr = treatment, W = within) • τ_i = treatment effect • x_{xxx} = measurement |
|--|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

REVIEW QUESTIONS AND EXERCISES

1. Define analysis of variance.
2. Describe the completely randomized design.
3. Describe the randomized block design.
4. Describe the repeated measures design.
5. Describe the factorial experiment as used in the completely randomized design.
6. What is the purpose of Tukey's HSD test?
7. What is an experimental unit?
8. What is the objective of the randomized complete block design?
9. What is interaction?
10. What is a mean square?
11. What is an ANOVA table?
12. For each of the following designs describe a situation in your particular field of interest where the design would be an appropriate experimental design. Use real or realistic data and do the appropriate analysis of variance for each one:
 - (a) Completely randomized design
 - (b) Randomized complete block design
 - (c) Completely randomized design with a factorial experiment
 - (d) Repeated measures designs
13. Werther et al. (A-28) examined the β -leucocyte count ($\times 10^9/L$) in 51 subjects with colorectal cancer and 19 healthy controls. The cancer patients were also classified into Dukes's classification (A, B, C) for colorectal cancer that gives doctors a guide to the risk, following surgery, of the cancer coming back or spreading to other parts of the body. An additional category (D) identified patients with disease that had not been completely resected. The results are displayed in the following table. Perform an analysis of these data in which you identify the sources of variability and specify the degrees of freedom for each. Do these data provide sufficient evidence to indicate that, on the average, leucocyte counts differ among the five categories? Let $\alpha = .01$ and find the p value. Use Tukey's procedure to test for significant differences between individual pairs of sample means.

Healthy	A	B	C	D
6.0	7.7	10.4	8.0	9.5
6.3	7.8	5.6	6.7	7.8
5.1	6.1	7.0	9.3	5.7
6.2	9.6	8.2	6.6	8.0
10.4	5.5	9.0	9.3	9.6
4.4	5.8	8.4	7.2	13.7
7.4	4.0	8.1	5.2	6.3
7.0	5.4	8.0	9.8	7.3
5.6		6.5	6.2	6.2
5.3		9.1	10.1	
2.6		11.0	9.3	
6.3		10.9	9.4	
6.1		10.6	6.5	
5.3		5.2	5.4	
5.4		7.9	7.6	
5.2		7.6	9.2	
4.3		5.8		
4.9		7.0		
7.3				
4.9				
6.9				
4.3				
5.6				
5.1				

Source: Data provided courtesy of Kim Werther, M.D., Ph.D.

14. In Example 8.4.1, we examined data from a study by Licciardone et al. (A-15) on osteopathic manipulation as a treatment for chronic back pain. At the beginning of that study, there were actually 91 subjects randomly assigned to one of three treatments: osteopathic manipulative treatment (OMT), sham manipulation (SHAM), or non-intervention (CONTROL). One important outcome variable was the rating of back pain at the beginning of the study. The researchers wanted to know if the treatment had essentially the same mean pain level at the start of the trial. The results are displayed in the following table. The researchers used a visual analog scale from 0 to 10 cm where 10 indicated “worst pain possible.” Can we conclude, on the basis of these data, that, on the average, pain levels differ in the three treatment groups? Let $\alpha = .05$ and find the p value. If warranted, use Tukey’s procedure to test for differences between individual pairs of sample means.

CONTROL	SHAM	OMT	
2.6	5.8	7.8	3.5
5.6	1.3	4.1	3.4
3.3	2.4	1.7	1.1
4.6	1.0	3.3	0.5
8.4	3.2	4.3	5.1
0.0	0.4	6.5	1.9
2.5	5.4	5.4	2.0

(Continued)

CONTROL	SHAM	OMT	
5.0	4.5	4.0	2.8
1.7	1.5	4.1	3.7
3.8	0.0	2.6	1.6
2.4	0.6	3.2	0.0
1.1	0.0	2.8	0.2
0.7	7.6	3.4	7.3
2.4	3.5	6.7	1.7
3.3	3.9	7.3	7.5
6.6	7.0	2.1	1.6
0.4	7.4	3.7	3.0
0.4	6.5	2.3	6.5
0.9	1.6	4.4	3.0
6.0	1.3	2.8	3.3
	6.6	0.4	7.3
	6.3	0.7	4.6
	7.0	7.9	4.8
		1.3	4.9

Source: Data provided courtesy of J. C. Licciardone, D.O.

15. The goal of a study conducted by Meshack and Norman (A-29) was to evaluate the effects of weights on postural hand tremor related to self-feeding in subjects with Parkinson's disease (PD). Each of the 16 subjects had the tremor amplitude measured (in mm) under three conditions: holding a built-up spoon (108 grams), holding a weighted spoon (248 grams), and holding the built-up spoon while wearing a weighted wrist cuff (470 grams). The data are displayed in the following table.

Tremor Amplitude (mm)			
Subject	Built-Up Spoon	Weighted Spoon	Built-Up Spoon + Wrist Cuff
1	.77	1.63	1.02
2	.78	.88	1.11
3	.17	.14	.14
4	.30	.27	.26
5	.29	.27	.28
6	1.60	1.49	1.73
7	.38	.39	.37
8	.24	.24	.24
9	.17	.17	.16
10	.38	.29	.27
11	.93	1.21	.90
12	.63	.52	.66
13	.49	.73	.76
14	.42	.60	.29
15	.19	.21	.21
16	.19	.20	.16

Source: Rubia P. Meshack and Kathleen E. Norman, "A Randomized Controlled Trial of the Effects of Weights on Amplitude and Frequency of Postural Hand Tremor in People with Parkinson's Disease," *Clinical Rehabilitation*, 16 (2003), 481-492.

Can one conclude on the basis of these data that the three experimental conditions, on the average, have different effects on tremor amplitude? Let $\alpha = .05$. Determine the p value.

16. In a study of pulmonary effects on guinea pigs, Lacroix et al. (A-30) exposed 18 ovalbumin-sensitized guinea pigs and 18 nonsensitized guinea pigs to regular air, benzaldehyde, and acetaldehyde. At the end of exposure, the guinea pigs were anesthetized and allergic responses were assessed in bronchoalveolar lavage (BAL). The following table shows the alveolar cell count ($\times 10^6$) by treatment group for the ovalbumin-sensitized and nonsensitized guinea pigs.

Ovalbumin-Sensitized	Treatment	Alveolar Count $\times 10^6$
no	acetaldehyde	49.90
no	acetaldehyde	50.60
no	acetaldehyde	50.35
no	acetaldehyde	44.10
no	acetaldehyde	36.30
no	acetaldehyde	39.15
no	air	24.15
no	air	24.60
no	air	22.55
no	air	25.10
no	air	22.65
no	air	26.85
no	benzaldehyde	31.10
no	benzaldehyde	18.30
no	benzaldehyde	19.35
no	benzaldehyde	15.40
no	benzaldehyde	27.10
no	benzaldehyde	21.90
yes	acetaldehyde	90.30
yes	acetaldehyde	72.95
yes	acetaldehyde	138.60
yes	acetaldehyde	80.05
yes	acetaldehyde	69.25
yes	acetaldehyde	31.70
yes	air	40.20
yes	air	63.20
yes	air	59.10
yes	air	79.60
yes	air	102.45
yes	air	64.60
yes	benzaldehyde	22.15
yes	benzaldehyde	22.75
yes	benzaldehyde	22.15
yes	benzaldehyde	37.85
yes	benzaldehyde	19.35
yes	benzaldehyde	66.70

Source: Data provided courtesy of G. Lacroix, Docteur en Toxicologie.

Test for differences (a) between ovalbumin-sensitized and nonsensitized outcomes, (b) among the three different exposures, and (c) interaction. Let $\alpha = .05$ for all tests.

17. Watanabe et al. (A-31) studied 52 healthy middle-aged male workers. The researchers used the Masstricht Vital Exhaustion Questionnaire to assess vital exhaustion. Based on the resultant scores, they assigned subjects into three groups: VE1, VE2, and VE3. VE1 indicates the fewest signs of exhaustion, and VE3 indicates the most signs of exhaustion. The researchers also asked subjects about their smoking habits. Smoking status was categorized as follows: SMOKE1 are nonsmokers, SMOKE2 are light smokers (20 cigarettes or fewer per day), SMOKE3 are heavy smokers (more than 20 cigarettes per day). One of the outcome variables of interest was the amplitude of the high-frequency spectral analysis of heart rate variability observed during an annual health checkup. This variable, HF-amplitude, was used as an index of parasympathetic nervous function. The data are summarized in the following table:

Vital Exhaustion Group	HF-Amplitude			
	Smoking Status			
	SMOKE1		SMOKE2	SMOKE3
VE1	23.33	13.37	16.14	16.83
	31.82	9.76	20.80	29.40
	10.61	22.24	15.44	6.50
	42.59	8.77	13.73	10.18
	23.15	20.28		13.86
	17.29			
VE2	20.69	11.67	44.92	27.91
	16.21	30.17		36.89
	28.49	29.20		16.80
	25.67	8.73		17.08
	15.29	9.08		18.77
	7.51	22.53		18.33
	22.03	17.19		
	10.27			
VE3	9.44		17.59	5.57
	19.16		18.90	13.51
	14.46			17.37
				10.63
			13.83	

Source: Data provided courtesy of Takemasa Watanabe, M.D., Ph.D.

Perform an analysis of variance on these data and test the three possible hypotheses. Let $\alpha' = \alpha'' = \alpha''' = .05$. Determine the p values.

18. The effects of thermal pollution on *Corbicula fluminea* (Asiatic clams) at three different geographical locations were analyzed by John Brooker (A-32). Sample data on clam shell length, width, and height are displayed in the following table. Determine if there is a significant difference in mean length, height, or width (measured in mm) of the clam shell at the three different locations by performing three analyses. What inferences can be made from your results? What are the assumptions underlying your inferences? What are the target populations?

Location 1			Location 2			Location 3		
Length	Width	Height	Length	Width	Height	Length	Width	Height
7.20	6.10	4.45	7.25	6.25	4.65	5.95	4.75	3.20
7.50	5.90	4.65	7.23	5.99	4.20	7.60	6.45	4.56
6.89	5.45	4.00	6.85	5.61	4.01	6.15	5.05	3.50
6.95	5.76	4.02	7.07	5.91	4.31	7.00	5.80	4.30
6.73	5.36	3.90	6.55	5.30	3.95	6.81	5.61	4.22
7.25	5.84	4.40	7.43	6.10	4.60	7.10	5.75	4.10
7.20	5.83	4.19	7.30	5.95	4.29	6.85	5.55	3.89
6.85	5.75	3.95	6.90	5.80	4.33	6.68	5.50	3.90
7.52	6.27	4.60	7.10	5.81	4.26	5.51	4.52	2.70
7.01	5.65	4.20	6.95	5.65	4.31	6.85	5.53	4.00
6.65	5.55	4.10	7.39	6.04	4.50	7.10	5.80	4.45
7.55	6.25	4.72	6.54	5.89	3.65	6.81	5.45	3.51
7.14	5.65	4.26	6.39	5.00	3.72	7.30	6.00	4.31
7.45	6.05	4.85	6.08	4.80	3.51	7.05	6.25	4.71
7.24	5.73	4.29	6.30	5.05	3.69	6.75	5.65	4.00
7.75	6.35	4.85	6.35	5.10	3.73	6.75	5.57	4.06
6.85	6.05	4.50	7.34	6.45	4.55	7.35	6.21	4.29
6.50	5.30	3.73	6.70	5.51	3.89	6.22	5.11	3.35
6.64	5.36	3.99	7.08	5.81	4.34	6.80	5.81	4.50
7.19	5.85	4.05	7.09	5.95	4.39	6.29	4.95	3.69
7.15	6.30	4.55	7.40	6.25	4.85	7.55	5.93	4.55
7.21	6.12	4.37	6.00	4.75	3.37	7.45	6.19	4.70
7.15	6.20	4.36	6.94	5.63	4.09	6.70	5.55	4.00
7.30	6.15	4.65				7.51	6.20	4.74
6.35	5.25	3.75				6.95	5.69	4.29
						7.50	6.20	4.65

Source: Data provided courtesy of John Brooker, M.S. and the Wright State University Statistical Consulting Center.

19. Eleftherios Kellis (A-33) conducted an experiment on 18 pubertal males. He recorded the electromyographic (EMG) activity at nine different angular positions of the biceps femoris muscle. The EMG values are expressed as a percent (0–100 percent) of the maximal effort exerted with the muscle and represent an average in a range of flexion angles. The nine positions correspond to testing knee flexion angles of 1–10°, 11–20°, 21–30°, 31–40°, 41–50°, 51–60°, 61–70°, 71–80°, and 81–90°. The results are displayed in the following table. For subject 1, for example, the value of 30.96 percent represents the average maximal percent of effort in angular positions from 1 to 10 degrees.

Subject	1-10°	11-20°	21-30°	31-40°	41-50°	51-60°	61-70°	71-80°	81-90°
1	30.96	11.32	4.34	5.99	8.43	10.50	4.49	10.93	33.26
2	3.61	1.47	3.50	10.25	3.30	3.62	10.14	11.05	8.78
3	8.46	2.94	1.83	5.80	11.59	15.17	13.04	10.57	8.22
4	0.69	1.06	1.39	1.08	0.96	2.52	2.90	3.27	5.52
5	4.40	3.02	3.74	3.83	3.73	10.16	9.31	12.70	11.45
6	4.59	9.80	10.71	11.64	9.78	6.91	8.53	8.30	11.75
7	3.31	3.31	4.12	12.56	4.60	1.88	2.42	2.46	2.19
8	1.98	6.49	2.61	3.28	10.29	7.56	16.68	14.52	13.49
9	10.43	4.96	12.37	24.32	17.16	34.71	35.30	37.03	45.65
10	20.91	20.72	12.70	15.06	12.03	11.31	28.47	26.81	25.08
11	5.59	3.13	2.83	4.31	6.37	13.95	13.48	11.15	30.97
12	8.67	4.32	2.29	6.20	13.01	19.30	9.33	12.30	12.20
13	2.11	1.59	2.40	2.56	2.83	2.55	5.84	5.23	8.84
14	3.82	5.04	6.81	10.74	10.10	13.14	19.39	13.31	12.02
15	39.51	62.34	70.46	20.48	17.38	54.04	25.76	50.32	46.84
16	3.31	4.95	12.49	9.18	14.00	16.17	25.75	11.82	13.17
17	11.42	7.53	4.65	4.70	7.57	9.86	5.30	4.47	3.99
18	2.97	2.18	2.36	4.61	7.83	17.49	42.55	61.84	39.70

Source: Data provided courtesy of Eleftherios Kellis, Ph.D.

Can we conclude on the basis of these data that the average EMG values differ among the nine angular locations? Let $\alpha = .05$.

20. In a study of Marfan syndrome, Pyeritz et al.(A-34) reported the following severity scores of patients with no, mild, and marked dural ectasia. May we conclude, on the basis of these data, that mean severity scores differ among the three populations represented in the study? Let $\alpha = .05$ and find the p value. Use Tukey's procedure to test for significant differences among individual pairs of sample means.

No dural ectasia: 18, 18, 20, 21, 23, 23, 24, 26, 26, 27, 28, 29, 29, 29, 30, 30, 30, 30, 32, 34, 34, 38

Mild dural ectasia: 10, 16, 22, 22, 23, 26, 28, 28, 28, 29, 29, 30, 31, 32, 32, 33, 33, 38, 39, 40, 47

Marked dural ectasia: 17, 24, 26, 27, 29, 30, 30, 33, 34, 35, 35, 36, 39

Source: Data provided courtesy of Reed E. Pyeritz, M.D., Ph.D.

21. The following table shows the arterial plasma epinephrine concentrations (nanograms per milliliter) found in 10 laboratory animals during three types of anesthesia:

Anesthesia	Animal									
	1	2	3	4	5	6	7	8	9	10
A	.28	.50	.68	.27	.31	.99	.26	.35	.38	.34
B	.20	.38	.50	.29	.38	.62	.42	.87	.37	.43
C	1.23	1.34	.55	1.06	.48	.68	1.12	1.52	.27	.35

Can we conclude from these data that the three types of anesthesia, on the average, have different effects? Let $\alpha = .05$.

22. The aim of a study by Hartman-Maeir et al. (A-35) was to evaluate the awareness of deficit profiles among stroke patients undergoing rehabilitation. She studied 35 patients with a stroke lesion in the right hemisphere and 19 patients with a lesion on the left hemisphere. She also grouped lesion size as

2 = “1-3 cm”, 3 = “3-5 cm”, and 4 = “5 cm or greater”

One of the outcome variables was a measure of each patient’s total unawareness of their own limitations. Scores ranged from 8 to 24, with higher scores indicating more unawareness.

Lesion Size Group	Unawareness Score		
	Left Hemisphere	Right Hemisphere	
2	11	10	8
	13	11	10
	10	13	9
	11	10	9
	9	13	9
	10	10	
	9	10	
	8	9	
	10	8	
3	13	11	10
	8	10	11
	10	10	12
	10	14	11
	10	8	
4	11	10	11
	13	13	9
	14	10	19
	13	10	10
	14	15	9
	8	10	

Source: Data provided courtesy of Adina Hartman-Maeir, Ph.D., O.T.R.

Test for a difference in lesion size, hemisphere, and interaction. Let $\alpha = .05$ for all tests.

23. A random sample of the records of single births was selected from each of four populations. The weights (grams) of the babies at birth were as follows:

Sample			
A	B	C	D
2946	3186	2300	2286
2913	2857	2903	2938
2280	3099	2572	2952
3685	2761	2584	2348

(Continued)

Sample			
A	B	C	D
2310	3290	2675	2691
2582	2937	2571	2858
3002	3347		2414
2408			2008
			2850
			2762

Do these data provide sufficient evidence to indicate, at the .05 level of significance, that the four populations differ with respect to mean birth weight? Test for a significant difference between all possible pairs of means.

24. The following table shows the aggression scores of 30 laboratory animals reared under three different conditions. One animal from each of 10 litters was randomly assigned to each of the three rearing conditions.

Litter	Rearing Condition		
	Extremely Crowded	Moderately Crowded	Not Crowded
1	30	20	10
2	30	10	20
3	30	20	10
4	25	15	10
5	35	25	20
6	30	20	10
7	20	20	10
8	30	30	10
9	25	25	10
10	30	20	20

Do these data provide sufficient evidence to indicate that level of crowding has an effect on aggression? Let $\alpha = .05$.

25. The following table shows the vital capacity measurements of 60 adult males classified by occupation and age group:

Age Group	Occupation			
	A	B	C	D
1	4.31	4.68	4.17	5.75
	4.89	6.18	3.77	5.70
	4.05	4.48	5.20	5.53
	4.44	4.23	5.28	5.97
	4.59	5.92	4.44	5.52

(Continued)

Age Group	Occupation			
	A	B	C	D
2	4.13	3.41	3.89	4.58
	4.61	3.64	3.64	5.21
	3.91	3.32	4.18	5.50
	4.52	3.51	4.48	5.18
	4.43	3.75	4.27	4.15
3	3.79	4.63	5.81	6.89
	4.17	4.59	5.20	6.18
	4.47	4.90	5.34	6.21
	4.35	5.31	5.94	7.56
	3.59	4.81	5.56	6.73

Test for differences among occupations, for differences among age groups, and for interaction. Let $\alpha = .05$ for all tests.

26. Complete the following ANOVA table and state which design was used.

Source	SS	<i>d.f.</i>	MS	V.R.	<i>p</i>
Treatments	154.9199	4			
Error					
Total	200.4773	39			

27. Complete the following ANOVA table and state which design was used.

Source	SS	<i>d.f.</i>	MS	V.R.	<i>p</i>
Treatments		3			
Blocks	183.5	3			
Error	26.0				
Total	709.0	15			

28. Consider the following ANOVA table.

Source	SS	<i>d.f.</i>	MS	V.R.	<i>p</i>
A	12.3152	2	6.15759	29.4021	<.005
B	19.7844	3	6.59481	31.4898	<.005
AB	8.94165	6	1.49027	7.11596	<.005
Treatments	41.0413	11			
Error	10.0525	48	0.209427		
Total	51.0938	59			

- (a) What sort of analysis was employed?
- (b) What can one conclude from the analysis? Let $\alpha = .05$.

29. Consider the following ANOVA table.

Source	SS	<i>d.f.</i>	MS	V.R.
Treatments	5.05835	2	2.52917	1.0438
Error	65.42090	27	2.4230	

- (a) What design was employed?
- (b) How many treatments were compared?
- (c) How many observations were analyzed?
- (d) At the .05 level of significance, can one conclude that there is a difference among treatments? Why?

30. Consider the following ANOVA table.

Source	SS	<i>d.f.</i>	MS	V.R.
Treatments	231.5054	2	115.7527	2.824
Blocks	98.5000	7	14.0714	
Error	573.7500	14	40.9821	

- (a) What design was employed?
- (b) How many treatments were compared?
- (c) How many observations were analyzed?
- (d) At the .05 level of significance, can one conclude that the treatments have different effects? Why?

31. In a study of the relationship between smoking and serum concentrations of high-density lipoprotein cholesterol (HDL-C), the following data (coded for ease of calculation) were collected from samples of adult males who were nonsmokers, light smokers, moderate smokers, and heavy smokers. We wish to know if these data provide sufficient evidence to indicate that the four populations differ with respect to mean serum concentration of HDL-C. Let the probability of committing a type I error be .05. If an overall significant difference is found, determine which pairs of individual sample means are significantly different.

Smoking Status				
Nonsmokers	Light	Moderate	Heavy	
12	9	5	3	
10	8	4	2	
11	5	7	1	
13	9	9	5	
9	9	5	4	
9	10	7	6	
12	8	6	2	

32. Polyzogopoulou et al. (A-36) report the effects of bariatric surgery on fasting glucose levels (mmol/L) on 12 obese subjects with type 2 diabetes at four points in time: pre-operation, at 3 months, 6 months, and 12 months. Can we conclude, after eliminating subject effects, that fasting glucose levels differ over time after surgery? Let $\alpha = .05$.

Subject No.	Pre-op	3 Months	6 Months	12 Months
1	108.0	200.0	94.3	92.0
2	96.7	119.0	84.0	93.0
3	77.0	130.0	76.0	74.0
4	92.0	181.0	82.5	80.5
5	97.0	134.0	81.0	76.0
6	94.0	163.0	96.0	71.0
7	76.0	125.0	74.0	75.5
8	100.0	189.0	97.0	88.5
9	82.0	282.0	91.0	93.0
10	103.5	226.0	86.0	80.5
11	85.5	145.0	83.5	83.0
12	74.5	156.0	71.0	87.0

Source: Data provided courtesy of Theodore K. Alexandrides, M.D.

33. Refer to Review Exercise 32. In addition to studying the 12 type 2 diabetes subjects (group 1), Polyzogopoulou et al. (A-36) studied five subjects with impaired glucose tolerance (group 2), and eight subjects with normal glucose tolerance (group 3). The following data are the 12-month post-surgery fasting glucose levels for the three groups.

Group	
1.0	92.0
1.0	93.0
1.0	74.0
1.0	80.5
1.0	76.0
1.0	71.0
1.0	75.5
1.0	88.5
1.0	93.0
1.0	80.5
1.0	83.0
1.0	87.0
2.0	79.0
2.0	78.0
2.0	100.0
2.0	76.5
2.0	68.0
3.0	81.5
3.0	75.0
3.0	76.5

(Continued)

Group	
3.0	70.5
3.0	69.0
3.0	73.8
3.0	74.0
3.0	80.0

Source: Data provided courtesy of Theodore K. Alexandrides, M.D.

Can we conclude that there is a difference among the means of the three groups? If so, which pairs of means differ? Let $\alpha = .05$ for all tests.

For exercises 34 to 38 do the following:

- (a) Indicate which technique studied in this chapter (the completely randomized design, the randomized block design, the repeated measures design, or the factorial experiment) is appropriate.
 - (b) Identify the response variable and treatment variables.
 - (c) As appropriate, identify the factors and the number of levels of each, the blocking variables, and the subjects.
 - (d) List any extraneous variables whose effects you think might be included in the error term.
 - (e) As appropriate, comment on carry-over and position effects.
 - (f) Construct an ANOVA table in which you indicate the sources of variability and the number of degrees of freedom for each.
34. Johnston and Bowling (A-37) studied the ascorbic acid content (vitamin C) in several orange juice products. One of the products examined was ready-to-drink juice packaged in a re-sealable, screw-top container. One analysis analyzed the juice for reduced and oxidized vitamin C content at time of purchase and reanalyzed three times weekly for 4 to 5 weeks.
 35. A study by Pittini et al. (A-38) assessed the effectiveness of a simulator-based curriculum on 30 trainees learning the basic practice of amniocentesis. Pre- and post-training performance were evaluated with the same instrument. The outcome variable was the post-training score—pretraining score. Trainees were grouped by years of postgraduate experience: PGY 0–2, PGY 3–5, Fellows, and Faculty.
 36. Anim-Nyame et al. (A-39) studied three sets of women in an effort to understand factors related to pre-eclampsia. Enrolled in the study were 18 women with pre-eclampsia, 18 normal pregnant women, and 18 nonpregnant female matched controls. Blood samples were obtained to measure plasma levels of vascular endothelial growth factor, leptin, TNF- α plasma protein concentrations, and full blood count.
 37. In a study by Iwamoto et al. (A-40) 26 women were randomly assigned to the medication alfacalcidol for treatment of lumbar bone mineral density (BMD). BMD of the lumbar spine was measured at baseline and every year for 5 years.
 38. Inoue et al. (A-41) studied donor cell type and genotype on the efficiency of mouse somatic cell cloning. They performed a factorial experiment with two donor cell types (Sertoli cells or cumulus) and six genotypes. Outcome variables were the cleavage rate and the birth rate of pups in each treatment combination.

For the studies described in Exercises 39 through 66, do the following:

- (a) Perform a statistical analysis of the data (including hypothesis testing and confidence interval construction) that you think would yield useful information for the researchers.
 - (b) Determine p values for each computed test statistic.
 - (c) State all assumptions that are necessary to validate your analysis.
 - (d) Describe the population(s) about which you think inferences based on your analysis would be applicable.
39. Shirakami et al. (A-42) investigated the clinical significance of endothelin (ET), natriuretic peptides, and the renin-angiotensin-aldosterone system in pediatric liver transplantation. Subjects were children ages 6 months to 12 years undergoing living-related liver transplantation due to congenital biliary atresia and severe liver cirrhosis. Among the data collected were the following serum total bilirubin (mg/dl) levels after transplantation (h=hours, d=days):

Preoperative	Liver Transection	Anhepatic Phase	Time After Reperfusion of Donor Liver						
			1 h	2 h	4 h	8 h	1 d	2 d	3 d
6.2	1.2	0.9	0.8	1.1	1.5	2	1.4	1.6	1.3
17.6	11.9	9.3	3.5	3	6.1	9	6.3	6.4	6.2
13.2	10.2	7.9	5.3	4.9	3.3	3.6	2.8	1.9	1.9
3.9	3.3	3	2.9	2.3	1.4	1.2	0.8	0.8	0.9
20.8	19.4	*	9.4	8.4	6.8	7.1	3.7	3.8	3.2
1.8	1.8	1.6	1.4	1.4	1.1	1.9	0.7	0.8	0.7
8.6	6.5	4.8	3.1	2.1	1	1.3	1.5	1.6	3.2
13.4	12	10.1	5.8	5.6	4.5	4.1	3	3.1	3.6
16.8	13.9	8.3	3.7	3.7	2.2	2.1	1.9	3.1	4.1
20.4	17.8	17	10.8	9.3	8.9	7	2.8	3.8	4.8
25	21.5	13.8	7.6	7	5	11.5	12.3	10.1	11.4
9.2	6.3	6.8	5.3	4.8	0.2	4	4.2	3.7	3.5
8	6.5	6.4	4.1	3.8	3.8	3.5	3.1	2.9	2.8
2.9	3	4.1	3.4	3.4	3.7	4.2	3.3	2	1.9
21.3	17.3	13.6	9.2	7.9	7.9	9.8	8.6	4.7	5.5
25	25	24	20.1	19.3	18.6	23.6	25	14.4	20.6
23.3	23.7	15.7	13.2	11	9.6	9.3	7.2	6.3	6.3
17.5	16.2	14.4	12.6	12.7	11.5	10	7.8	5.5	4.9

*Missing observation.

Source: Data provided courtesy of Dr. Gotaro Shirakami.

Note that there is a missing observation in the data set. You may handle this problem in at least three ways.

- (a) Omit the subject whose datum is missing, and analyze the data of the remaining 17 subjects.
- (b) Use a computer package that automatically deals with missing data.
- (c) Analyze the data using a missing data procedure. For such a procedure, see Jerome L. Myers and Arnold D. Well, *Research Design and Statistical Analysis*, Erlbaum Associates, Hillsdale, NJ, 1995, pp. 256–258.

40. The purpose of a study by Sakakibara and Hayano (A-43) was to examine the effect of voluntarily slowed respiration on the cardiac parasympathetic response to a threat (the anticipation of an electric shock). Subjects were 30 healthy college students whose mean age was 23 years with a standard deviation of 1.5 years. An equal number of subjects were randomly assigned to slow (six males, four females), fast (seven males, three females), and nonpaced (five males, five females) breathing groups. Subjects in the slow- and fast-paced breathing groups regulated their breathing rate to 8 and 30 cpm, respectively. The nonpaced group breathed spontaneously. The following are the subjects' scores on the State Anxiety Score of State-Trait Anxiety Inventory after baseline and period of threat:

Slow paced		Fast paced		Nonpaced	
Baseline	Threat	Baseline	Threat	Baseline	Threat
39	59	37	49	36	51
44	47	40	42	34	71
48	51	39	48	50	37
50	61	47	57	49	53
34	48	45	49	38	52
54	69	43	44	39	56
34	43	32	45	66	67
38	52	27	54	39	49
44	48	44	44	45	65
39	65	41	61	42	57

Source: Data provided courtesy of Dr. Masahito Sakakibara.

41. Takahashi et al. (A-44) investigated the correlation of magnetic resonance signal intensity with spinal cord evoked potentials and spinal cord morphology after 5 hours of spinal cord compression in cats. Twenty-four adult cats were divided into four groups on the basis of a measure of spinal cord function plus a control group that did not undergo spinal compression. Among the data collected were the following compression ratio [(sagittal diameter/transverse diameter) \times 100] values after 5 hours of compression:

Control	80.542986	Group III	36.923077
	79.111111		31.304348
	70.535714		53.333333
	87.323944		55.276382
	80.000000		40.725806
Group I	82.222222	Group IV	66.666667
	83.928571		29.565217
	84.183673		12.096774
	48.181818		34.274194
Group II	98.461538		24.000000
	30.263158		
	34.865900		
	43.775100		
	82.439024		

Source: Data provided courtesy of Dr. Toshiaki Takahashi.

42. The objective of a study by Yamashita et al. (A-45) was to investigate whether pentoxifylline administered in the flush solution or during reperfusion would reduce ischemia-reperfusion lung injury in preserved canine lung allografts. Three groups of animals were studied. Pentoxifylline was not administered to animals in group 1 (C), was administered only during the reperfusion period (P) to animals in group 2, and was administered only in the flush solution to animals in group 3 (F). A total of 14 left lung allotransplantations were performed. The following are the aortic pressure readings for each animal during the 6-hour assessment period:

Group	0 min	60 min	120 min	180 min	240 min	300 min	360 min
C	85.0	100.0	120.0	80.0	72.0	75.0	*
C	85.0	82.0	80.0	80.0	85.0	80.0	80.0
C	100.0	75.0	85.0	98.0	85.0	80.0	82.0
C	57.0	57.0	57.0	30.0	*	*	*
C	57.0	75.0	52.0	56.0	65.0	95.0	75.0
P	112.0	67.0	73.0	90.0	71.0	70.0	66.0
P	92.0	70.0	90.0	80.0	75.0	80.0	² *
P	105.0	62.0	73.0	75.0	70.0	55.0	50.0
P	80.0	73.0	50.0	35.0	*	*	*
F	70.0	95.0	105.0	115.0	110.0	105.0	100.0
F	60.0	63.0	140.0	135.0	125.0	130.0	120.0
F	67.0	65.0	75.0	75.0	80.0	80.0	80.0
F	115.0	107.0	90.0	103.0	110.0	112.0	95.0
F	90.0	99.0	102.0	110.0	117.0	118.0	103.0

*Missing observation.

Source: Data provided courtesy of Dr. Motohiro Yamashita.

43. In a study investigating the relative bioavailability of beta-carotene (BC) and alpha-carotene (AC) from different sources of carrots, Zhou et al. (A-46) used ferrets as experimental animals. Among the data collected were the following concentrations of BC, AC, and AC/BC molar ratios in the sera of 24 ferrets provided with different sources of carotenoids for 3 days in their drinking water:

BC ($\mu\text{mol/g}$)	AC ($\mu\text{mol/g}$)	AC/BC (mol/mol)
Unheated Juice		
0.637	0.506	0.795
0.354	0.297	0.840
0.287	0.249	0.869
0.533	0.433	0.813
0.228	0.190	0.833
0.632	0.484	0.767
Heated Juice		
0.303	0.266	0.878
0.194	0.180	0.927

(Continued)

BC ($\mu\text{mol/g}$)	AC ($\mu\text{mol/g}$)	AC/BC (mol/mol)
Heated Juice		
0.293	0.253	0.864
0.276	0.238	0.859
0.226	0.207	0.915
0.395	0.333	0.843
Unheated Chromoplast		
0.994	0.775	0.780
0.890	0.729	0.819
0.809	0.661	0.817
0.321	0.283	0.882
0.712	0.544	0.763
0.949	0.668	0.704
Heated Chromoplast		
0.933	0.789	0.845
0.280	0.289	1.031
0.336	0.307	0.916
0.678	0.568	0.837
0.714	0.676	0.947
0.757	0.653	0.862

Source: Data provided
courtesy of Dr. Jin-R. Zhou.

44. Potteiger et al. (A-47) wished to determine if sodium citrate ingestion would improve cycling performance and facilitate favorable metabolic conditions during the cycling ride. Subjects were eight trained male competitive cyclists whose mean age was 25.4 years with a standard deviation of 6.5. Each participant completed a 30-km cycling time trial under two conditions, following ingestion of sodium citrate and following ingestion of a placebo. Blood samples were collected prior to treatment ingestion (PRE-ING); prior to exercising (PRE-EX); during the cycling ride at completion of 10, 20, and 30 km; and 15 minutes after cessation of exercise (POST-EX). The following are the values of partial pressures of oxygen (P_{O_2}) and carbon dioxide (P_{CO_2}) for each subject, under each condition, at each measurement time:

(P_{O_2}) (mm Hg)							
Subject	Treatment ^a	Measurement Times					
		PRE-ING	PRE-EX	10-km	20-km	30-km	15-POST-EX
1	1	42.00	20.00	53.00	51.00	56.00	41.00
1	2	43.00	29.00	58.00	49.00	55.00	56.00
2	1	44.00	38.00	66.00	66.00	76.00	58.00
2	2	40.00	26.00	57.00	47.00	46.00	45.00
3	1	37.00	22.00	59.00	58.00	56.00	52.00

(Continued)

(P_{O₂}) (mm Hg)							
Subject	Treatment^a	Measurement Times					
		PRE-ING	PRE-EX	10-km	20-km	30-km	15-POST-EX
3	2	36.00	30.00	52.00	65.00	65.00	36.00
4	1	34.00	21.00	65.00	62.00	62.00	59.00
4	2	46.00	36.00	65.00	72.00	72.00	66.00
5	1	36.00	24.00	41.00	43.00	50.00	46.00
5	2	41.00	25.00	52.00	60.00	67.00	54.00
6	1	28.00	31.00	52.00	60.00	53.00	46.00
6	2	34.00	21.00	57.00	58.00	57.00	41.00
7	1	39.00	28.00	72.00	69.00	65.00	72.00
7	2	40.00	27.00	64.00	61.00	57.00	60.00
8	1	49.00	27.00	67.00	61.00	51.00	49.00
8	2	27.00	22.00	56.00	64.00	49.00	34.00

(P_{CO₂}) (mm Hg)							
Subject	Treatment^a	Measurement Times					
		PRE-ING	PRE-EX	10-km	20-km	30-km	15-POST-EX
1	1	31.70	30.20	28.20	29.80	28.20	30.10
1	2	24.60	24.40	34.40	35.20	30.90	34.00
2	1	27.10	35.90	31.30	35.40	34.10	42.00
2	2	21.70	37.90	31.90	39.90	45.10	48.00
3	1	37.40	49.60	39.90	39.70	39.80	42.80
3	2	38.40	42.10	40.90	37.70	37.70	45.60
4	1	36.60	45.50	34.80	33.90	34.00	40.50
4	2	39.20	40.20	31.90	32.30	33.70	45.90
5	1	33.70	39.50	32.90	30.50	28.50	37.20
5	2	31.50	37.30	32.40	31.90	30.20	31.70
6	1	35.00	41.00	38.70	37.10	35.80	40.00
6	2	27.20	36.10	34.70	36.30	34.10	40.60
7	1	28.00	36.50	30.70	34.60	34.30	38.60
7	2	28.40	31.30	48.10	43.70	35.10	34.70
8	1	22.90	28.40	25.70	28.20	32.30	34.80
8	2	41.40	41.80	29.50	29.90	31.30	39.00

^a 1 = Sodium citrate; 2 = placebo.

Source: Data provided courtesy of Dr. Jeffrey A. Potteiger.

45. Teitge et al. (A-48) describe a radiographic method to demonstrate patellar instability. The 90 subjects ranged in age from 13 to 52 years and were divided into the following four groups on the basis of clinical findings regarding the nature of instability of the knee: normal (no symptoms or signs related to the knee), lateral, medial, and multidirectional instability. Among the data collected were the following radiographic measurements of the congruence angle (degrees):

<u>Normal</u>	<u>Lateral</u>		<u>Medial</u>		<u>Multidirectional</u>
-8	4	12	-16	10	15
-16	18	-8	-25	-5	-26
-22	5	-8	20	-10	-8
-26	-6	-20	-8	-12	-12
-8	32	-5	8	-14	-40
12	30	-10	-14		-20
-8	-10		-18		-16
12	28		-4		-34
-20	6		-20		-14
-20	9		-20		-6
-5	10		-20		-35
10	20		-22		-24
-4	-9		-15		-25
-2	-10		-10		10
-6	12		-5		-16
-7	0		-5		-30
0	35		-6		-30
-2	-1		-15		
-15	5		-25		
-5	22		-10		
	22		-20		

Source: Data provided courtesy of Dr. Robert A. Teitge.

46. A study by Ikeda et al. (A-49) was designed to determine the dose of ipratropium bromide aerosol that improves exercise performance using progressive cycle ergometry in patients with stable chronic obstructive pulmonary disease. The mean age of the 20 male subjects was 69.2 years with a standard deviation of 4.6 years. Among the data collected were the following maximum ventilation ($VE_{\max,L}/\text{min}$) values at maximum achieved exercise for different ipratropium bromide dosage levels (μg):

<u>Placebo</u>	<u>40</u>	<u>80</u>	<u>160</u>	<u>240</u>
26	24	23	25	28
38	39	43	43	37
49	46	54	57	52
37	39	39	38	38
34	33	37	37	41
42	38	44	44	42
23	26	28	27	22
38	41	44	37	40
37	37	36	38	39
33	35	34	38	36
40	37	40	46	40
52	58	48	58	63
45	48	47	51	38

(Continued)

Placebo	40	80	160	240
24	30	23	27	30
41	37	39	46	42
56	54	51	58	58
35	51	49	51	46
28	41	37	33	38
28	34	34	35	35
38	40	43	39	45

Source: Data provided courtesy of Dr. Akihiko Ikeda.

47. Pertovaara et al. (A-50) compared the effect of skin temperature on the critical threshold temperature eliciting heat pain with the effect of skin temperature on the response latency to the first heat pain sensation. Subjects were healthy adults between the ages of 23 and 54 years. Among the data collected were the following latencies (seconds) to the first pain response induced by radiant heat stimulation at three different skin temperatures:

Subject	25°C	30°C	35°C
1	6.4	4.5	3.6
2	8.1	5.7	6.3
3	9.4	6.8	3.2
4	6.75	4.6	3.9
5	10	6.2	6.2
6	4.5	4.2	3.4

Source: Data provided courtesy of Dr. Antti Pertovaara.

48. A study for the development and validation of a sensitive and specific method for quantifying total activin-A concentrations has been reported on by Knight et al. (A-51). As part of the study they collected the following peripheral serum concentrations of activin-A in human subjects of differing reproductive status: normal follicular phase (FP), normal luteal phase (LP), pregnant (PREG), ovarian hyperstimulated for in vivo fertilization (HYP), postmenopausal (PM), and normal adult males. Hint: Convert responses to logarithms before performing analysis.

FP	LP	PREG	HYP	PM	Male
134.5	78.0	2674.0	253.1	793.1	196.7
159.2	130.4	945.6	294.3	385.1	190.6
133.2	128.3	5507.6	170.2	270.9	185.3
225.0	166.4	7796.5	219.8	640.3	335.4
146.4	115.2	5077.5	165.8	459.8	214.6
180.5	148.9	4541.9	159.0		

Source: Data provided courtesy of Dr. Philip G. Knight.

49. The purpose of a study by Maheux et al. (A-52) was to evaluate the effect of labor on glucose production and glucose utilization. Subjects were six normal pregnant women. Among the data

collected were the following glucose concentrations during four stages of labor: latent (A1) and active (A2) phases of cervical dilatation, fetal expulsion (B), and placental expulsion (C).

A1	A2	B	C
3.60	4.40	5.30	6.20
3.53	3.70	4.10	3.80
4.02	4.80	5.40	5.27
4.90	5.33	6.30	6.20
4.06	4.65	6.10	6.90
3.97	5.20	4.90	4.60

Source: Data provided courtesy of Dr. Pierre C. Maheux.

50. Trachtman et al. (A-53) conducted studies (1) to assess the effect of recombinant human (rh) IGF-I on chronic puromycin aminonucleoside (PAN) nephropathy and (2) to compare the results of rhIGF-I versus rhGH treatment in a model of focal segmental glomerulosclerosis. As part of the studies, male Sprague-Dawley rats were divided into four groups: PAN (IA), PAN + rhIGF-I (IB), normal (IIA), and normal + rhIGF-I (IIB). The animals yielded the following data on creatinine levels before (pre) and after 4, 8, and 12 weeks of treatment:

Group			
IA	IB	IIA	IIB
Pre			
44	44	44	35
44	44	44	44
44	44	44	44
53	44	44	35
44	44		
44	53		
4 Weeks			
97	44	53	44
88	35	44	53
62	44	44	53
53	35	53	44
62	62		
53	53		
8 Weeks			
53	53	62	44
53	53	53	62
44	53	62	44
53	44	53	44
62	53		
70	62		

(Continued)

Group			
IA	IB	IIA	IIB
12 Weeks			
88	79	53	53
70	79	62	62
53	79	53	53
70	62	62	53
88	79		
88	70		

Source: Data provided courtesy of Dr. Howard Trachtman.

51. Twelve healthy men, ages 22 through 35 years, yielded the following serum T_3 (nmol/L) levels at 0800 hours after 8 (day 1), 32 (day 2), and 56 (day 3) hours of fasting, respectively. Subjects were participants in a study of fasting-induced alterations in pulsatile glycoprotein secretion conducted by Samuels and Kramer (A-54).

Subject	T_3	Day	Subject	T_3	Day	Subject	T_3	Day	Subject	T_3	Day
1	88	1	2	115	1	3	119	1	4	164	1
1	73	2	2	77	2	3	93	2	4	120	2
1	59	3	2	75	3	3	65	3	4	86	3
Subject	T_3	Day	Subject	T_3	Day	Subject	T_3	Day	Subject	T_3	Day
5	93	1	6	119	1	7	152	1	8	121	1
5	91	2	6	57	2	7	70	2	8	107	2
5	113	3	6	44	3	7	74	3	8	133	3
Subject	T_3	Day	Subject	T_3	Day	Subject	T_3	Day	Subject	T_3	Day
9	108	1	10	124	1	11	102	1	12	131	1
9	93	2	10	97	2	11	56	2	12	83	2
9	75	3	10	74	3	11	58	3	12	66	3

Source: Data provided courtesy of Dr. Mary H. Samuels.

52. To determine the nature and extent to which neurobehavioral changes occur in association with the toxicity resulting from exposure to excess dietary iron (Fe), Sobotka et al. (A-55) used weanling male Sprague-Dawley rats as experimental subjects. The researchers randomly assigned the animals, according to ranked body weights, to one of five diet groups differentiated on the basis of amount of Fe present: Control—35 (1), 350 (2), 3500 (3), 4 (iron deficient) (4), and 20,000 (5) ppm, respectively. The following are the body weights of the animals (grams) at the end of 10 weeks.

Diet	Weight	Diet	Weight	Diet	Weight
1	396	1	335	1	373
2	368	2	349	4	292
3	319	3	302	5	116

(Continued)

Diet	Weight	Diet	Weight	Diet	Weight
4	241	4	220	4	291
5	138	5	118	5	154
1	331	1	394	4	281
2	325	2	300	5	118
3	331	3	285	4	250
4	232	4	237	5	119
5	116	5	113	4	242
1	349	1	377	5	118
2	364	2	366	4	277
3	392	3	269	5	104
4	310	4	344	5	120
5	131	5	Dead	5	102
1	341	1	336		
2	399	2	379		
3	274	3	195		
4	319	4	277		
5	131	5	148		
1	419	1	301		
2	373	2	368		
3	Dead	3	308		
4	220	4	299		
5	146	5	Dead		

Source: Data provided courtesy of Dr. Thomas J. Sobotka.

53. Hansen (A-56) notes that brain bilirubin concentrations are increased by hyperosmolality and hypercarbia, and that previous studies have not addressed the question of whether increased brain bilirubin under different conditions is due to effects on the entry into or clearance of bilirubin from brain. In a study, he hypothesized that the kinetics of increased brain bilirubin concentration would differ in respiratory acidosis (hypercarbia) and hyperosmolality. Forty-four young adult male Sprague-Dawley rats were sacrificed at various time periods following infusion with bilirubin. The following are the blood bilirubin levels ($\mu\text{mol/L}$) of 11 animals just prior to sacrifice 60 minutes after the start of bilirubin infusion:

Controls	Hypercarbia	Hyperosmolality
30	48	102
94	20	118
78	58	74
52	74	

Source: Data provided courtesy of Dr. Thor Willy Ruud Hansen.

54. Johansson et al. (A-57) compared the effects of short-term treatments with growth hormone (GH) and insulin-like growth factor I (IGF-I) on biochemical markers of bone metabolism in men with idiopathic osteoporosis. Subjects ranged in age from 32 to 57 years. Among the data collected were the following serum concentrations of IGF binding protein-3 at 0 and 7 days after first injection and 1, 4, 8, and 12 weeks after last injection with GH and IGF-I.

Patient No.	Treatment	0 Day	7 Days	1 Week	4 Weeks	8 Weeks	12 Weeks
1	GH	4507	4072	3036	2484	3540	3480
1	IGF-I	3480	3515	4003	3667	4263	4797
2	GH	2055	4095	2315	1840	2483	2354
2	IGF-I	2354	3570	3630	3666	2700	2782
3	GH	3178	3574	3196	2365	4136	3088
3	IGF-I	3088	3405	3309	3444	2357	3831
4	GH	3464	5874	2929	3903	3367	2938
4	IGF-I	2905	2888	2797	3083	3376	3464
5	GH	4142	4465	3967	4213	4321	4990
5	IGF-I	4990	4590	2989	4081	4806	4435
6	GH	3622	6800	6185	4247	4450	4199
6	IGF-I	3504	3529	4093	4114	4445	3622
7	GH	5390	5188	4788	4602	4926	5793
7	IGF-I	5130	4784	4093	4852	4943	5390
8	GH	3161	4942	3222	2699	3514	2963
8	IGF-I	3074	2691	2614	3003	3145	3161
9	GH	3228	5995	3315	2919	3235	4379
9	IGF-I	4379	3548	3339	2379	2783	3000
10	GH	5628	6152	4415	5251	3334	3910
10	IGF-I	5838	5025	4137	5777	5659	5628
11	GH	2304	4721	3700	3228	2440	2698
11	IGF-I	2698	2621	3072	2383	3075	2822

Source: Data provided courtesy of Dr. Anna G. Johansson.

55. The objective of a study by Srijbos et al. (A-58) was to compare the results of a 12-week hospital-based outpatient rehabilitation program (group 1) with those of a 12-week home-care rehabilitation program (group 2) in chronic obstructive pulmonary disease with moderate to severe airflow limitation. A control group (group 3) did not receive rehabilitation therapy. Among the data collected were the following breathing frequency scores of subjects 18 months after rehabilitation:

Group			Group		
1	2	3	1	2	3
12	16	24	12	16	24
16	14	16	12	12	14
16	12	18	14	12	15
14	12	18	16	12	16
12	18	24	12	12	16
12	12	24	12	15	18
12	10	18		20	16

Source: Data provided courtesy of Dr. Jaap H. Srijbos.

56. Seven healthy males (mean age 27.4 years with a standard deviation of 4.4) participated in a study by Lambert et al. (A-59), who measured intestinal absorption following oral ingestion and intestinal perfusion of a fluid. As part of the study the researchers recorded the following percent changes in

plasma volume at six points during 85 minutes of cycle exercise in the drinking and infusion experiments:

	Subject	1	2	3
Drinking	1	-8.4151514	-7.4902674	-8.02277330
	2	-12.1966790	-5.1496679	-10.46486300
	3	-9.7418719	-5.9062747	-7.06516950
	4	-15.0291920	-14.4165470	-16.61268200
	5	-5.8845683	-5.8845683	-3.57781750
	6	-9.7100000	-7.5700000	-3.52995560
	7	-6.9787024	-6.5752716	-5.07020210
Infusion	1	-13.5391010	-11.7186910	-10.77312900
	2	-8.8259516	-8.9029745	-6.38160030
	3	-4.2410016	-1.3448910	-2.49740390
	4	-10.7192870	-9.7651132	-11.12140900
	5	-6.9487760	-2.9830660	1.77828157
	6	-7.1160660	-5.4111706	-7.07086340
	7	-7.0497788	-5.7725485	-5.18045500
	Subject	4	5	6
Drinking	1	-7.35202650	-7.89172340	-7.84726700
	2	-8.40517240	-9.02789810	5.13333985
	3	-4.19974130	-3.33795970	-5.65380700
	4	-15.36239700	-17.63314100	-14.43982000
	5	-5.50433470	-5.12242600	-6.26313790
	6	-4.22938570	-7.86923080	-7.51168220
	7	-5.94416340	-5.21535350	-6.34285620
Infusion	1	-11.64145400	-12.40814000	-8.26411320
	2	-5.69396590	-6.38160030	-7.37350920
	3	-1.01234570	-5.58572150	-2.81811090
	4	-12.13053100	-15.98360700	-12.64667500
	5	2.28844839	2.59034233	1.56622058
	6	-8.35430040	-10.60663700	-9.45689580
	7	-7.92841880	-8.38462720	-8.44542770

Source: Data provided courtesy of Dr. C. V. Gisolfi.

57. Roemer et al. (A-60) developed a self-report measure of generalized anxiety disorder (GAD) for use with undergraduate populations. In reliability studies the undergraduate subjects completed the GAD questionnaire (GAD-Q) as well as the Penn State Worry Questionnaire (PSWQ). The following are the PSWQ scores made by four groups of subjects determined by their GAD status: GAD by questionnaire, Study II (group 1); non-GAD by questionnaire, Study II (group 2); GAD by questionnaire, Study I (group 3); and clinical GAD (group 4).

Group				
1	2	3	4	
59.0	50.0	46.0	65.0	65.0
51.0	28.0	77.0	62.0	66.0
58.0	43.0	80.0	76.0	69.0
61.0	36.0	60.0	66.0	73.0
64.0	36.0	59.0	78.0	67.0
68.0	30.0	56.0	76.0	78.0
64.0	24.0	44.0	74.0	76.0
67.0	39.0	71.0	73.0	66.0
56.0	29.0	54.0	61.0	55.0
78.0	48.0	64.0	63.0	59.0
48.0	36.0	66.0	75.0	44.0
62.0	38.0	59.0	63.0	68.0
77.0	42.0	68.0	55.0	64.0
72.0	26.0	59.0	67.5	41.0
59.0	35.0	61.0	70.0	54.0
	32.0	78.0	70.0	72.0
	43.0	70.0	55.0	74.0
	55.0	74.0	73.0	59.0
	42.0	73.0	80.0	63.0
	37.0		79.0	51.0
	36.0		79.0	72.0
	41.0		61.0	63.0
	36.0		61.0	58.0
	34.0		72.0	71.0
	42.0		67.0	
	35.0		74.0	
	51.0		65.0	
	37.0		68.0	
	50.0		72.0	
	39.0		75.0	
			56.0	

Source: Data provided courtesy of Dr. T. D. Borkovec.

58. Noting that non-Hodgkin's lymphomas (NHL) represent a heterogeneous group of diseases in which prognosis is difficult to predict, Christiansen et al. (A-61) report on the prognostic aspects of soluble intercellular adhesion molecule-1 (sICAM-1) in NHL. Among the data collected were the following serum sICAM-1 (ng/ml) levels in four groups of subjects: healthy controls (C), high-grade NHL (hNHL), low-grade NHL (lNHL), and patients with hairy cell leukemia (HCL).

C	hNHL			lNHL		HCL
309	460	844	824	961	581	382
329	222	503	496	1097	601	975
314	663	764	656	1099	572	663

(Continued)

C	hNHL			INHL		HCL
254	1235	1088	1038	625	439	429
304	500	470	1050	473	1135	1902
335	739	806	446	654	590	1842
381	1847	482	1218	508	404	314
456	477	734	511	454	382	430
294	818	616	317	889	692	645
450	585	836	334	805	484	637
422	1837	1187	1026	541	438	712
528	362	581	534	655	787	581
461	671	381	292	654	77	860
286	375	699	782	1859	478	448
309	543	1854	1136	619	602	735
226	352	769	476	1837	802	
388	443	510		534	568	
377	359	571		424	665	
310	383	1248		571		
261	587	784		420		
350	648	514		408		
405	782	678		391		
319	472	1264		493		
289	506	618		1162		
310	663	1123		460		
227	873	912		1113		
206	987	520		572		
226	859	1867		653		
309	1193	485		1340		
382	1836	287		656		
325	691	455				
		522				

Source: Data provided courtesy of Dr. Ilse Christiansen.

59. Cossette et al. (A-62) examined gender and kinship with regard to caregivers' use of informal and formal support and to two models of support. Among the data collected were the following ages of three groups of caregivers of a demented relative living at home: husbands, wives, and adult daughters.

Husband		Wife			Daughter	
64	66	73	59	67	40	50
70	58	71	66	67	47	58
55	81	70	80	57	46	46
67	77	71	76	53	45	47
79	76	56	68	50	69	50
67	64	68	53	70	48	53
77	82	76	78	70	53	57
68	85	67	75		50	65
72	63	66	74		47	50

(Continued)

Husband		Wife		Daughter	
67	72	67	86	62	43
77	77	72	63	55	59
70	79	72	52	49	44
65	63	70	55	43	45
65	80	66	71	44	41
74	70	73	67	47	50
86	85	78	78	57	58
72	76	64	70	49	35
71	67	78	68		50
78	72	59	78		59
71	60	71	59		45
88	74	70	72		50
77	65	67	73		48
75	53	78	75		51
66	70	67	54		46
80	72	55	65		62
76	74	64	67		55
67	79	69	83		50
65	63	59	70		43
62	77	55	72		39
82	78	75	71		50
75	69	68	76		50
80		65	74		43
74		81	68		28
70		79	69		
		75	72		

Source: Data provided courtesy of Sylvie Cossette, M.Sc., R.N.

60. Tasaka et al. (A-63) note that *Corynebacterium parvum* (CP) increases susceptibility to endotoxin, which is associated with increased production of tumor necrosis factor (TNF). They investigated the effect of CP-priming on the pathogenesis of acute lung injury caused by intratracheal *Escherichia coli* endotoxin (lipopolysaccharide [LPS]). Experimental animals consisted of female guinea pigs divided into four groups. Animals in two groups received a 4-mg/kg treatment of CP 7 days before the study. Subsequently, nonpretreated animals received either saline alone (Control) or endotoxin (LPS-alone). The pretreated groups received either saline (CP-alone) or LPS (CP + LPS). Among the data collected were the following values of lung tissue-to-plasma ratio of radio-iodized serum albumin assay:

Control	CP-alone	LPS-alone	CP + LPS
0.12503532	0.18191647	0.17669093	0.3651166
0.10862729	0.30887462	0.25344761	0.64062964
0.10552931	0.25011885	0.17372285	0.39208734
0.15587316	0.23858085	0.1786867	0.49942059
0.13672624	0.26558231	0.22209666	0.85718475
0.11290446	0.32298454	0.27064831	0.93030465

Source: Data provided courtesy of Dr. Sadatomo Tasaka.

61. According to Takahashi et al. (A-64) research indicates that there is an association between alterations in calcium metabolism and various bone diseases in patients with other disabilities. Using subjects with severe mental retardation (mean age 16 years) who had been living in institutions for most of their lives, Takahashi et al. examined the relationship between bone change and other variables. Subjects were divided into groups on the basis of severity of bone change. Among the data collected were the following serum alkaline phosphatase (IU/L) values:

Grade I: 109, 86, 79, 103, 47, 105, 188, 96, 249

Grade II: 86, 106, 164, 146, 111, 263, 162, 111

Grade III: 283, 201, 208, 301, 135, 192, 135, 83, 193, 175, 174, 193, 224, 192, 233

Source: Data provided courtesy of Dr. Mitsugi Takahashi.

62. Research indicates that dietary copper deficiency reduces growth rate in rats. In a related study, Allen (A-65) assigned weanling male Sprague-Dawley rats to one of three food groups: copper-deficient (CuD), copper-adequate (CuA), and pair-fed (PF). Rats in the PF group were initially weight-matched to rats of the CuD group and then fed the same weight of the CuA diet as that consumed by their CuD counterparts. After 20 weeks, the rats were anesthetized, blood samples were drawn, and organs were harvested. As part of the study the following data were collected:

Rat	Diet	Body weight (BW)(g)	Heart weight (HW)(g)	Liver weight (LW)(g)	Kidney weight (KW)(g)	Spleen weight (SW)(g)
1	CuD	253.66	0.89	2.82	1.49	0.41
2		400.93	1.41	3.98	2.15	0.76
3		355.89	1.24	5.15	2.27	0.69
4		404.70	2.18	4.77	2.99	0.76
6	PF	397.28	0.99	2.34	1.84	0.50
7		421.88	1.20	3.26	2.32	0.79
8		386.87	0.88	3.05	1.86	0.84
9		401.74	1.02	2.80	2.06	0.76
10		437.56	1.22	3.94	2.25	0.75
11	CuA	490.56	1.21	4.51	2.30	0.78
12		528.51	1.34	4.38	2.75	0.76
13		485.51	1.36	4.40	2.46	0.82
14		509.50	1.27	4.67	2.50	0.79
15		489.62	1.31	5.83	2.74	0.81
Rat	Diet	HW/BW (g/100 g)	LW/BW (g/100 g)	KW/BW (g/100 g)	SW/BW (g/100 g)	Ceruloplasmin (mg/dl)
1	CuD	0.00351	0.01112	0.00587	0.00162	nd
2		0.00352	0.00993	0.00536	0.00190	5.27
3		0.00348	0.01447	0.00638	0.00194	4.80
4		0.00539	0.01179	0.00739	0.00188	4.97

(Continued)

Rat	Diet	HW/BW (g/100 g)	LW/BW (g/100 g)	KW/BW (g/100 g)	SW/BW (g/100 g)	Ceruloplasmin (mg/dl)
6		0.00249	0.00589	0.00463	0.00126	35.30
7		0.00284	0.00773	0.00550	0.00187	39.00
8	PF	0.00227	0.00788	0.00481	0.00217	28.00
9		0.00254	0.00697	0.00513	0.00189	34.20
10		0.00279	0.00900	0.00514	0.00171	45.20
11		0.00247	0.00919	0.00469	0.00159	34.60
12		0.00254	0.00829	0.00520	0.00144	39.00
13	CuA	0.00280	0.00906	0.00507	0.00169	37.10
14		0.00249	0.00917	0.00491	0.00155	33.40
15		0.00268	0.01191	0.00560	0.00165	37.30

nd, no data.

Source: Data provided courtesy of Corrie B. Allen.

63. Hughes et al. (A-66) point out that systemic complications in acute pancreatitis are largely responsible for mortality associated with the disease. They note further that proinflammatory cytokines, particularly $\text{TNF}\alpha$, may play a central role in acute pancreatitis by mediating the systemic sequelae. In their research they used a bile-infusion model of acute pancreatitis to show amelioration of disease severity as well as an improvement in overall survival by $\text{TNF}\alpha$ inhibition. Experimental material consisted of adult male Sprague-Dawley rats weighing between 250 and 300 grams divided into three groups: untreated (bile solution infused without treatment); treated (bile solution infused preceded by treatment with polyclonal anti – $\text{TNF}\alpha$ antibody); and sham (saline infused). Among the data collected were the following hematocrit (%) values for animals surviving more than 48 hours:

Sham	Untreated	Treated
38	56	40
40	60	42
32	50	38
36	50	46
40	50	36
40		35
38		40
40		40
38		55
40		35
		36
		40
		40
		35
		45

Source: Data provided courtesy of
Dr. A. Osama Gaber.

64. A study by Smárason et al. (A-67) was motivated by the observations of other researchers that sera from pre-eclamptic women damaged cultured human endothelial cells. Subjects for the present study were women with pre-eclampsia, matched control women with normal pregnancies, and nonpregnant

women of childbearing age. Among the data collected were the following observations on a relevant variable measured on subjects in the three groups.

Pre-Eclampsia	Pregnant Controls	Nonpregnant Controls
113.5	91.4	94.5
106.6	95.6	115.9
39.1	113.1	107.2
95.5	100.8	103.2
43.5	88.2	104.7
49.2	92.2	94.9
99.5	78.6	93.0
102.9	96.9	100.4
101.2	91.6	107.1
104.9	108.6	105.5
75.4	77.3	119.3
71.1	100.0	88.2
73.9	61.7	82.2
76.0	83.3	125.0
81.3	103.6	126.1
72.7	92.3	129.1
75.3	98.6	106.9
55.2	85.0	110.0
90.5	128.2	127.3
55.8	88.3	128.6

Source: Data provided courtesy of Dr. Alexander Smáráson.

65. The objective of a study by LeRoith et al. (A-68) was to evaluate the effect of a 7-week administration of recombinant human GH (rhGH) and recombinant human insulin-like growth factor (rhIGF-I) separately and in combination on immune function in elderly female rhesus monkeys. The assay for the *in vivo* function of the immune system relied on the response to an immunization with tetanus toxoid. The following are the responses for the three treatment groups and a control group:

Saline	rhIGF-I	rhGH	rhIGF-I + rhGH
11.2	12.2	12.15	11.5
9.0	9.4	11.20	12.4
10.8	10.7	10.60	10.8
10.0	10.8	11.30	11.9
	9.1	11.00	11.0
	12.6		

Source: Data provided courtesy of Dr. Jack A. Yanowski.

66. Hampl et al. (A-69) note that inhaled nitric oxide (NO) is a selective pulmonary vasodilator. They hypothesized that a nebulized diethylenetriamine/NO (DETA/NO) would stay in the lower airways and continuously supply sufficient NO to achieve sustained vasodilation in chronic pulmonary hypertension. Experimental material consisted of adult, male, specific pathogen-free Sprague-Dawley rats randomly divided into four groups: untreated, pulmonary normotensive controls;

monocrotaline-injected (to induce hypertension) with no treatment (MCT); monocrotaline-injected treated with either a $5 - \mu\text{mol}$ dose or a $50 - \mu\text{mol}$ dose of DETA/NO. Nineteen days after inducing pulmonary hypertension in the two groups of rats, the researchers began the treatment procedure, which lasted for 4 days. They collected, among other data, the following measurements on cardiac output for the animals in the four groups:

Control	MCT	MCT + DETA/NO	
		$5\mu\text{mol}$	$50\mu\text{mol}$
71.8	42.8	72.5	47.1
66.1	53.2	62.9	86.6
67.6	56.1	58.9	56.0
66.4	56.5	69.3	

Source: Data provided courtesy of Dr. Stephen L. Archer.

Exercises for Use with Large Data Sets Available on the Following Website:
www.wiley.com/college/daniel

1. In Kreiter et al. (A-70) medical school exams were delivered via computer format. Because there were not enough computer stations to test the entire class simultaneously, the exams were administered over 2 days. Both students and faculty wondered if students testing on day 2 might have an advantage due to extra study time or a breach in test security. Thus, the researchers examined a large medical class ($n = 193$) tested over 2 days with three 2-hour 80-item multiple-choice exams. Students were assigned testing days via pseudorandom assignment. Of interest was whether taking a particular exam on day 1 or day 2 had a significant impact on scores. Use the data set MEDSCORES to determine if test, day, or interaction has significant impact on test scores. Let $\alpha = .05$.
2. Refer to the serum lipid-bound sialic acid data on 1400 subjects (LSADATA). We wish to conduct a study to determine if the measurement of serum lipid-bound sialic acid (LSA) might be of use in the detection of breast cancer. The LSA measurements (mg/dl) are for four populations of subjects: normal controls, A; patients with benign breast disease, B; patients with primary breast cancer, C; and patients with recurrent metastatic breast cancer, D. Select a simple random sample of size 10 from each population and perform an appropriate analysis to determine if we may conclude that the four population means are different. Let $\alpha = .05$ and determine the p value. Test all possible pairs of sample means for significance. What conclusions can one draw from the analysis? Prepare a verbal report of the findings. Compare your results with those of your classmates.
3. Refer to the serum angiotensin-converting enzyme data on 1600 subjects (SACEDATA). Sarcoidosis, found throughout the world, is a systemic granulomatous disease of unknown cause. The assay of serum angiotensin-converting enzyme (SACE) is helpful in the diagnosis of active sarcoidosis. The activity of SACE is usually increased in patients with the disease, while normal levels occur in subjects who have not had the disease, those who have recovered, and patients with other granulomatous disorders. The data are the SACE values for four populations of subjects classified according to status regarding sarcoidosis: never had, A; active, B; stable, C; recovered, D. Select a simple random sample of 15 subjects from each population and perform an analysis to determine if you can conclude that the population means are different. Let $\alpha = .05$. Use Tukey's test to test for significant differences among individual pairs of means. Prepare a written report on your findings. Compare your results with those of your classmates.

4. Refer to the urinary colony-stimulating factor data on 1500 subjects (CSFDATA). The data are the urinary colony-stimulating factor (CSF) levels in five populations: normal subjects and subjects with four different diseases. Each observation represents the mean colony count of four plates from a single urine specimen from a given subject. Select a simple random sample of size 15 from each of the five populations and perform an analysis of variance to determine if one may conclude that the population means are different. Let $\alpha = .05$. Use Tukey's HSD statistic to test for significant differences among all possible pairs of sample means. Prepare a narrative report on the results of your analysis. Compare your results with those of your classmates.
5. Refer to the red blood cell data on 1050 subjects (RBCDATA). Suppose that you are a statistical consultant to a medical researcher who is interested in learning something about the relationship between blood folate concentrations in adult females and the quality of their diet. The researcher has available three populations of subjects: those whose diet quality is rated as good, those whose diets are fair, and those with poor diets. For each subject there is also available her red blood cell (RBC) folate value (in $\mu\text{g/liter}$ of red cells). Draw a simple random sample of size 10 from each population and determine whether the researcher can conclude that the three populations differ with respect to mean RBC folate value. Use Tukey's test to make all possible comparisons. Let $\alpha = .05$ and find the p value for each test. Compare your results with those of your classmates.
6. Refer to the serum cholesterol data on 350 subjects under three diet regimens (SERUMCHO). A total of 347 adult males between the ages of 30 and 65 participated in a study to investigate the relationship between the consumption of meat and serum cholesterol levels. Each subject ate beef as his only meat for a period of 20 weeks, pork as his only meat for another period of 20 weeks, and chicken or fish as his only meat for another 20-week period. At the end of each period serum cholesterol determinations (mg/100ml) were made on each subject. Select a simple random sample of 10 subjects from the population of 350. Use two-way analysis of variance to determine whether one should conclude that there is a difference in population mean serum cholesterol levels among the three diets. Let $\alpha = .05$. Compare your results with those of your classmates.

REFERENCES

Methodology References

1. R. A. FISHER, *The Design of Experiments*, Eighth Edition, Oliver and Boyd, Edinburgh, 1966.
2. R. A. FISHER, *Contributions to Mathematical Statistics*, Wiley, New York, 1950.
3. R. A. FISHER, *Statistical Methods for Research Workers*, Fourteenth Edition, Hafner, New York, 1973.
4. KLAUS HINKELMANN and OSCAR KEMPTHORNE, *Design and Analysis of Experiments*, Revised Edition, Wiley, New York, 1994.
5. DOUGLAS C. MONTGOMERY, *Design and Analysis of Experiments*, Fourth Edition, Wiley, New York, 1997.
6. JEROME L. MYERS and ARNOLD D. WELL, *Research Design and Statistical Analysis*, Earlbaum Associates, Hillsdale, NJ, 1995.
7. JASON C. HSU, *Multiple Comparisons: Theory and Methods*, Chapman & Hall, New York, 1996.
8. PETER H. WESTFALL, RANDALL D. TOBIAS, DROR ROM, RUSSELL D. WOLFINGER, and YOSEF HOCHBERG, *Multiple Comparisons and Multiple Tests*, SAS Publishing, Cary, NC, 2001.
9. J. W. TUKEY, "The Problem of Multiple Comparisons," Mimeographed Monograph, Princeton University, 1953; cited in Yosef Hochberg and Ajit C. Tamhane, *Multiple Comparison Procedures*, John Wiley and Sons, New York, 1987.
10. B. J. WINER, *Statistical Principles in Experimental Design*, Second Edition, McGraw-Hill, New York, 1971.
11. C. Y. KRAMER, "Extension of Multiple Range Tests to Group Means with Unequal Sample Sizes," *Biometrics*, 63 (1956), 307–310.

Applications References

- A-1 DAVID H. HOLBEN, "Selenium Content of Venison, Squirrel, and Beef Purchased or Produced in Ohio, a Low Selenium Region of the United States," *Journal of Food Science*, 67 (2002), 431–433.
- A-2 W. D. MEMBER, W. M. MURRY, S. I. RINGLEB, K. L. KILGORE, and S. A. SNYDER, "A Transducer to Measure Isometric Elbow Moments," *Clinical Biomechanics*, 16 (2001), 918–920.
- A-3 JOHN P. HOLCOMB and RALPH J. ROTHENBERG, "The Effectiveness of Alendronate and Etidronate in Clinical Practice," (unpublished manuscript).
- A-4 JASMINKA ILICH-ERNST, RHONDA A. BROWNBILL, MARTHA A. LUDEMANN, and RONGWEI FU, "Critical Factors for Bone Health in Women across the Age Span: How Important Is Muscle Mass?" *Medscape Women's Health eJournal*, 7 (2002).
- A-5 PAUL B. GOLD, ROBERT N. RUBEY, and RICHARD T. HARVEY, "Naturalistic, Self-Assignment Comparative Trial of Bupropion SR, a Nicotine Patch, or Both for Smoking Cessation Treatment in Primary Care," *American Journal on Addictions*, 11 (2002), 315–331.
- A-6 X. WANG, X. SHEN, X. LI, and C. MAULI AGRAWAL, "Age-Related Changes in the Collagen Network and Toughness of Bone," *Bone*, 31 (2002), 1–7.
- A-7 FARHAD ATASSI, "Oral Home Care and the Reasons for Seeking Dental Care by Individuals on Renal Dialysis," *Journal of Contemporary Dental Practice*, 3 (2002), 031–041.
- A-8 SIMONE RAGAZZI, AGOSTINO PIERRO, MARK PETERS, LORELLA FASOLI, and SIMON EATON, "Early Full Blood Count and Severity of Disease in Neonates with Necrotizing Enterocolitis," *Pediatric Surgery International*, 19 (2003), 376–379.
- A-9 ROBERTA ROMITO, MARIA IDA PANSINI, FRANCESCO PERTICONE, GIANFRANCO ANTONELLI, MARIAVITTORIA PITZALIS, and PAOLO RIZZON, "Comparative Effect of Lercanidipine, Felodipine, and Nifedipine FITS on Blood Pressure and Heart Rate in Patients with Mild to Moderate Arterial Hypertension: The Lercanidipine in Adults (LEAD) Study," *Journal of Clinical Hypertension*, 5 (2003), 249–253.
- A-10 L. KOSMISKI, D. KURITZKES, J. HAMILTON, T. SHARP, K. LICHTENSTIEN, J. HILL, and R. ECKEL, "Fat Distribution Is Altered in HIV-Infected Men Without Clinical Evidence of the HIV Lipodystrophy Syndrome," *HIV Medicine*, 4 (2003), 235–240.
- A-11 B. M. BROOKS, F. D. ROSE, E. A. ATTREE, and A. ELLIOT-SQUARE, "An Evaluation of the Efficacy of Training People with Learning Disabilities in a Virtual Environment," *Disability and Rehabilitation*, 24 (2002), 622–626.
- A-12 BRIAN J. MCCONVILLE, M. HAROLD FOGELSON, ANDREW B. NORMAN, WILLIAM M. KLYKYLO, PAT Z. MANDERSCHIED, KAREN W. PARKER, and PAUL R. SANBERG, "Nicotine Potentiation of Haloperidol in Reducing Tic Frequency in Tourette's Disorder," *American Journal of Psychiatry*, 148 (1991), 793–794.
- A-13 M. E. VALENCIA, G. MCNEILL, J. M. BROCKWAY, and J. S. SMITH, "The Effect of Environmental Temperature and Humidity on 24h Energy Expenditure in Men," *British Journal of Nutrition*, 68 (September 1992), 319–327.
- A-14 D. S. HODGSON, C. I. DUNLOR, P. L. CHAPMAN, and J. L. GRANDY, "Cardiopulmonary Responses to Experimentally Induced Gastric Dilatation in Isoflurane-Anesthetized Dogs," *American Journal of Veterinary Research*, 53 (June 1992), 938–943.
- A-15 JOHN C. LICCIARDONE, SCOTT T. STOLL, KIMBERLY G. FULDA, DAVID P. RUSSO, JEFF SIU, WILLIAM WINN, and JOHN SWIFT, JR., "Osteopathic Manipulative Treatment for Chronic Low Back Pain: A Randomized Controlled Trial," *Spine*, 28 (2003), 1355–1362.
- A-16 Mid-Michigan Medical Center, Midland, Michigan, 1999; A study of oral condition of cancer patients. Available online at: http://calcnet.mth.cmich.edu/org/spss/Prj_cancer_data.htm.
- A-17 YITAO LIU, LUDMILA BELAYEV, WEIZHAO ZHAO, RAUL BUSTO, and MYRON D. GINSBURG, "MRZ 2/579, a Novel Uncompetitive N-Methyl-D-Aspartate Antagonist, Reduces Infarct Volume and Brain Swelling and Improves Neurological Deficit After Focal Cerebral Ischemia in Rats," *Brain Research*, 862 (2000), 111–119.
- A-18 DAVID W. STARCH, JERRY W. ALEXANDER, PHILIP C. NOBLE, SURAJ REDDY, and DAVID M. LINTNER, "Multistranded Hamstring Tendon Graft Fixation with a Central Four-Quadrant or a Standard Tibial Interference Screw for Anterior Cruciate Ligament Reconstruction," *American Journal of Sports Medicine*, 31 (2003), 338–344.
- A-19 D. HOLBEN, A. SMITH, J. ILICH, J. LANDOLL, J. HOLCOMB, and V. MATKOVIK, "Puberty Does Not Negatively Affect Selenium Status in Young Females," *Journal of the American Dietetic Association*, 102 (2002), 1082–1087.
- A-20 AXEL LINKE, GONG ZHAO, FABIO A. RECCHIA, JEFFREY WILLIAMS, XIAOBIN XU, and THOMAS H. HINTZE, "Shift in Metabolic Substrate Uptake by the Heart During Development of Alloxan-Induced Diabetes," *American Journal of Physiology—Heart and Circulatory Physiology*, 285 (2003), H1007–H1014.

- A-21 K. WERTHER, I. J. CHRISTENSEN, and H. J. NIELSEN, "Determination of Vascular Endothelial Growth Factor (VEGF) in Circulating Blood: Significance of VEGF in Various Leucocytes and Platelets," *Scandinavian Journal of Clinical and Laboratory Investigation*, 62 (2002), 343–350.
- A-22 CAROLYN B. YUCHA, SUSAN KOWALSKI, and CHAD L. CROSS. "Student Stress and Academic Performance: Home Hospital Program," *Journal of Nursing Education*, 48 (2009), 631–637.
- A-23 SHIGEKU URYU, JUN HARADA, MARIE HISAMOTO, and TOMICHIRO ODA, "Troglitazone Inhibits Both Postglutamate Neurotoxicity and Low-Potassium-Induced Apoptosis in Cerebellar Granule Neurons," *Brain Research*, 924 (2002), 229–236.
- A-24 A. MAIN, H. ABU-SAAD, R. SALT, I. VLACHONIKOLIS, and A. DOWSON, "Management by Nurses of Primary Headache: A Pilot Study," *Current Medical Research Opinion*, 18 (2002), 471–478.
- A-25 SIMONA PORCELLINI, GIULIANA VALLANTI, SILVIA NOZZA, GUIDO POLI, ADRIANO LAZZARIN, GUISEPPE TABUSSI, and ANTONIO GRASSIA, "Improved Thymopoietic Potential in Aviremic HIV Infected Individuals with HAART by Intermittent IL-2 Administration," *AIDS*, 17 (2003), 1621–1630.
- A-26 DOROTA GÓRECKA, MICHAŁ BENAREK, ADAM NOWINSKI, ELZBIETA PUSCINSKA, ANNA GOLJAN-GEREMEK, and JAN ZIELINSKI, "Diagnosis of Airflow Limitation Combined with Smoking Cessation Advice Increases Stop-Smoking Rate," *Chest*, 123 (2003), 1916–1923.
- A-27 ELI MELTZER, BRUCE M. PRENNER, and ANJULI NAYAK, "Efficacy and Tolerability of Once-Daily 5 mg of Desloratadine, an H1-Receptor Antagonist, in Patients with Seasonal Allergic Rhinitis Assessment During the Spring and Fall Allergy Seasons," *Clinical Drug Investigations*, 21 (2001), 25–32.
- A-28 K. WERTHER, I. J. CHRISTENSEN, and H. J. NIELSEN, "Determination of Vascular Endothelial Growth Factor (VEGF) in Circulating Blood: Significance of VEGF in Various Leucocytes and Platelets," *Scandinavian Journal of Clinical and Laboratory Investigation*, 62 (2002), 343–350.
- A-29 RUBIA P. MESHACK and KATHLEEN E. NORMAN, "A Randomized Controlled Trial of the Effects of Weights on Amplitude and Frequency of Postural Hand Tremor in People with Parkinson's Disease," *Clinical Rehabilitation*, 16 (2003), 481–492.
- A-30 G. LACROIX, S. TISSOT, F. ROGERIEUX, R. BEAULIEU, L. CORNU, C. GILLET, F. R. OBIDEL, J. P. LEFEVRE, and F. Y. BOIS, "Decrease in Ovalbumin-Induced Pulmonary Allergic Response by Benzaldehyde but Not Acetaldehyde Exposure in a Guinea Pig Model," *Journal of Toxicology and Environmental Health, Part A* (2002), 995–1012.
- A-31 TAKEMASA WATANABE, YOSHIKI SUGIYAMA, YOSHIKO SUMI, MISUZU WATANABE, KIYOMI TAKEUCHI, FUMIO KOBAYASHI, and KOICHI KONO, "Effects of Vital Exhaustion on Cardiac Autonomic Nervous Functions Assessed by Heart Rate Variability at Rest in Middle-Aged Male Workers," *International Journal of Behavioral Medicine*, 9 (2002), 68–75.
- A-32 JOHN BROOKER, Data analyzed by the Wright State University Statistical Consulting Center, Wright State University, Dayton, OH (1997).
- A-33 ELEFTHERIOS KELLIS, Ph.D., "Pubertal Children Are Unable to Relax During the Passive Gravity Correction Procedure on the Isokinetic Dynamometer," *Isokinetics and Exercise Science*, 10 (2002), 97–105.
- A-34 REED E. PYERITZ, ELLIOT K. FISHMAN, BARBARA A. BERNHARDT, and STANLEY S. SIEGELMAN, "Dural Ectasia Is a Common Feature of the Marfan Syndrome," *American Journal of Human Genetics*, 43 (1988), 726–732.
- A-35 ADINA HARTMAN-MAEIR, NACHUM SOROKER, HAIM RING, and NOOMI KATZ, "Awareness of Deficits in Stroke Rehabilitation," *Journal of Rehabilitation Medicine*, 34 (2002), 158–164.
- A-36 EFTIHIA V. POLYZOGOPOULOU, FOTIOS KALFARENTZOS, APOSTOLOS G. VAGENAKIS, and THEODORE K. ALEXANDRIDES, "Restoration of Euglycemia and Normal Acute Insulin Response to Glucose in Obese Subjects with Type 2 Diabetes Following Bariatric Surgery," *Diabetes*, 53 (2003), 1098–1103.
- A-37 CAROL S. JOHNSTON and D. L. BOWLING, "Stability of Ascorbic Acid in Commercially Available Orange Juices," *Journal of the American Dietetic Association*, 102 (2002), 525–529.
- A-38 R. PITTINI, D. OEPKESS, K. MACRURY, R. REZNICK, J. BEYENE, and R. WINDRIM, "Teaching Invasive Perinatal Procedures: Assessment of a High Fidelity Simulator-Based Curriculum," *Ultrasound Obstetrics and Gynecology*, 19 (2002), 478–483.
- A-39 NICK ANIM-NYAME, JOHN GAMBLE, SUREN R. SOORANNA, MARK R. JOHNSON, and PHILIP J. STEER, "Microvascular Permeability Is Related to Circulating Levels of Tumour Necrosis Factor- α in Pre-Eclampsia," *Cardiovascular Research*, 58 (2003), 162–169.
- A-40 JUN IWAMOTO, TSUYOSHI TAKEDA, SHOICHI ICHIMURA, and MITSUYOSHI UZAWA, "Effects of 5-Year Treatment with Elcatonin and Alfacalcidol on Lumbar Bone Mineral Density and the Incidence of Vertebral Fractures in

- Postmenopausal Women with Osteoporosis: A Retrospective Study," *Journal of Orthopaedic Science*, 7 (2002), 637–643.
- A-41 KIMIKO INOUE, NARUMI OGOBUKI, KEIJI MOCHIDA, YOSHIE YAMAMOTO, KAORU TAKANO, FUMITOSHI ISHINO, and ATSUO OGURA, "Effects of Donor Cell Type and Genotype on the Efficiency of Mouse Somatic Cell Cloning," *Biology of Reproduction*, 69 (2003), 1394–1400.
- A-42 GOTARO SHIRAKAMI, MASAHIRO MURAKAWA, KOH SHINGU, SUSUMU MASHIMA, SHIN-ICHI SUGA, TATSUO MAGARIBUCHI, TOYOHICO O'HIGASHI, TSUTOMU SHICHINO, KAZUWA NAKAO, and KENJIRO MORI, "Perioperative Plasma Concentrations of Endothelin and Natriuretic Peptides in Children Undergoing Living-Related Liver Transplantation," *Anesthesia and Analgesia*, 82 (1996), 235–240.
- A-43 MASAHIRO SAKAKIBARA and JUNICHIRO HAYANO, "Effect of Slowed Respiration on Cardiac Parasympathetic Response to Threat," *Psychosomatic Medicine*, 58 (1996), 32–37.
- A-44 TOSHIKI TAKAHASHI, YUJI SUTO, SHINSUKE KATO, and EISAKU OHAMA, "Experimental Acute Dorsal Compression of Cat Spinal Cord," *Spine*, 21 (1996), 166–173.
- A-45 MOTOHIRO YAMASHITA, RALPH A. SCHMID, KAN OKABAYASHI, KOEI ANDO, JUN KOBAYASHI, JOEL D. COOPER, and G. ALEXANDER PATTERSON, "Pentoxifylline in Flush Solution Improves Early Lung Allograft Function," *Annals of Thoracic Surgery*, 61 (1996), 1055–1061.
- A-46 JIN-R. ZHOU, ERIC T. GUGGER, and JOHN W. ERDMAN, JR., "The Crystalline Form of Carotenes and the Food Matrix in Carrot Root Decrease the Relative Bioavailability of Beta- and Alpha-Carotene in the Ferret Model," *Journal of the American College of Nutrition*, 15 (1996), 84–91.
- A-47 J. A. POTTEIGER, G. L. NICKEL, M. J. WEBSTER, M. D. HAUB, and R. J. PALMER, "Sodium Citrate Ingestion Enhances 30 km Cycling Performance," *International Journal of Sports Medicine*, 17 (1996), 7–11.
- A-48 ROBERT A. TEITGE, WADE FAERBER, PATRICIA DES MADRYL, and THOMAS M. MATELIC, "Stress Radiographs of the Patellofemoral Joint," *Journal of Bone and Joint Surgery*, 78-A (February 1996), 193–203.
- A-49 AKIHIKO IKEDA, KOICHI NISHIMURA, HIROSHI KOYAMA, MITSUHIRO TSUKINO, MICHIAKI MISHIMA, and TAKATERU IZUMI, "Dose Response Study of Ipratropium Bromide Aerosol on Maximum Exercise Performance in Stable Patients with Chronic Obstructive Pulmonary Disease," *Thorax*, 51 (1996), 48–53.
- A-50 ANTTI PERTOVAARA, TIMO KAUPPILA, and MINNA M. HÄMÄLÄINEN, "Influence of Skin Temperature on Heat Pain Threshold in Humans," *Experimental Brain Research*, 107 (1996), 497–503.
- A-51 P. G. KNIGHT, S. MUTTUKRISHNA, and N. P. GROOME, "Development and Application of a Two-Site Enzyme Immunoassay for the Determination of 'Total' Activin-A Concentrations in Serum and Follicular Fluid," *Journal of Endocrinology*, 148 (1996), 267–279.
- A-52 PIERRE C. MAHEUX, BRIGITTE BONIN, ANNE DIZAZO, PIERRE GUIMOND, DANIELLE MONIER, JOSÉE BOURQUE, and JEAN-LOUIS CHIASSON, "Glucose Homeostasis During Spontaneous Labor in Normal Human Pregnancy," *Journal of Clinical Endocrinology and Metabolism*, 81 (1996), 209–215.
- A-53 HOWARD TRACHTMAN, JOHN MAESAKA, STEPHEN FUTTERWEIT, CHAN MA, ELSA VALDERRAMA, JAMES C. M. CHAN, and WINNIE CHAN, "The Effect of Recombinant Human Insulin-Like Growth Factor-I on Chronic Puromycin Aminonucleoside Nephropathy in Rats," *Pediatric Research*, 39 (1996), 13–19.
- A-54 M. H. SAMUELS and P. KRAMER, "Differential Effects of Short-Term Fasting on Pulsatile Thyrotropin, Gonadotropin, and α -Subunit Secretion in Healthy Men—A Clinical Research Center Study," *Journal of Clinical Endocrinology and Metabolism*, 81 (1996), 32–36.
- A-55 T. J. SOBOTKA, P. WHITTAKER, J. M. SOBOTKA, R. E. BRODIE, D. Y. QUANDER, M. ROBL, M. BRYANT, and C. N. BARTON, "Neurobehavioral Dysfunctions Associated with Dietary Iron Overload," *Physiology and Behavior*, 58 (1996), 213–219.
- A-56 THOR WILLY RUUD HANSEN, "Bilirubin Entry into and Clearance from Rat Brain During Hypercarbia and Hyperosmolality," *Pediatric Research*, 39 (1996), 72–76.
- A-57 ANNA G. JOHANSSON, ERIK LINDH, WERNER F. BLUM, GINA KOLLERUP, OLE H. SØRENSEN, and SVERKER LJUNGHALL, "Effects of Growth Hormone and Insulin-Like Growth Factor I in Men with Idiopathic Osteoporosis," *Journal of Clinical Endocrinology and Metabolism*, 81 (1996), 44–48.
- A-58 JAAP H. STRUBOS, DIRKJE S. POSTMA, RICHARD ALTENA, FERNANDO GIMENO, and GERARD H. KOËTER, "A Comparison Between an Outpatient Hospital-Based Pulmonary Rehabilitation Program and a Home-Care Pulmonary Rehabilitation Program in Patients with COPD," *Chest*, 109 (1996), 366–372.
- A-59 G. P. LAMBERT, R. T. CHANG, D. JOENSEN, X. SHI, R. W. SUMMERS, H. P. SCHEDL, and C. V. GISOLFI, "Simultaneous Determination of Gastric Ingestion and Intestinal Absorption During Cycle Exercise in Humans," *International Journal of Sports Medicine*, 17 (1996), 48–55.

- A-60 LIZABETH ROEMER, MARY BORKOVEC, SHARON POSA, and T. D. BORKOVEC, "A Self-Diagnostic Measure of Generalized Anxiety Disorder," *Journal of Behavior Therapy & Experimental Psychiatry*, 26 (1995), 345–350.
- A-61 ILSE CHRISTIANSEN, CECILIA GIDLÖF, KARL-MICHAEL KÄLKNER, HANS HAGBERG, HELGE BENNMARKER, and THOMAS TÖTTERMAN, "Elevated Serum Levels of Soluble ICAM-1 in Non-Hodgkin's Lymphomas Correlate with Tumour Burden, Disease Activity and Other Prognostic Markers," *British Journal of Haematology*, 92 (1996), 639–646.
- A-62 SYLVIE COSSETTE, LOUISE LÉVESQUE, and LIANE LAURIN, "Informal and Formal Support for Caregivers of a Demented Relative: Do Gender and Kinship Make a Difference?" *Research in Nursing & Health*, 18 (1995), 437–451.
- A-63 SADATOMO TASAKA, AKITOSHI ISHIZAKA, KOICHI SAYAMA, FUMIO SAKAMAKI, HIDETOSHI NAKAMURA, TAKESHI TERASHIMA, YASUHIRO WAKI, KENZO SOEJIMA, MORIO NAKAMURA, HIROAKI MATSUBARA, SEITARO FUJISHIMA, and MINORU KANAZAWA, "Heat-Killed *Corynebacterium parvum* Enhances Endotoxin Lung Injury with Increased TNF Production in Guinea Pigs," *American Journal of Respiratory and Critical Care Medicine*, 153 (1996), 1047–1055.
- A-64 MITSUGI TAKAHASHI, HIDEO NAGAO, and HIROSI MATSUDA, "Estimating Bone Change in Patients with Severe Disability," *Development Medicine and Child Neurology*, 37 (1995), 1085–1093.
- A-65 CORRIE B. ALLEN, "Effects of Dietary Copper Deficiency on Relative Food Intake and Growth Efficiency in Rats," *Physiology & Behavior*, 59 (1996), 247–253.
- A-66 CHRISTOPHER B. HUGHES, LILLIAN W. GABER, ABOU BAKR MOHEY EL-DIN, HANI P. GREWAL, MALAK KOTB, LINDA MANN, and A. OSAMA GABER, "Inhibition of TNF α Improves Survival in an Experimental Model of Acute Pancreatitis," *American Surgeon*, 62 (1996), 8–13.
- A-67 ALEXANDER K. SMÁRASON, IAN L. SARGENT, and CHRISTOPHER W. G. REDMAN, "Endothelial Cell Proliferation Is Suppressed by Plasma but Not Serum from Women with Preeclampsia," *American Journal of Obstetrics and Gynecology*, 174 (1996), 787–793.
- A-68 DEREK LEROITH, JACK YANOWSKI, ERIC P. KALDJIAN, ELAINE S. JAFFE, TANYA LEROITH, KATHY PURDUE, BARBARA D. COOPER, ROBERT PYLE, and WILLIAM ADLER, "The Effects of Growth Hormone and Insulin-Like Growth Factor I on the Immune System of Aged Female Monkeys," *Endocrinology*, 137 (1996), 1071–1079.
- A-69 VÁCLAV HAMPL, MARTIN TRISTANI-FIROUZI, THOMAS C. HUTSELL, and STEPHEN L. ARCHER, "Nebulized Nitric Oxide/Nucleophile Adduct Reduces Chronic Pulmonary Hypertension," *Cardiovascular Research*, 31 (1996), 55–62.
- A-70 C. KREITER, M. W. PETERSON, K. FERGUSON, and S. ELLIOTT, "The Effects of Testing in Shifts on a Clinical In-Course Computerized Exam," *Medical Education*, 37 (2003), 202–204.

SIMPLE LINEAR REGRESSION AND CORRELATION

CHAPTER OVERVIEW

This chapter provides an introduction and overview of two common techniques for exploring the strength of the relationship between two variables. The first technique, linear regression, will help us find an objective way to predict or estimate the value of one variable given a value of another variable. The second technique, correlation, will help us find an objective measure of the strength of the relationship between two variables.

TOPICS

- 9.1 INTRODUCTION
- 9.2 THE REGRESSION MODEL
- 9.3 THE SAMPLE REGRESSION EQUATION
- 9.4 EVALUATING THE REGRESSION EQUATION
- 9.5 USING THE REGRESSION EQUATION
- 9.6 THE CORRELATION MODEL
- 9.7 THE CORRELATION COEFFICIENT
- 9.8 SOME PRECAUTIONS
- 9.9 SUMMARY

LEARNING OUTCOMES

After studying this chapter, the student will

1. be able to obtain a simple linear regression model and use it to make predictions.
2. be able to calculate the coefficient of determination and to interpret tests of regression coefficients.
3. be able to calculate correlations among variables.
4. understand how regression and correlation differ and when the use of each is appropriate.

9.1 INTRODUCTION

In analyzing data for the health sciences disciplines, we find that it is frequently desirable to learn something about the relationship between two numeric variables. We may, for example, be interested in studying the relationship between blood pressure and age, height and weight, the concentration of an injected drug and heart rate, the consumption level of some nutrient and weight gain, the intensity of a stimulus and reaction time, or total family income and medical care expenditures. The nature and strength of the relationships between variables such as these may be examined using linear models such as *regression* and *correlation* analysis, two statistical techniques that, although related, serve different purposes.

Regression Regression analysis is helpful in assessing specific forms of the relationship between variables, and the ultimate objective when this method of analysis is employed usually is to *predict* or *estimate* the value of one variable corresponding to a given value of another variable. The ideas of regression were first elucidated by the English scientist Sir Francis Galton (1822–1911) in reports of his research on heredity—first in sweet peas and later in human stature. He described a tendency of adult offspring, having either short or tall parents, to revert back toward the average height of the general population. He first used the word *reversion*, and later *regression*, to refer to this phenomenon.

Correlation Correlation analysis, on the other hand, is concerned with measuring the strength of the relationship between variables. When we compute measures of correlation from a set of data, we are interested in the degree of the *correlation* between variables. Again, the concepts and terminology of correlation analysis originated with Galton, who first used the word *correlation* in 1888.

In this chapter our discussion is limited to the exploration of the linear relationship between two variables. The concepts and methods of regression are covered first, beginning in the next section. In Section 9.6 the ideas and techniques of correlation are introduced. In the next chapter we consider the case where there is an interest in the relationships among three or more variables.

Regression and correlation analysis are areas in which the speed and accuracy of a computer are most appreciated. The data for the exercises of this chapter, therefore, are presented in a way that makes them suitable for computer processing. As is always the case, the input requirements and output features of the particular programs and software packages to be used should be studied carefully.

9.2 THE REGRESSION MODEL

In the typical regression problem, as in most problems in applied statistics, researchers have available for analysis a sample of observations from some real or hypothetical population. Based on the results of their analysis of the sample data, they are interested in reaching decisions about the population from which the sample is presumed to have been drawn. It is important, therefore, that the researchers understand the nature of the population in which they are interested. They should know enough about the population to be able either to construct a mathematical model for its representation or to determine if it reasonably fits

some established model. A researcher about to analyze a set of data by the methods of simple linear regression, for example, should be secure in the knowledge that the simple linear regression model is, at least, an approximate representation of the population. It is unlikely that the model will be a perfect portrait of the real situation, since this characteristic is seldom found in models of practical value. A model constructed so that it corresponds precisely with the details of the situation is usually too complex to yield any information of value. On the other hand, the results obtained from the analysis of data that have been forced into a model that does not fit are also worthless. Fortunately, however, a perfectly fitting model is not a requirement for obtaining useful results. Researchers, then, should be able to distinguish between the occasion when their chosen models and the data are sufficiently compatible for them to proceed and the case where their chosen model must be abandoned.

Assumptions Underlying Simple Linear Regression In the simple linear regression model two variables, usually labeled X and Y , are of interest. The letter X is usually used to designate a variable referred to as the *independent variable*, since frequently it is controlled by the investigator; that is, values of X may be selected by the investigator and, corresponding to each preselected value of X , one or more values of another variable, labeled Y , are obtained. The variable, Y , accordingly, is called the *dependent variable*, and we speak of the regression of Y on X . The following are the assumptions underlying the simple linear regression model.

1. Values of the independent variable X are said to be “fixed.” This means that the values of X are preselected by the investigator so that in the collection of the data they are not allowed to vary from these preselected values. In this model, X is referred to by some writers as a *nonrandom variable* and by others as a *mathematical variable*. It should be pointed out at this time that the statement of this assumption classifies our model as the *classical regression model*. Regression analysis also can be carried out on data in which X is a random variable.
2. The variable X is measured without error. Since no measuring procedure is perfect, this means that the magnitude of the measurement error in X is negligible.
3. For each value of X there is a subpopulation of Y values. For the usual inferential procedures of estimation and hypothesis testing to be valid, these subpopulations must be normally distributed. In order that these procedures may be presented it will be assumed that the Y values are normally distributed in the examples and exercises that follow.
4. The variances of the subpopulations of Y are all equal and denoted by σ^2 .
5. The means of the subpopulations of Y all lie on the same straight line. This is known as the *assumption of linearity*. This assumption may be expressed symbolically as

$$\mu_{y|x} = \beta_0 + \beta_1 x \quad (9.2.1)$$

where $\mu_{y|x}$ is the mean of the subpopulation of Y values for a particular value of X , and β_0 and β_1 are called the population regression coefficients. Geometrically, β_0 and β_1 represent the y -intercept and slope, respectively, of the line on which all of the means are assumed to lie.

6. The Y values are statistically independent. In other words, in drawing the sample, it is assumed that the values of Y chosen at one value of X in no way depend on the values of Y chosen at another value of X .

These assumptions may be summarized by means of the following equation, which is called the simple linear regression model:

$$y = \beta_0 + \beta_1 x + \epsilon \quad (9.2.2)$$

where y is a typical value from one of the subpopulations of Y , β_0 and β_1 are as defined for Equation 9.2.1, and ϵ is called the error term. If we solve 9.2.2 for ϵ , we have

$$\begin{aligned} \epsilon &= y - (\beta_0 + \beta_1 x) \\ &= y - \mu_{y|x} \end{aligned} \quad (9.2.3)$$

and we see that ϵ shows the amount by which y deviates from the mean of the subpopulation of Y values from which it is drawn. As a consequence of the assumption that the subpopulations of Y values are normally distributed with equal variances, the ϵ 's for each subpopulation are normally distributed with a variance equal to the common variance of the subpopulations of Y values.

The following acronym will help the reader remember most of the assumptions necessary for inference in linear regression analysis:

LINE [*L*inear (assumption 5), *I*ndependent (assumption 6), *N*ormal (assumption 3), *E*qual variances (assumption 4)]

A graphical representation of the regression model is given in Figure 9.2.1.

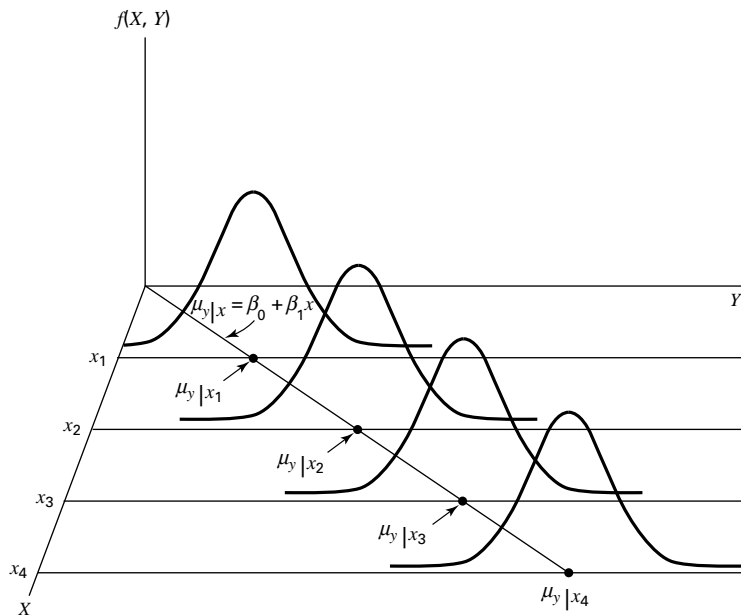


FIGURE 9.2.1 Representation of the simple linear regression model.

9.3 THE SAMPLE REGRESSION EQUATION

In simple linear regression the object of the researcher's interest is the population regression equation—the equation that describes the true relationship between the dependent variable Y and the independent variable X . The variable designated by Y is sometimes called the *response variable* and X is sometimes called the *predictor variable*.

In an effort to reach a decision regarding the likely form of this relationship, the researcher draws a sample from the population of interest and using the resulting data, computes a sample regression equation that forms the basis for reaching conclusions regarding the unknown population regression equation.

Steps in Regression Analysis In the absence of extensive information regarding the nature of the variables of interest, a frequently employed strategy is to assume initially that they are linearly related. Subsequent analysis, then, involves the following steps.

1. Determine whether or not the assumptions underlying a linear relationship are met in the data available for analysis.
2. Obtain the equation for the line that best fits the sample data.
3. Evaluate the equation to obtain some idea of the strength of the relationship and the usefulness of the equation for predicting and estimating.
4. If the data appear to conform satisfactorily to the linear model, use the equation obtained from the sample data to predict and to estimate.

When we use the regression equation to *predict*, we will be predicting the value Y is likely to have when X has a given value. When we use the equation to *estimate*, we will be estimating the mean of the subpopulation of Y values assumed to exist at a given value of X . Note that the sample data used to obtain the regression equation consist of known values of both X and Y . When the equation is used to predict and to estimate Y , only the corresponding values of X will be known. We illustrate the steps involved in simple linear regression analysis by means of the following example.

EXAMPLE 9.3.1

Després et al. (A-1) point out that the topography of adipose tissue (AT) is associated with metabolic complications considered as risk factors for cardiovascular disease. It is important, they state, to measure the amount of intraabdominal AT as part of the evaluation of the cardiovascular-disease risk of an individual. Computed tomography (CT), the only available technique that precisely and reliably measures the amount of deep abdominal AT, however, is costly and requires irradiation of the subject. In addition, the technique is not available to many physicians. Després and his colleagues conducted a study to develop equations to predict the amount of deep abdominal AT from simple anthropometric measurements. Their subjects were men between the ages of 18 and 42 years who were free from metabolic disease that would require treatment. Among the measurements taken on each subject were deep abdominal AT obtained by CT and waist circumference as

shown in Table 9.3.1. A question of interest is how well one can predict and estimate deep abdominal AT from knowledge of the waist circumference. This question is typical of those that can be answered by means of regression analysis. Since deep abdominal AT is the variable about which we wish to make predictions and estimations, it is the dependent variable. The variable waist measurement, knowledge of which will be used to make the predictions and estimations, is the independent variable.

TABLE 9.3.1 Waist Circumference (cm), X , and Deep Abdominal AT, Y , of 109 Men

Subject	X	Y	Subject	X	Y	Subject	X	Y
1	74.75	25.72	38	103.00	129.00	75	108.00	217.00
2	72.60	25.89	39	80.00	74.02	76	100.00	140.00
3	81.80	42.60	40	79.00	55.48	77	103.00	109.00
4	83.95	42.80	41	83.50	73.13	78	104.00	127.00
5	74.65	29.84	42	76.00	50.50	79	106.00	112.00
6	71.85	21.68	43	80.50	50.88	80	109.00	192.00
7	80.90	29.08	44	86.50	140.00	81	103.50	132.00
8	83.40	32.98	45	83.00	96.54	82	110.00	126.00
9	63.50	11.44	46	107.10	118.00	83	110.00	153.00
10	73.20	32.22	47	94.30	107.00	84	112.00	158.00
11	71.90	28.32	48	94.50	123.00	85	108.50	183.00
12	75.00	43.86	49	79.70	65.92	86	104.00	184.00
13	73.10	38.21	50	79.30	81.29	87	111.00	121.00
14	79.00	42.48	51	89.80	111.00	88	108.50	159.00
15	77.00	30.96	52	83.80	90.73	89	121.00	245.00
16	68.85	55.78	53	85.20	133.00	90	109.00	137.00
17	75.95	43.78	54	75.50	41.90	91	97.50	165.00
18	74.15	33.41	55	78.40	41.71	92	105.50	152.00
19	73.80	43.35	56	78.60	58.16	93	98.00	181.00
20	75.90	29.31	57	87.80	88.85	94	94.50	80.95
21	76.85	36.60	58	86.30	155.00	95	97.00	137.00
22	80.90	40.25	59	85.50	70.77	96	105.00	125.00
23	79.90	35.43	60	83.70	75.08	97	106.00	241.00
24	89.20	60.09	61	77.60	57.05	98	99.00	134.00
25	82.00	45.84	62	84.90	99.73	99	91.00	150.00
26	92.00	70.40	63	79.80	27.96	100	102.50	198.00
27	86.60	83.45	64	108.30	123.00	101	106.00	151.00
28	80.50	84.30	65	119.60	90.41	102	109.10	229.00
29	86.00	78.89	66	119.90	106.00	103	115.00	253.00
30	82.50	64.75	67	96.50	144.00	104	101.00	188.00
31	83.50	72.56	68	105.50	121.00	105	100.10	124.00
32	88.10	89.31	69	105.00	97.13	106	93.30	62.20
33	90.80	78.94	70	107.00	166.00	107	101.80	133.00
34	89.40	83.55	71	107.00	87.99	108	107.90	208.00
35	102.00	127.00	72	101.00	154.00	109	108.50	208.00
36	94.50	121.00	73	97.00	100.00			
37	91.00	107.00	74	100.00	123.00			

Source: Data provided courtesy of Jean-Pierre Després, Ph.D.

The Scatter Diagram

A first step that is usually useful in studying the relationship between two variables is to prepare a *scatter diagram* of the data such as is shown in Figure 9.3.1. The points are plotted by assigning values of the independent variable X to the horizontal axis and values of the dependent variable Y to the vertical axis.

The pattern made by the points plotted on the scatter diagram usually suggests the basic nature and strength of the relationship between two variables. As we look at Figure 9.3.1, for example, the points seem to be scattered around an invisible straight line. The scatter diagram also shows that, in general, subjects with large waist circumferences also have larger amounts of deep abdominal AT. These impressions suggest that the relationship between the two variables may be described by a straight line crossing the Y -axis below the origin and making approximately a 45-degree angle with the X -axis. It looks as if it would be simple to draw, freehand, through the data points the line that describes the relationship between X and Y . It is highly unlikely, however, that the lines drawn by any two people would be exactly the same. In other words, for every person drawing such a line by eye, or freehand, we would expect a slightly different line. The question then arises as to which line best describes the relationship between the two variables. We cannot obtain an answer to this question by inspecting the lines. In fact, it is not likely that any freehand line

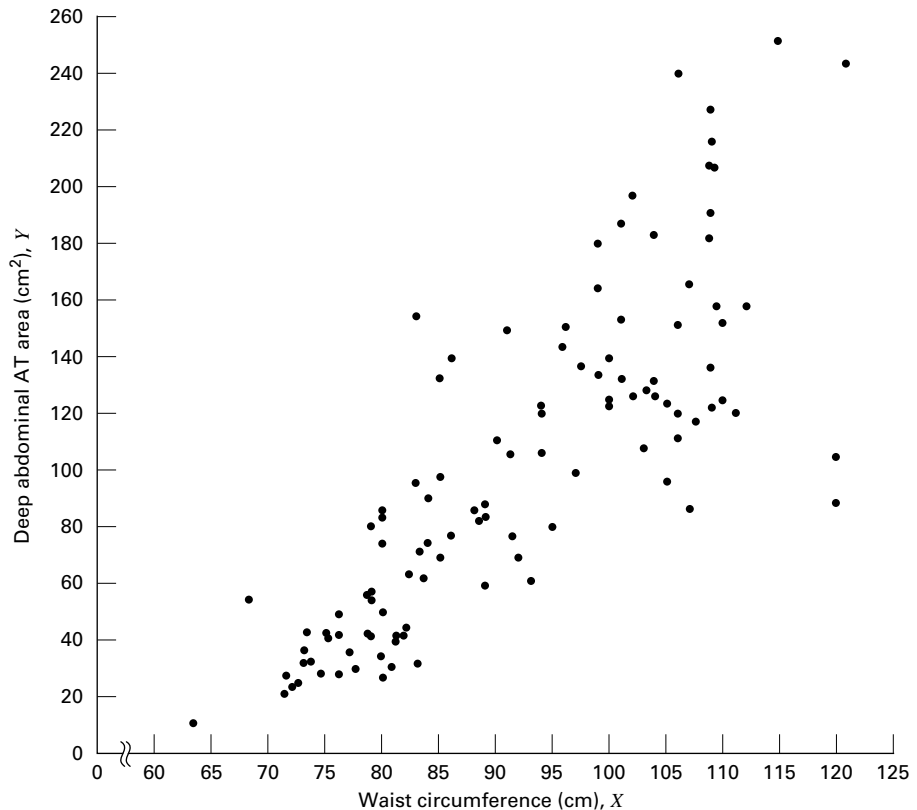


FIGURE 9.3.1 Scatter diagram of data shown in Table 9.3.1.

drawn through the data will be the line that best describes the relationship between X and Y , since freehand lines will reflect any defects of vision or judgment of the person drawing the line. Similarly, when judging which of two lines best describes the relationship, subjective evaluation is liable to the same deficiencies.

What is needed for obtaining the desired line is some method that is not fraught with these difficulties.

The Least-Squares Line

The method commonly employed for obtaining the desired line is known as the *method of least squares*, and the resulting line is called the *least-squares line*. The reason for calling the method by this name will be explained in the discussion that follows.

We recall from algebra that the general equation for a straight line may be written as

$$y = a + bx \quad (9.3.1)$$

where y is a value on the vertical axis, x is a value on the horizontal axis, a is the point where the line crosses the vertical axis, and b shows the amount by which y changes for each unit change in x . We refer to a as the y -intercept and b as the *slope* of the line. To draw a line based on Equation 9.3.1, we need the numerical values of the constants a and b . Given these constants, we may substitute various values of x into the equation to obtain corresponding values of y . The resulting points may be plotted. Since any two such coordinates determine a straight line, we may select any two, locate them on a graph, and connect them to obtain the line corresponding to the equation.

Obtaining the Least-Square Line

The least-squares regression line equation may be obtained from sample data by simple arithmetic calculations that may be carried out by hand using the following equations

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (9.3.2)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (9.3.3)$$

where x_i and y_i are the corresponding values of each data point (X , Y), \bar{x} and \bar{y} are the means of the X and Y sample data values, respectively, and $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimates of the intercept β_0 and slope β_1 , respectively, of the population regression line. Since the necessary hand calculations are time consuming, tedious, and subject to error, the regression line equation is best obtained through the use of a computer software package. Although the typical researcher need not be concerned with the arithmetic involved, the interested reader will find them discussed in references listed at the end of this chapter.

For the data in Table 9.3.1 we obtain the least-squares regression equation by means of MINITAB. After entering the X values in Column 1 and the Y values in Column 2 we proceed as shown in Figure 9.3.2.

For now, the only information from the output in Figure 9.3.2 that we are interested in is the regression equation. Other information in the output will be discussed later.

Dialog box:

Stat > Regression > Regression
 Type y in **Response** and x in **Predictors**.
 Click **Storage**. Check **Residuals** and **Fits**.
 Click **OK**.

Session command:

```
MTB > Name C3 = 'FITS1' C4 = 'RESI1'
MTB > Regress 'y' 1 'x';
SUBC> Fits 'FITS1';
SUBC> Constant;
SUBC> Residuals 'RESI1'.
```

Output:**Regression Analysis: y versus x**

The regression equation is
 $y = -216 + 3.46 x$

Predictor	Coef	Stdev	t-ratio	p
Constant	-215.98	21.80	-9.91	0.000
x	3.4589	0.2347	14.74	0.000

$s = 33.06$ $R\text{-sq} = 67.0\%$ $R\text{-sq(adj)} = 66.7\%$

Analysis of Variance

SOURCE	DF	SS	MS	F	p
Regression	1	237549	237549	217.28	0.000
Error	107	116982	1093		
Total	108	354531			

Unusual Observations

Obs.	x	y	Fit	Stdev.Fit	Residual	St.Resid
58	86	155.00	82.52	3.43	72.48	2.20R
65	120	90.41	197.70	7.23	-107.29	-3.33R
66	120	106.00	198.74	7.29	-92.74	-2.88R
71	107	87.99	154.12	4.75	-66.13	-2.02R
97	106	241.00	150.66	4.58	90.34	2.76R
102	109	229.00	161.38	5.13	67.62	2.07R
103	115	253.00	181.79	6.28	71.21	2.19R

R denotes an obs. with a large st. resid.

FIGURE 9.3.2 MINITAB procedure and output for obtaining the least-squares regression equation from the data in Table 9.3.1.

From Figure 9.3.2 we see that the linear equation for the least-squares line that describes the relationship between waist circumference and deep abdominal AT may be written, then, as

$$\hat{y} = -216 + 3.46x$$

This equation tells us that since $\hat{\beta}_0$ is negative, the line crosses the Y -axis below the origin, and that since $\hat{\beta}_1$ the slope, is positive, the line extends from the lower left-hand corner of the graph to the upper right-hand corner. We see further that for each unit increase in x , y increases by an amount equal to 3.46. The symbol \hat{y} denotes a value of y computed from the equation, rather than an observed value of Y .

By substituting two convenient values of X into Equation 9.3.2, we may obtain the necessary coordinates for drawing the line. Suppose, first, we let $X = 70$ and obtain

$$\hat{y} = -216 + 3.46(70) = 26.2$$

If we let $X = 110$ we obtain

$$\hat{y} = -216 + 3.46(110) = 164$$

The line, along with the original data, is shown in Figure 9.3.3.

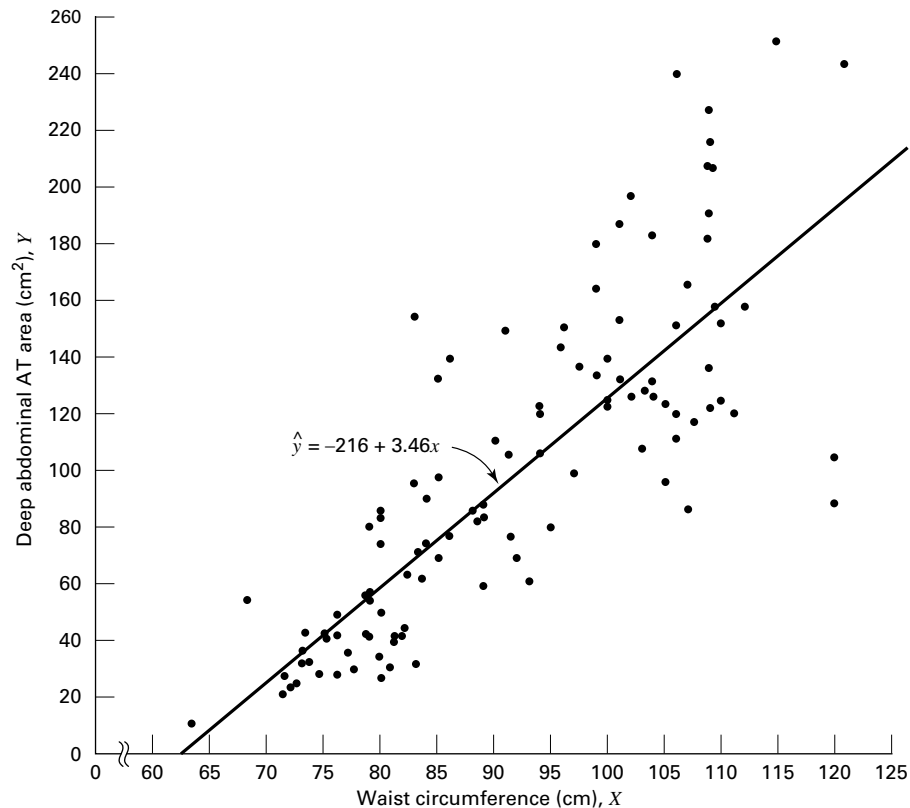


FIGURE 9.3.3 Original data and least-squares line for Example 9.3.1.

The Least-Squares Criterion Now that we have obtained what we call the “best fit” line for describing the relationship between our two variables, we need to determine by what criterion it is considered best. Before the criterion is stated, let us examine Figure 9.3.3. We note that generally the least-squares line does not pass through the observed points that are plotted on the scatter diagram. In other words, most of the observed points *deviate* from the line by varying amounts.

The line that we have drawn through the points is best in this sense:

The sum of the squared vertical deviations of the observed data points (y_i) from the least-squares line is smaller than the sum of the squared vertical deviations of the data points from any other line.

In other words, if we square the vertical distance from each observed point (y_i) to the least-squares line and add these squared values for all points, the resulting total will be smaller than the similarly computed total for any other line that can be drawn through the points. For this reason the line we have drawn is called the least-squares line.

EXERCISES

9.3.1 Plot each of the following regression equations on graph paper and state whether X and Y are directly or inversely related.

(a) $\hat{y} = -3 + 2x$

(b) $\hat{y} = 3 + 0.5x$

(c) $\hat{y} = 10 - 0.75x$

9.3.2 The following scores represent a nurse’s assessment (X) and a physician’s assessment (Y) of the condition of 10 patients at time of admission to a trauma center.

X :	18	13	18	15	10	12	8	4	7	3
Y :	23	20	18	16	14	11	10	7	6	4

(a) Construct a scatter diagram for these data.

(b) Plot the following regression equations on the scatter diagram and indicate which one you think best fits the data. State the reason for your choice.

(1) $\hat{y} = 8 + 0.5x$

(2) $\hat{y} = -10 + 2x$

(3) $y = 1 + 1x$

For each of the following exercises (a) draw a scatter diagram and (b) obtain the regression equation and plot it on the scatter diagram.

9.3.3 Methadone is often prescribed in the treatment of opioid addiction and chronic pain. Krantz et al. (A-2) studied the relationship between dose of methadone and the corrected QT (QT_c) interval for 17 subjects who developed *torsade de pointes* (ventricular tachycardia nearly always due to medications). QT_c is calculated from an electrocardiogram and is measured in mm/sec. A higher QT_c value indicates a higher risk of cardiovascular mortality. A question of interest is how well one can predict and estimate the QT_c value from a knowledge of methadone dose. This question is typical of those that can be answered by means of regression analysis. Since QT_c is the variable

about which we wish to make predictions and estimations, it is the dependent variable. The variable methadone dose, knowledge of which will be used to make the predictions and estimations, is the independent variable.

Methadone Dose		Methadone Dose	
(mg/day)	QTc (mm/sec)	(mg/day)	QTc (mm/sec)
1000	600	650	785
550	625	600	765
97	560	660	611
90	585	270	600
85	590	680	625
126	500	540	650
300	700	600	635
110	570	330	522
65	540		

Source: Mori J. Krantz, Ilana B. Kutinsky, Alastair D. Roberston, and Philip S. Mehler, "Dose-Related Effects of Methadone on QT Prolongation in a Series of Patients with Torsade de Pointes," *Pharmacotherapy*, 23 (2003), 802–805.

- 9.3.4** Reiss et al. (A-3) compared point-of-care and standard hospital laboratory assays for monitoring patients receiving a single anticoagulant or a regimen consisting of a combination of anticoagulants. It is quite common when comparing two measuring techniques, to use regression analysis in which one variable is used to predict another. In the present study, the researchers obtained measures of international normalized ratio (INR) by assay of capillary and venous blood samples collected from 90 subjects taking warfarin. INR, used especially when patients are receiving warfarin, measures the clotting ability of the blood. Point-of-care testing for INR was conducted with the CoaguChek assay product. Hospital testing was done with standard hospital laboratory assays. The authors used the hospital assay INR level to predict the CoaguChek INR level. The measurements are given in the following table.

CoaguChek (Y)	Hospital (X)	CoaguChek (Y)	Hospital (X)	CoaguChek (Y)	Hospital (X)
1.8	1.6	2.4	1.2	3.1	2.4
1.6	1.9	2.3	2.3	1.7	1.8
2.5	2.8	2.0	1.6	1.8	1.6
1.9	2.4	3.3	3.8	1.9	1.7
1.3	1.5	1.9	1.6	5.3	4.2
2.3	1.8	1.8	1.5	1.6	1.6
1.2	1.3	2.8	1.8	1.6	1.4
2.3	2.4	2.5	1.5	3.3	3.3
2.0	2.1	0.8	1.0	1.5	1.5
1.5	1.5	1.3	1.2	2.2	2.8
2.1	2.4	3.7	1.4	1.1	1.6
1.5	1.5	2.4	1.6	2.6	2.6

(Continued)

CoaguChek (Y)	Hospital (X)	CoaguChek (Y)	Hospital (X)	CoaguChek (Y)	Hospital (X)
1.5	1.7	4.1	3.2	6.4	5.0
1.8	2.1	2.4	1.2	1.5	1.4
1.0	1.2	2.3	2.3	3.0	2.8
2.1	1.9	3.1	1.6	2.6	2.3
1.6	1.6	1.5	1.4	1.2	1.2
1.7	1.6	3.6	2.1	2.1	1.9
2.0	1.9	2.5	1.7	1.1	1.1
1.8	1.6	2.1	1.7	1.0	1.0
1.3	4.1	1.8	1.2	1.4	1.5
1.5	1.9	1.5	1.3	1.7	1.3
3.6	2.1	2.5	1.1	1.2	1.1
2.4	2.2	1.5	1.2	2.5	2.4
2.2	2.3	1.5	1.1	1.2	1.3
2.7	2.2	1.6	1.2	2.5	2.9
2.9	3.1	1.4	1.4	1.9	1.7
2.0	2.2	4.0	2.3	1.8	1.7
1.0	1.2	2.0	1.2	1.2	1.1
2.4	2.6	2.5	1.5	1.3	1.1

Source: Data provided courtesy of Curtis E. Haas, Pharm.D.

- 9.3.5** Digoxin is a drug often prescribed to treat heart ailments. The purpose of a study by Parker et al. (A-4) was to examine the interactions of digoxin with common grapefruit juice. In one experiment, subjects took digoxin with water for 2 weeks, followed by a 2-week period during which digoxin was withheld. During the next 2 weeks subjects took digoxin with grapefruit juice. For seven subjects, the average peak plasma digoxin concentration (Cmax) when taking water is given in the first column of the following table. The second column contains the percent change in Cmax concentration when subjects were taking the digoxin with grapefruit juice [GFJ (%) change]. Use the Cmax level when taking digoxin with water to predict the percent change in Cmax concentration when taking digoxin with grapefruit juice.

Cmax (ngl/ml) with Water	Change in Cmax with GFJ (%)
2.34	29.5
2.46	40.7
1.87	5.3
3.09	23.3
5.59	-45.1
4.05	-35.3
6.21	-44.6
2.34	29.5

Source: Data provided courtesy of Robert B. Parker, Pharm.D.

- 9.3.6** Evans et al. (A-5) examined the effect of velocity on ground reaction forces (GRF) in dogs with lameness from a torn cranial cruciate ligament. The dogs were walked and trotted over a force

platform and the GRF recorded (in newtons) during the stance phase. The following table contains 22 measurements of force expressed as the mean of five force measurements per dog when walking and the mean of five force measurements per dog when trotting. Use the GRF value when walking to predict the GRF value when trotting.

GRF-Walk	GRF-Trot	GRF-Walk	GRF-Trot
31.5	50.8	24.9	30.2
33.3	43.2	33.6	46.3
32.3	44.8	30.7	41.8
28.8	39.5	27.2	32.4
38.3	44.0	44.0	65.8
36.9	60.1	28.2	32.2
14.6	11.1	24.3	29.5
27.0	32.3	31.6	38.7
32.8	41.3	29.9	42.0
27.4	38.2	34.3	37.6
31.5	50.8	24.9	30.2

Source: Data provided courtesy of Richard Evans, Ph.D.

- 9.3.7** Glomerular filtration rate (GFR) is the most important parameter of renal function assessed in renal transplant recipients. Although inulin clearance is regarded as the gold standard measure of GFR, its use in clinical practice is limited. Krieser et al. (A-6) examined the relationship between the inverse of Cystatin C (a cationic basic protein measured in mg/L) and inulin GFR as measured by technetium radionuclide labeled diethylenetriamine penta-acetic acid (DTPA GFR) clearance (ml/min/1.73 m²). The results of 27 tests are shown in the following table. Use DTPA GFR as the predictor of inverse Cystatin C.

DTPA GFR	1/Cystatin C	DTPA GFR	1/Cystatin C
18	0.213	42	0.485
21	0.265	42	0.427
21	0.446	43	0.562
23	0.203	43	0.463
27	0.369	48	0.549
27	0.568	48	0.538
30	0.382	51	0.571
32	0.383	55	0.546
32	0.274	58	0.402
32	0.424	60	0.592
36	0.308	62	0.541
37	0.498	67	0.568
41	0.398	68	0.800
		88	0.667

Source: Data provided courtesy of David Krieser, M.D.

9.4 EVALUATING THE REGRESSION EQUATION

Once the regression equation has been obtained it must be evaluated to determine whether it adequately describes the relationship between the two variables and whether it can be used effectively for prediction and estimation purposes.

When $H_0 : \beta_1 = 0$ Is Not Rejected If in the population the relationship between X and Y is linear, β_1 , the slope of the line that describes this relationship, will be either positive, negative, or zero. If β_1 is zero, sample data drawn from the population will, in the long run, yield regression equations that are of little or no value for prediction and estimation purposes. Furthermore, even though we assume that the relationship between X and Y is linear, it may be that the relationship could be described better by some nonlinear model. When this is the case, sample data when fitted to a linear model will tend to yield results compatible with a population slope of zero. Thus, following a test in which the null hypothesis that β_1 equals zero is not rejected, we may conclude (assuming that we have not made a type II error by accepting a false null hypothesis) either (1) that although the relationship between X and Y may be linear it is not strong enough for X to be of much value in predicting and estimating Y , or (2) that the relationship between X and Y is not linear; that is, some curvilinear model provides a better fit to the data. Figure 9.4.1 shows the kinds of relationships between X and Y in a population that may prevent rejection of the null hypothesis that $\beta_1 = 0$.

When $H_0 : \beta_1 = 0$ Is Rejected Now let us consider the situations in a population that may lead to rejection of the null hypothesis that $\beta_1 = 0$. Assuming that we do not commit a type I error, rejection of the null hypothesis that $\beta_1 = 0$ may be attributed to one of the following conditions in the population: (1) the relationship is linear and of sufficient strength to justify the use of sample regression equations to predict and estimate Y for given values of X ; and (2) there is a good fit of the data to a linear model, but some curvilinear model might provide an even better fit. Figure 9.4.2 illustrates the two population conditions that may lead to rejection of $H_0 : \beta_1 = 0$.

Thus, we see that before using a sample regression equation to predict and estimate, it is desirable to test $H_0 : \beta_1 = 0$. We may do this either by using analysis of variance and the F statistic or by using the t statistic. We will illustrate both methods. Before we do this, however, let us see how we may investigate the strength of the relationship between X and Y .

The Coefficient of Determination One way to evaluate the strength of the regression equation is to compare the scatter of the points about the regression line with the scatter about \bar{y} , the mean of the sample values of Y . If we take the scatter diagram for Example 9.3.1 and draw through the points a line that intersects the Y -axis at \bar{y} and is parallel to the X -axis, we may obtain a visual impression of the relative magnitudes of the scatter of the points about this line and the regression line. This has been done in Figure 9.4.3.

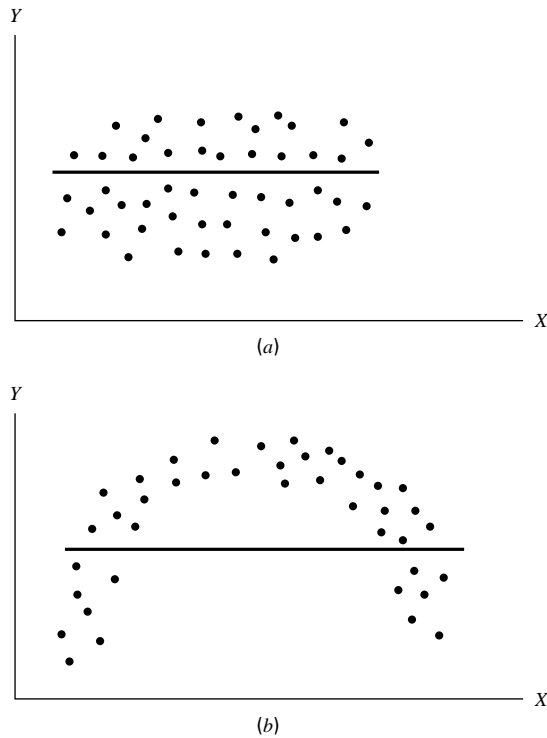


FIGURE 9.4.1 Conditions in a population that may prevent rejection of the null hypothesis that $\beta_1 = 0$. (a) The relationship between X and Y is linear, but β_1 is so close to zero that sample data are not likely to yield equations that are useful for predicting Y when X is given. (b) The relationship between X and Y is not linear; a curvilinear model provides a better fit to the data; sample data are not likely to yield equations that are useful for predicting Y when X is given.

It appears rather obvious from Figure 9.4.3 that the scatter of the points about the regression line is much less than the scatter about the \bar{y} line. We would not wish, however, to decide on this basis alone that the equation is a useful one. The situation may not be always this clear-cut, so that an objective measure of some sort would be much more desirable. Such an objective measure, called the *coefficient of determination*, is available.

The Total Deviation Before defining the coefficient of determination, let us justify its use by examining the logic behind its computation. We begin by considering the point corresponding to any observed value, y_i , and by measuring its vertical distance from the \bar{y} line. We call this the *total deviation* and designate it $(y_i - \bar{y})$.

The Explained Deviation If we measure the vertical distance from the regression line to the \bar{y} line, we obtain $(\hat{y}_i - \bar{y})$, which is called the *explained deviation*, since it shows by how much the total deviation is reduced when the regression line is fitted to the points.

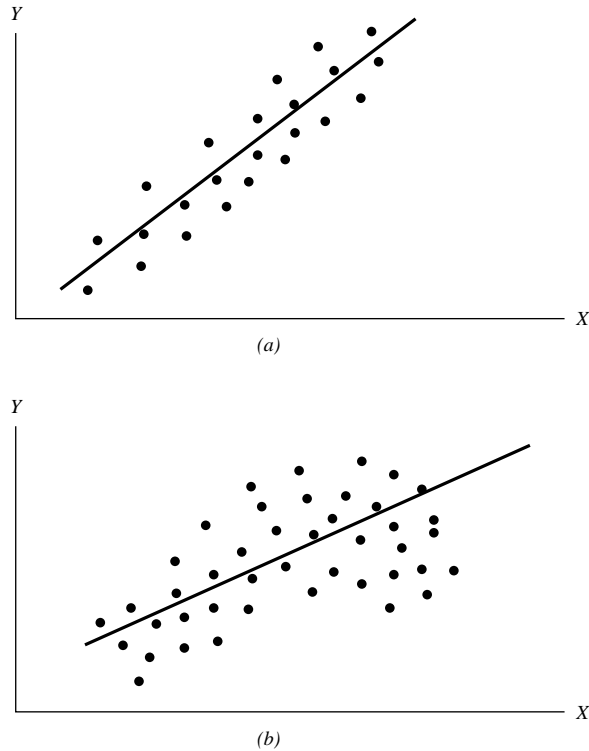


FIGURE 9.4.2 Population conditions relative to X and Y that may cause rejection of the null hypothesis that $\beta_1 = 0$. (a) The relationship between X and Y is linear and of sufficient strength to justify the use of a sample regression equation to predict and estimate Y for given values of X . (b) A linear model provides a good fit to the data, but some curvilinear model would provide an even better fit.

Unexplained Deviation Finally, we measure the vertical distance of the observed point from the regression line to obtain $(y_i - \hat{y}_i)$, which is called the *unexplained deviation*, since it represents the portion of the total deviation not “explained” or accounted for by the introduction of the regression line. These three quantities are shown for a typical value of Y in Figure 9.4.4. The difference between the observed value of Y and the predicted value of Y , $(y_i - \hat{y}_i)$, is also referred to as a *residual*. The set of residuals can be used to test the underlying linearity and equal-variances assumptions of the regression model described in Section 9.2. This procedure is illustrated at the end of this section.

It is seen, then, that the total deviation for a particular y_i is equal to the sum of the explained and unexplained deviations. We may write this symbolically as

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \quad (9.4.1)$$

total deviation	explained deviation	unexplained deviation
--------------------	------------------------	--------------------------

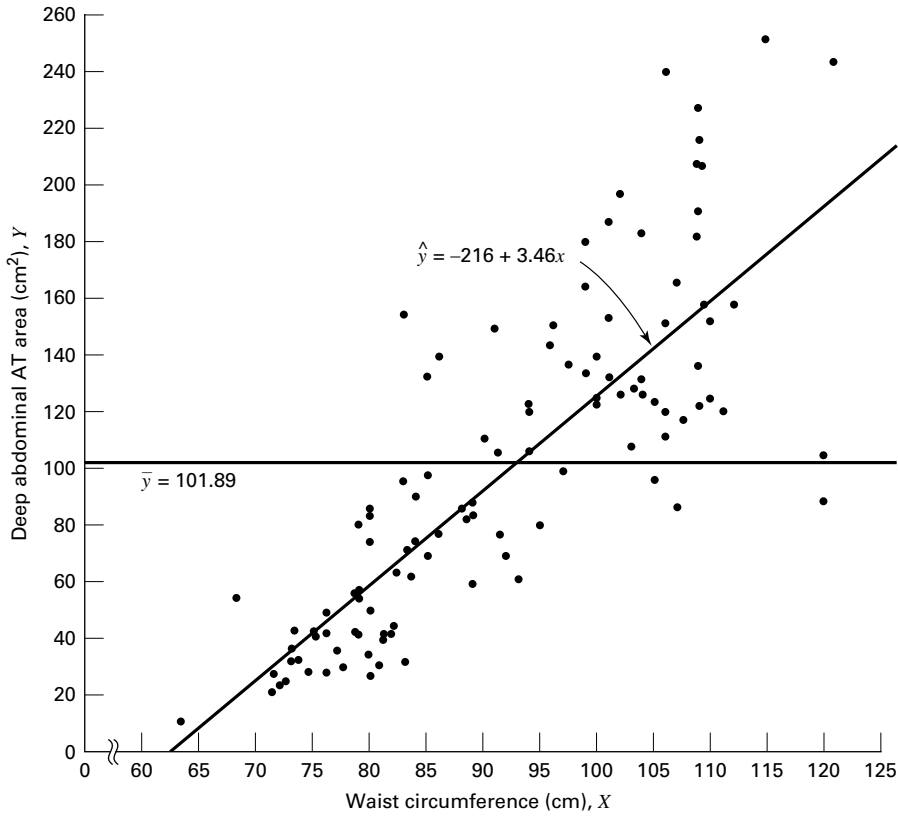


FIGURE 9.4.3 Scatter diagram, sample regression line, and \bar{y} line for Example 9.3.1.

If we measure these deviations for each value of y_i and \hat{y}_i , square each deviation, and add up the squared deviations, we have

$$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2 \quad (9.4.2)$$

total	explained	unexplained	
sum	sum	sum	
of squares	of squares	of squares	

These quantities may be considered measures of dispersion or variability.

Total Sum of Squares The *total sum of squares* (SST), for example, is a measure of the dispersion of the observed values of Y about their mean \bar{y} ; that is, this term is a measure of the total variation in the observed values of Y . The reader will recognize this term as the numerator of the familiar formula for the sample variance.

Explained Sum of Squares The *explained sum of squares* measures the amount of the total variability in the observed values of Y that is accounted for by the linear relationship between the observed values of X and Y . This quantity is referred to also as the *sum of squares due to linear regression* (SSR).

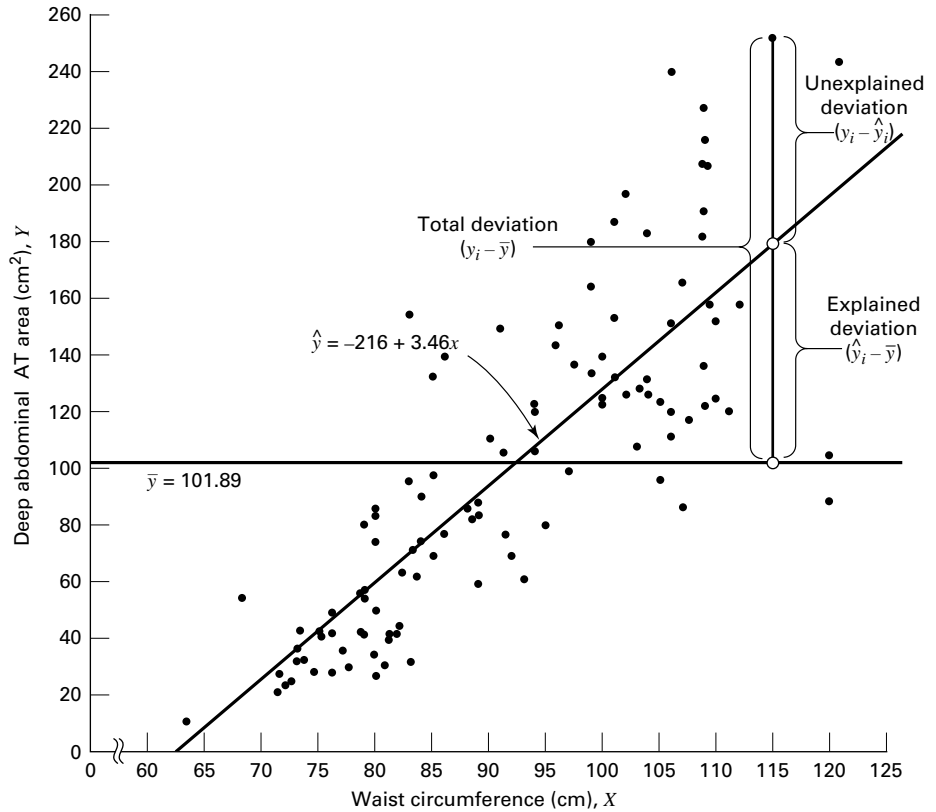


FIGURE 9.4.4 Scatter diagram showing the total, explained, and unexplained deviations for a selected value of Y , Example 9.3.1.

Unexplained Sum of Squares The *unexplained sum of squares* is a measure of the dispersion of the observed Y values about the regression line and is sometimes called the *error sum of squares*, or the *residual sum of squares* (SSE). It is this quantity that is minimized when the least-squares line is obtained.

We may express the relationship among the three sums of squares values as

$$SST = SSR + SSE$$

The numerical values of these sums of squares for our illustrative example appear in the analysis of variance table in Figure 9.3.2. Thus, we see that $SST = 354531$, $SSR = 237549$, $SSE = 116982$, and

$$354531 = 237549 + 116982$$

$$354531 = 354531$$

Calculating r^2 It is intuitively appealing to speculate that if a regression equation does a good job of describing the relationship between two variables, the explained or regression sum of squares should constitute a large proportion of the total sum of

squares. It would be of interest, then, to determine the magnitude of this proportion by computing the ratio of the explained sum of squares to the total sum of squares. This is exactly what is done in evaluating a regression equation based on sample data, and the result is called the sample *coefficient of determination*, r^2 . That is,

$$r^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{SSR}{SST}$$

In our present example we have, using the sums of squares values from Figure 9.3.2,

$$r^2 = \frac{237549}{354531} = .67$$

The sample coefficient of determination measures the closeness of fit of the sample regression equation to the observed values of Y . When the quantities $(y_i - \hat{y}_i)$, the vertical distances of the observed values of Y from the equations, are small, the unexplained sum of squares is small. This leads to a large explained sum of squares that leads, in turn, to a large value of r^2 . This is illustrated in Figure 9.4.5.

In Figure 9.4.5(a) we see that the observations all lie close to the regression line, and we would expect r^2 to be large. In fact, the computed r^2 for these data is .986, indicating that about 99 percent of the total variation in the y_i is explained by the regression.

In Figure 9.4.5(b) we illustrate a case in which the y_i are widely scattered about the regression line, and there we suspect that r^2 is small. The computed r^2 for the data is .403; that is, less than 50 percent of the total variation in the y_i is explained by the regression.

The largest value that r^2 can assume is 1, a result that occurs when all the variation in the y_i is explained by the regression. When $r^2 = 1$ all the observations fall on the regression line. This situation is shown in Figure 9.4.5(c).

The lower limit of r^2 is 0. This result is obtained when the regression line and the line drawn through \bar{y} coincide. In this situation none of the variation in the y_i is explained by the regression. Figure 9.4.5(d) illustrates a situation in which r^2 is close to zero.

When r^2 is large, then, the regression has accounted for a large proportion of the total variability in the observed values of Y , and we look with favor on the regression equation. On the other hand, a small r^2 which indicates a failure of the regression to account for a large proportion of the total variation in the observed values of Y , tends to cast doubt on the usefulness of the regression equation for predicting and estimating purposes. We do not, however, pass final judgment on the equation until it has been subjected to an objective statistical test.

Testing $H_0 : \beta_1 = 0$ with the F Statistic The following example illustrates one method for reaching a conclusion regarding the relationship between X and Y .

EXAMPLE 9.4.1

Refer to Example 9.3.1. We wish to know if we can conclude that, in the population from which our sample was drawn, X and Y are linearly related.

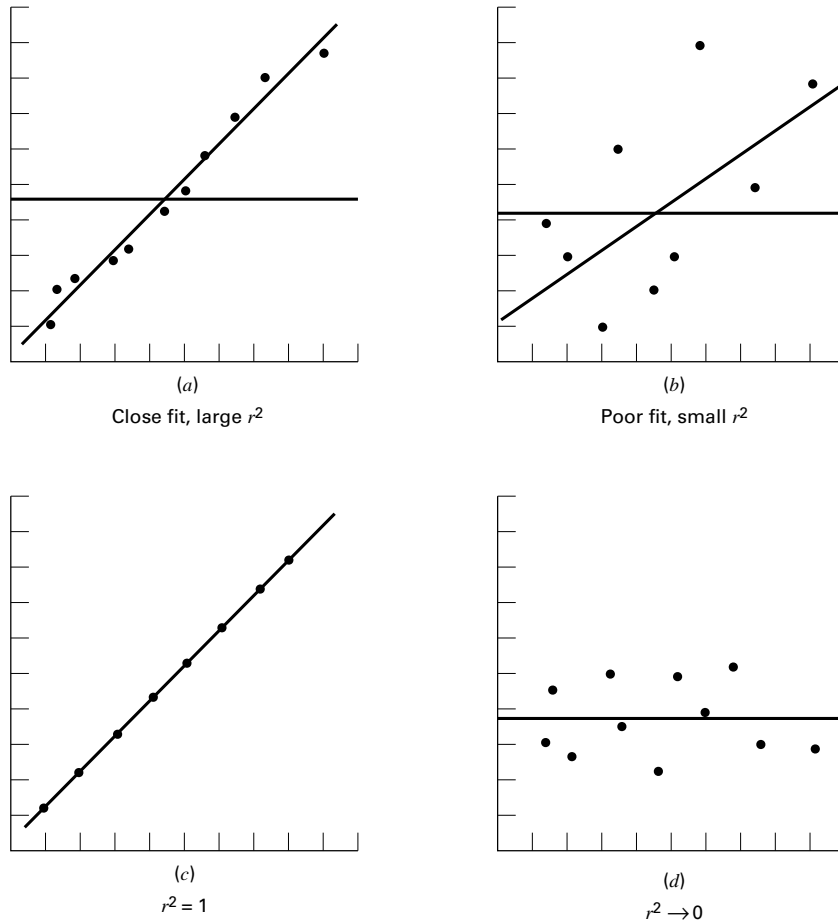


FIGURE 9.4.5 r^2 as a measure of closeness-of-fit of the sample regression line to the sample observations.

Solution: The steps in the hypothesis testing procedure are as follows:

1. **Data.** The data were described in the opening statement of Example 9.3.1.
2. **Assumptions.** We presume that the simple linear regression model and its underlying assumptions as given in Section 9.2 are applicable.
3. **Hypotheses.**

$$H_0 : \beta_1 = 0$$

$$H_A : \beta_1 \neq 0$$

$$\alpha = .05$$

TABLE 9.4.1 ANOVA Table for Simple Linear Regression

Source of Variation	SS	d.f.	MS	V.R.
Linear regression	SSR	1	$MSR = SSR/1$	MSR/MSE
Residual	SSE	$n - 2$	$MSE = SSE/(n - 2)$	
Total	SST	$n - 1$		

- 4. Test statistic.** The test statistic is V.R. as explained in the discussion that follows.

From the three sums-of-squares terms and their associated degrees of freedom the analysis of variance table of Table 9.4.1 may be constructed.

In general, the degrees of freedom associated with the sum of squares due to regression is equal to the number of constants in the regression equation minus 1. In the simple linear case we have two estimates, β_0 and β_1 ; hence the degrees of freedom for regression are $2 - 1 = 1$.

- 5. Distribution of test statistic.** It can be shown that when the hypothesis of no linear relationship between X and Y is true, and when the assumptions underlying regression are met, the ratio obtained by dividing the regression mean square by the residual mean square is distributed as F with 1 and $n - 2$ degrees of freedom.
- 6. Decision rule.** Reject H_0 if the computed value of V.R. is equal to or greater than the critical value of F .
- 7. Calculation of test statistic.** As shown in Figure 9.3.2, the computed value of F is 217.28.
- 8. Statistical decision.** Since 217.28 is greater than 3.94, the critical value of F (obtained by interpolation) for 1 and 107 degrees of freedom, the null hypothesis is rejected.
- 9. Conclusion.** We conclude that the linear model provides a good fit to the data.
- 10. p value.** For this test, since $217.28 > 8.25$, we have $p < .005$.
Examining Figure 9.3.2, we see that, in fact, $p < .001$. ■

Estimating the Population Coefficient of Determination The sample coefficient of determination provides a point estimate of ρ^2 the *population coefficient of determination*. The population coefficient of determination, ρ^2 has the same function relative to the population as r^2 has to the sample. It shows what proportion of the total population variation in Y is explained by the regression of Y on X . When the number of degrees of freedom is small, r^2 is positively biased. That is, r^2 tends to be large. An unbiased estimator of ρ^2 is provided by

$$\tilde{r}^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / (n - 2)}{\sum (y_i - \bar{y})^2 / (n - 1)} \quad (9.4.3)$$

Observe that the numerator of the fraction in Equation 9.4.3 is the unexplained mean square and the denominator is the total mean square. These quantities appear in the analysis of variance table. For our illustrative example we have, using the data from Figure 9.3.2,

$$\tilde{r}^2 = 1 - \frac{116982/107}{354531/108} = .66695$$

This quantity is labeled R-sq(adj) in Figure 9.3.2 and is reported as 66.7 percent. We see that this value is less than

$$r^2 = 1 - \frac{116982}{354531} = .67004$$

We see that the difference in r^2 and \tilde{r}^2 is due to the factor $(n-1)/(n-2)$. When n is large, this factor will approach 1 and the difference between r^2 and \tilde{r}^2 will approach zero.

Testing $H_0 : \beta_1 = 0$ with the t Statistic When the assumptions stated in Section 9.2 are met, $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased point estimators of the corresponding parameters β_0 and β_1 . Since, under these assumptions, the subpopulations of Y values are normally distributed, we may construct confidence intervals for and test hypotheses about β_0 and β_1 . When the assumptions of Section 9.2 hold true, the sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ are each normally distributed with means and variances as follows:

$$\mu_{\hat{\beta}_0} = \beta_0 \quad (9.4.4)$$

$$\sigma_{\hat{\beta}_0}^2 = \frac{\sigma_{y/x}^2 \sum x_i^2}{n \sum (x_i - \bar{x})^2} \quad (9.4.5)$$

$$\mu_{\hat{\beta}_1} = \beta_1 \quad (9.4.6)$$

and

$$\sigma_{\hat{\beta}_1}^2 = \frac{\sigma_{y/x}^2}{\sum (x_i - \bar{x})^2} \quad (9.4.7)$$

In Equations 9.4.5 and 9.4.7 $\sigma_{y/x}^2$ is the unexplained variance of the subpopulations of Y values.

With knowledge of the sampling distributions of $\hat{\beta}_0$ and $\hat{\beta}_1$ we may construct confidence intervals and test hypotheses relative to β_0 and β_1 in the usual manner. Inferences regarding α are usually not of interest. On the other hand, as we have seen, a great deal of interest centers on inferential procedures with respect to $\hat{\beta}_1$. The reason for this is the fact that β_1 tells us so much about the form of the relationship between X and Y . When X and Y are linearly related a positive $\hat{\beta}_1$ indicates that, in general, Y increases as X increases, and we say that there is a *direct linear relationship* between X and Y . A negative $\hat{\beta}_1$ indicates that values of Y tend to decrease as values of X increase, and we say

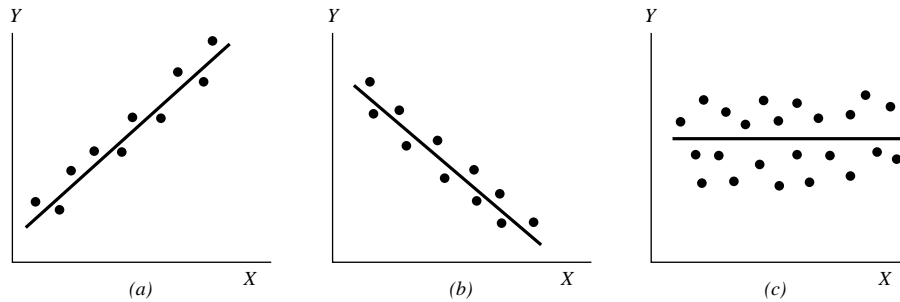


FIGURE 9.4.6 Scatter diagrams showing (a) direct linear relationship, (b) inverse linear relationship, and (c) no linear relationship between X and Y .

that there is an *inverse linear relationship* between X and Y . When there is no linear relationship between X and Y , β_1 is equal to zero. These three situations are illustrated in Figure 9.4.6.

The Test Statistic For testing hypotheses about β_1 the test statistic when $\sigma_{y|x}^2$ is known is

$$z = \frac{\hat{\beta}_1 - (\beta_1)_0}{\sigma_{\hat{\beta}_1}} \quad (9.4.8)$$

where $(\beta_1)_0$ is the hypothesized value of β_1 . The hypothesized value of β_1 does not have to be zero, but in practice, more often than not, the null hypothesis of interest is that $\beta_1 = 0$.

As a rule $\sigma_{y|x}^2$ is unknown. When this is the case, the test statistic is

$$t = \frac{\hat{\beta}_1 - (\beta_1)_0}{s_{\hat{\beta}_1}} \quad (9.4.9)$$

where $s_{\hat{\beta}_1}$ is an estimate of $\sigma_{\hat{\beta}_1}$ and t is distributed as Student's t with $n - 2$ degrees of freedom.

If the probability of observing a value as extreme as the value of the test statistic computed by Equation 9.4.9 when the null hypothesis is true is less than $\alpha/2$ (since we have a two-sided test), the null hypothesis is rejected.

EXAMPLE 9.4.2

Refer to Example 9.3.1. We wish to know if we can conclude that the slope of the population regression line describing the relationship between X and Y is zero.

Solution:

1. **Data.** See Example 9.3.1.
2. **Assumptions.** We presume that the simple linear regression model and its underlying assumptions are applicable.

3. Hypotheses.

$$H_0 : \beta_1 = 0$$

$$H_A : \beta_1 \neq 0$$

$$\alpha = .05$$

4. **Test statistic.** The test statistic is given by Equation 9.4.9.
5. **Distribution of test statistic.** When the assumptions are met and H_0 is true, the test statistic is distributed as Student's t with $n - 2$ degrees of freedom.
6. **Decision rule.** Reject H_0 if the computed value of t is either greater than or equal to 1.9826 or less than or equal to -1.9826 .
7. **Calculation of statistic.** The output in Figure 9.3.2 shows that $\hat{\beta}_1 = 3.4589$, $s_{\hat{\beta}_1} = .2347$, and

$$t = \frac{3.4589 - 0}{.2347} = 14.74$$

8. **Statistical decision.** Reject H_0 because $14.74 > 1.9826$.
9. **Conclusion.** We conclude that the slope of the true regression line is not zero.
10. **p value.** The p value for this test is less than .01, since, when H_0 is true, the probability of getting a value of t as large as or larger than 2.6230 (obtained by interpolation) is .005, and the probability of getting a value of t as small as or smaller than -2.6230 is also .005. Since 14.74 is greater than 2.6230, the probability of observing a value of t as large as or larger than 14.74 (when the null hypothesis is true) is less than .005. We double this value to obtain $2(.005) = .01$.

Either the F statistic or the t statistic may be used for testing $H_0 : \beta_1 = 0$. The value of the variance ratio is equal to the square of the value of the t statistic (i.e., $t^2 = F$) and, therefore, both statistics lead to the same conclusion. For the current example, we see that $(14.74)^2 = 217.27$, the value obtained by using the F statistic in Example 9.4.1. Hence, the corresponding p value will be the same for with the f statistic and the t statistic.

The practical implication of our results is that we can expect to get better predictions and estimates of Y if we use the sample regression equation than we would get if we ignore the relationship between X and Y . The fact that b is positive leads us to believe that β_1 is positive and that the relationship between X and Y is a direct linear relationship. ■

As has already been pointed out, Equation 9.4.9 may be used to test the null hypothesis that β_1 is equal to some value other than 0. The hypothesized value for β_1 , $(\beta_1)_0$ is substituted into Equation 9.4.9. All other quantities, as well as the computations, are the same as in the illustrative example. The degrees of freedom and the method of determining significance are also the same.

A Confidence Interval for β_1 Once we determine that it is unlikely, in light of sample evidence, that β_1 is zero, we may be interested in obtaining an interval estimate of β_1 . The general formula for a confidence interval,

$$\text{estimator} \pm (\text{reliability factor})(\text{standard error of the estimate})$$

may be used. When obtaining a confidence interval for β_1 , the estimator is $\hat{\beta}_1$, the reliability factor is some value of z or t (depending on whether or not $\sigma_{y|x}^2$ is known), and the standard error of the estimator is

$$\sigma_{\hat{\beta}_1} = \sqrt{\frac{\sigma_{y|x}^2}{\sum (x_i - \bar{x})^2}}$$

When $\sigma_{y|x}^2$ is unknown, σ_b is estimated by

$$s_{\hat{\beta}_1} = \sqrt{\frac{s_{y|x}^2}{\sum (x_i - \bar{x})^2}}$$

where $s_{y|x}^2 = MSE$

In most practical situations our $100(1 - \alpha)$ percent confidence interval for β is

$$\hat{\beta}_1 \pm t_{(1-\alpha/2)} s_{\hat{\beta}_1} \quad (9.4.10)$$

For our illustrative example we construct the following 95 percent confidence interval for β :

$$3.4589 \pm 1.9826(.2347) \\ (2.99, 3.92)$$

We interpret this interval in the usual manner. From the probabilistic point of view we say that in repeated sampling 95 percent of the intervals constructed in this way will include β_1 . The practical interpretation is that we are 95 percent confident that the single interval constructed includes β_1 .

Using the Confidence Interval to Test $H_0 : \beta_1 = 0$ It is instructive to note that the confidence interval we constructed does not include zero, so that zero is not a candidate for the parameter being estimated. We feel, then, that it is unlikely that $\beta_1 = 0$. This is compatible with the results of our hypothesis test in which we rejected the null hypothesis that $\beta_1 = 0$. Actually, we can always test $H_0 : \beta_1 = 0$ at the α significance level by constructing the $100(1 - \alpha)$ percent confidence interval for β_1 , and we can reject or fail to reject the hypothesis on the basis of whether or not the interval includes zero. If the interval contains zero, the null hypothesis is not rejected; and if zero is not contained in the interval, we reject the null hypothesis.

Interpreting the Results It must be emphasized that failure to reject the null hypothesis that $\beta_1 = 0$ does not mean that X and Y are not related. Not only is it possible that a type II error may have been committed but it may be true that X and Y are related in some nonlinear manner. On the other hand, when we reject the null hypothesis that $\beta_1 = 0$, we cannot conclude that the *true* relationship between X and Y is linear. Again, it may be

that although the data fit the linear regression model fairly well (as evidenced by the fact that the null hypothesis that $\beta_1 = 0$ is rejected), some nonlinear model would provide an even better fit. Consequently, when we reject H_0 that $\beta_1 = 0$, the best we can say is that more useful results (discussed below) may be obtained by taking into account the regression of Y on X than in ignoring it.

Testing the Regression Assumptions The values of the set of residuals, $(y_i - \hat{y}_i)$, for a data set are often used to test the linearity and equal-variances assumptions (assumptions 4 and 5 of Section 9.2) underlying the regression model. This is done by plotting the values of the residuals on the y -axis and the predicted values of y on the x -axis. If these plots show a relatively random scatter of points above and below a horizontal line at $(y_i - \hat{y}_i) = 0$, these assumptions are assumed to have been met for a given set of data. A non-random pattern of points can indicate violation of the linearity assumption, and a funnel-shaped pattern of the points can indicate violation of the equal-variances assumption. Examples of these patterns are shown in Figure 9.4.7.

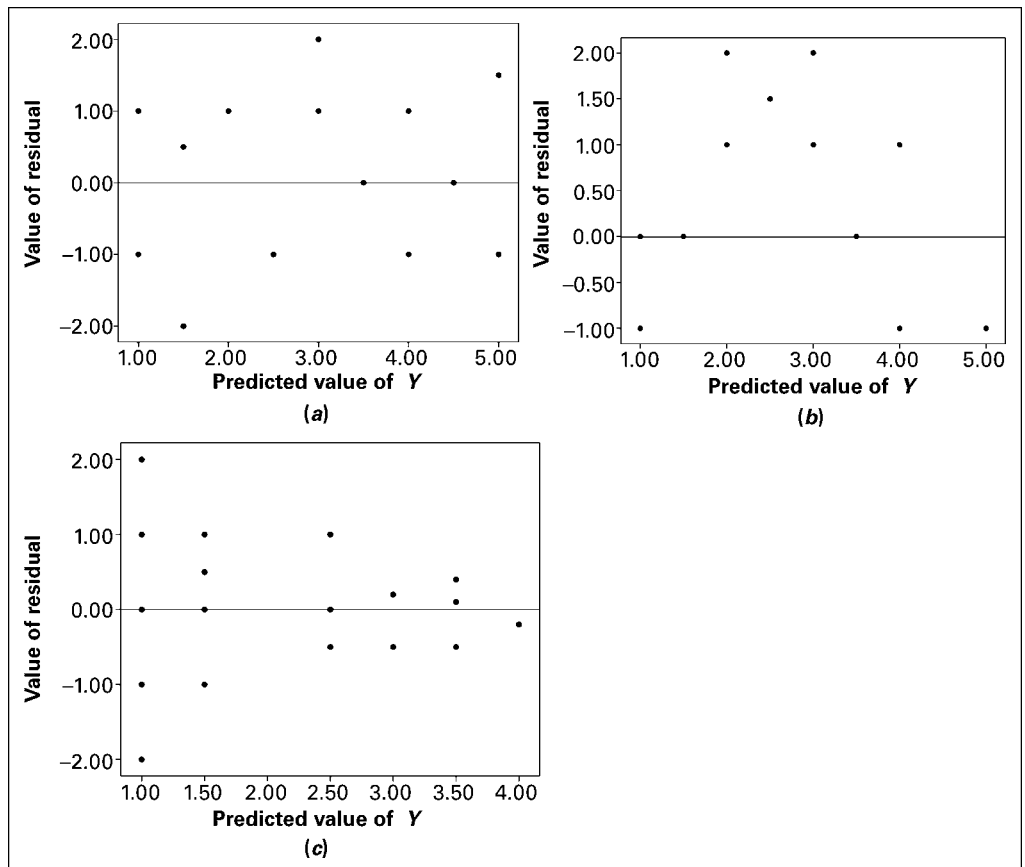


FIGURE 9.4.7 Residual plots useful for testing the linearity and equal-variances assumptions of the regression model. (a) A random pattern of points illustrating non-violation of the assumptions. (b) A non-random pattern illustrating a likely violation of the linearity assumption. (c) A funneling pattern illustrating a likely violation of the equal-variances assumption.

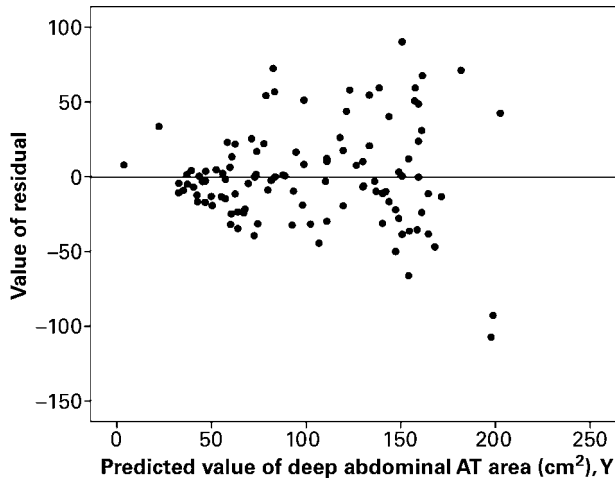


FIGURE 9.4.8 Residual plot of data from Example 9.3.1.

Many computer packages will provide residual plots automatically. These plots often use standardized values (i.e., e_i/\sqrt{MSE}) of the residuals and predicted values, but are interpreted in the same way as are plots of unstandardized values.

EXAMPLE 9.4.3

Refer to Example 9.3.1. We wish to use residual plots to test the assumptions of linearity and equal variances in the data.

Solution: A residual plot is shown in Figure 9.4.8.

Since there is a relatively equal and random scatter of points above and below the residual $(y_i - \hat{y}_i) = 0$ line, the linearity assumption is presumed to be valid. However, the funneling tendency of the plot suggests that as the predicted value of deep abdominal AT area increases, so does the amount of error. This indicates that the assumption of equal variances may not be valid for these data. ■

EXERCISES

9.4.1 to 9.4.5 Refer to Exercises 9.3.3 to 9.3.7, and for each one do the following:

- Compute the coefficient of determination.
- Prepare an ANOVA table and use the F statistic to test the null hypothesis that $\beta_1 = 0$. Let $\alpha = .05$.
- Use the t statistic to test the null hypothesis that $\beta_1 = 0$ at the .05 level of significance.
- Determine the p value for each hypothesis test.
- State your conclusions in terms of the problem.
- Construct the 95 percent confidence interval for β_1 .

9.5 USING THE REGRESSION EQUATION

If the results of the evaluation of the sample regression equation indicate that there is a relationship between the two variables of interest, we can put the regression equation to practical use. There are two ways in which the equation can be used. It can be used to *predict* what value Y is likely to assume given a particular value of X . When the normality assumption of Section 9.2 is met, a *prediction interval* for this predicted value of Y may be constructed.

We may also use the regression equation to *estimate* the mean of the subpopulation of Y values assumed to exist at any particular value of X . Again, if the assumption of normally distributed populations holds, a confidence interval for this parameter may be constructed. The predicted value of Y and the point estimate of the mean of the subpopulation of Y will be numerically equivalent for any particular value of X but, as we will see, the prediction interval will be wider than the confidence interval.

Predicting Y for a Given X If it is known, or if we are willing to assume that the assumptions of Section 9.2 are met, and when $\sigma_{y|x}^2$ is unknown, then the $100(1 - \alpha)$ percent prediction interval for Y is given by

$$\hat{y} \pm t_{(1-\alpha/2), S_{y|x}} \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \quad (9.5.1)$$

where x_p is the particular value of x at which we wish to obtain a prediction interval for Y and the degrees of freedom used in selecting t are $n - 2$.

Estimating the Mean of Y for a Given X The $100(1 - \alpha)$ percent confidence interval for $\mu_{y|x}$, when $\sigma_{y|x}^2$ is unknown, is given by

$$\hat{y} \pm t_{(1-\alpha/2), S_{y|x}} \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}} \quad (9.5.2)$$

We use MINITAB to illustrate, for a specified value of X , the calculation of a 95 percent confidence interval for the mean of Y and a 95 percent prediction interval for an individual Y measurement.

Suppose, for our present example, we wish to make predictions and estimates about AT for a waist circumference of 100 cm. In the regression dialog box click on “Options.” Enter 100 in the “Prediction interval for new observations” box. Click on “Confidence limits,” and click on “Prediction limits.”

We obtain the following output:

Fit	Stdev.Fit	95.0% C.I.	95.0% P.I.
129.90	3.69	(122.58, 137.23)	(63.93, 195.87)

We interpret the 95 percent confidence interval (C.I.) as follows.

If we repeatedly drew samples from our population of men, performed a regression analysis, and estimated $\mu_{y|x=100}$ with a similarly constructed confidence interval, about 95 percent of such intervals would include the mean amount of deep abdominal AT for the population. For this reason we are 95 percent confident that the single interval constructed contains the population mean and that it is somewhere between 122.58 and 137.23.

Our interpretation of a prediction interval (P.I.) is similar to the interpretation of a confidence interval. If we repeatedly draw samples, do a regression analysis, and construct prediction intervals for men who have a waist circumference of 100 cm, about 95 percent of them will include the man's deep abdominal AT value. This is the probabilistic interpretation. The practical interpretation is that we are 95 percent confident that a man who has a waist circumference of 100 cm will have a deep abdominal AT area of somewhere between 63.93 and 195.87 square centimeters.

Simultaneous confidence intervals and prediction intervals can be calculated for all possible points along a fitted regression line. Plotting lines through these points will then provide a graphical representation of these intervals. Since the mean data point (\bar{X}, \bar{Y}) is always included in the regression equation, as illustrated by equations 9.3.2 and 9.3.3, plots of the simultaneous intervals will always provide the best estimates at the middle of the line and the error will increase toward the ends of the line. This illustrates the fact that estimation within the bounds of the data set, called *interpolation*, is acceptable, but that estimation outside of the bounds of the data set, called *extrapolation*, is not advisable since the prediction error can be quite large. See Figure 9.5.1.

Figure 9.5.2 contains a partial printout of the SAS[®] simple linear regression analysis of the data of Example 9.3.1.

Resistant Line Frequently, data sets available for analysis by linear regression techniques contain one or more “unusual” observations; that is, values of x or y , or both, may be either considerably larger or considerably smaller than most of the other measurements. In the output of Figure 9.3.2, we see that the computer detected seven

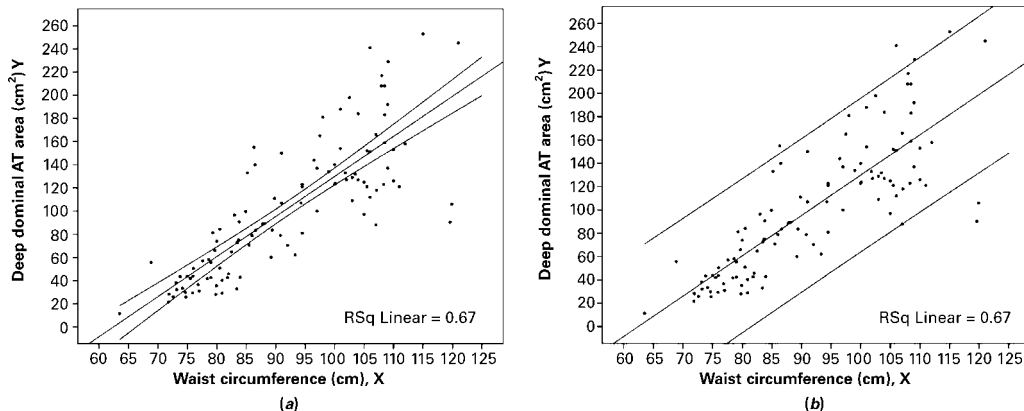


FIGURE 9.5.1 Simultaneous confidence intervals (a) and prediction intervals (b) for the data in Example 9.3.1.

The SAS System					
Model: MODEL1					
Dependent Variable: Y					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	237548.51620	237548.51620	217.279	0.0001
Error	107	116981.98602	1093.28959		
C Total	108	354530.50222			
Root MSE		33.06493	R-square	0.6700	
Dep Mean		101.89404	Adj R-sq	0.6670	
C.V.		32.45031			
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter = 0	Prob > T
INTERCEP	1	-215.981488	21.79627076	-9.909	0.0001
X	1	3.458859	0.23465205	14.740	0.0001

FIGURE 9.5.2 Partial printout of the computer analysis of the data given in Example 9.3.1, using the SAS® software package.

unusual observations in the waist circumference and deep abdominal AT data shown in Table 9.3.1.

The least-squares method of fitting a straight line to data is sensitive to unusual observations, and the location of the fitted line can be affected substantially by them. Because of this characteristic of the least-squares method, the resulting least-squares line is said to lack *resistance* to the influence of unusual observations. Several methods have been devised for dealing with this problem, including one developed by John W. Tukey. The resulting line is variously referred to as *Tukey's line* and the *resistant line*.

Based on medians, which, as we have seen, are descriptive measures that are themselves resistant to extreme values, the resistant line methodology is an exploratory data analysis tool that enables the researcher to quickly fit a straight line to a set of data consisting of paired x, y measurements. The technique involves partitioning, on the basis of the independent variable, the sample measurements into three groups of as near equal size as possible: the smallest measurements, the largest measurements, and those in between. The resistant line is the line fitted in such a way that there are an equal number of values

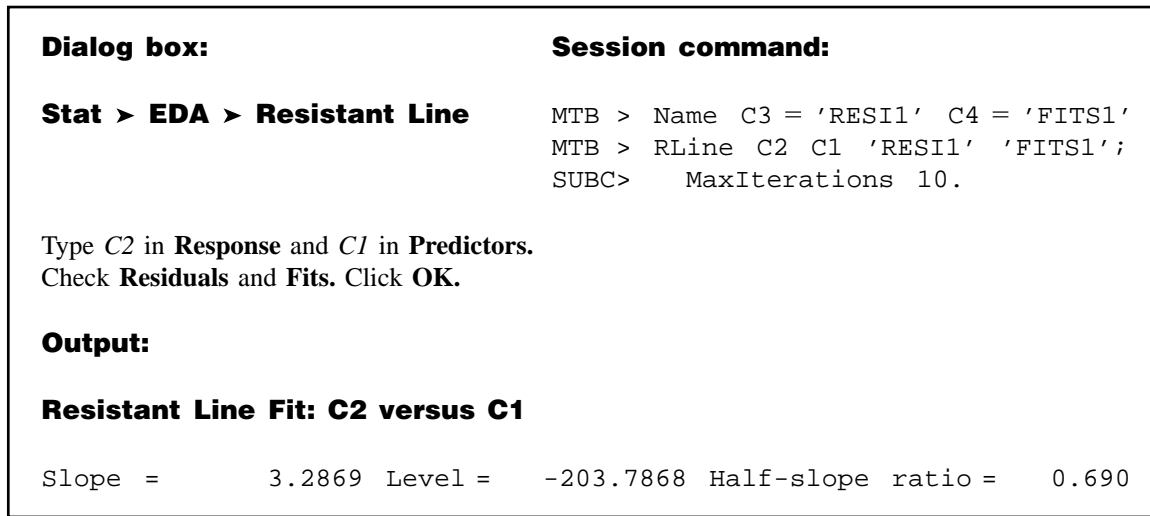


FIGURE 9.5.3 MINITAB resistant line procedure and output for the data of Table 9.3.1.

above and below it in both the smaller group and the larger group. The resulting slope and y -intercept estimates are resistant to the effects of either extreme y values, extreme x values, or both. To illustrate the fitting of a resistant line, we use the data of Table 9.3.1 and MINITAB. The procedure and output are shown in Figure 9.5.3.

We see from the output in Figure 9.5.3 that the resistant line has a slope of 3.2869 and a y -intercept of -203.7868 . The *half-slope ratio*, shown in the output as equal to .690, is an indicator of the degree of linearity between x and y . A slope, called a half-slope, is computed for each half of the sample data. The ratio of the right half-slope, b_R , and the left half-slope, b_L , is equal to b_R/b_L . If the relationship between x and y is straight, the half-slopes will be equal, and their ratio will be 1. A half-slope ratio that is not close to 1 indicates a lack of linearity between x and y .

The resistant line methodology is discussed in more detail by Hartwig and Dearing (1), Johnstone and Velleman (2), McNeil (3), and Velleman and Hoaglin (4).

EXERCISES

In each exercise refer to the appropriate previous exercise and, for the value of X indicated, (a) construct the 95 percent confidence interval for $\mu_{y|x}$ and (b) construct the 95 percent prediction interval for Y .

- 9.5.1 Refer to Exercise 9.3.3 and let $X = 400$.
- 9.5.2 Refer to Exercise 9.3.4 and let $X = 1.6$.
- 9.5.3 Refer to Exercise 9.3.5 and let $X = 4.16$.
- 9.5.4 Refer to Exercise 9.3.6 and let $X = 29.4$.
- 9.5.5 Refer to Exercise 9.3.7 and let $X = 35$.

9.6 THE CORRELATION MODEL

In the classic regression model, which has been the underlying model in our discussion up to this point, only Y , which has been called the dependent variable, is required to be random. The variable X is defined as a fixed (nonrandom or mathematical) variable and is referred to as the independent variable. Recall, also, that under this model observations are frequently obtained by preselecting values of X and determining corresponding values of Y .

When both Y and X are random variables, we have what is called the *correlation model*. Typically, under the correlation model, sample observations are obtained by selecting a random sample of the *units of association* (which may be persons, places, animals, points in time, or any other element on which the two measurements are taken) and taking on each a measurement of X and a measurement of Y . In this procedure, values of X are not preselected but occur at random, depending on the unit of association selected in the sample.

Although correlation analysis cannot be carried out meaningfully under the classic regression model, regression analysis can be carried out under the correlation model. Correlation involving two variables implies a co-relationship between variables that puts them on an equal footing and does not distinguish between them by referring to one as the dependent and the other as the independent variable. In fact, in the basic computational procedures, which are the same as for the regression model, we may fit a straight line to the data either by minimizing $\sum (y_i - \hat{y}_i)^2$ or by minimizing $\sum (x_i - \hat{x}_i)^2$. In other words, we may do a regression of X on Y as well as a regression of Y on X . The fitted line in the two cases in general will be different, and a logical question arises as to which line to fit.

If the objective is solely to obtain a measure of the strength of the relationship between the two variables, it does not matter which line is fitted, since the measure usually computed will be the same in either case. If, however, it is desired to use the equation describing the relationship between the two variables for the purposes discussed in the preceding sections, it does matter which line is fitted. The variable for which we wish to estimate means or to make predictions should be treated as the dependent variable; that is, this variable should be regressed on the other variable.

The Bivariate Normal Distribution Under the correlation model, X and Y are assumed to vary together in what is called a *joint distribution*. If this joint distribution is a normal distribution, it is referred to as a *bivariate normal distribution*. Inferences regarding this population may be made based on the results of samples properly drawn from it. If, on the other hand, the form of the joint distribution is known to be nonnormal, or if the form is unknown and there is no justification for assuming normality, inferential procedures are invalid, although descriptive measures may be computed.

Correlation Assumptions The following assumptions must hold for inferences about the population to be valid when sampling is from a bivariate distribution.

1. For each value of X there is a normally distributed subpopulation of Y values.
2. For each value of Y there is a normally distributed subpopulation of X values.
3. The joint distribution of X and Y is a normal distribution called the *bivariate normal distribution*.

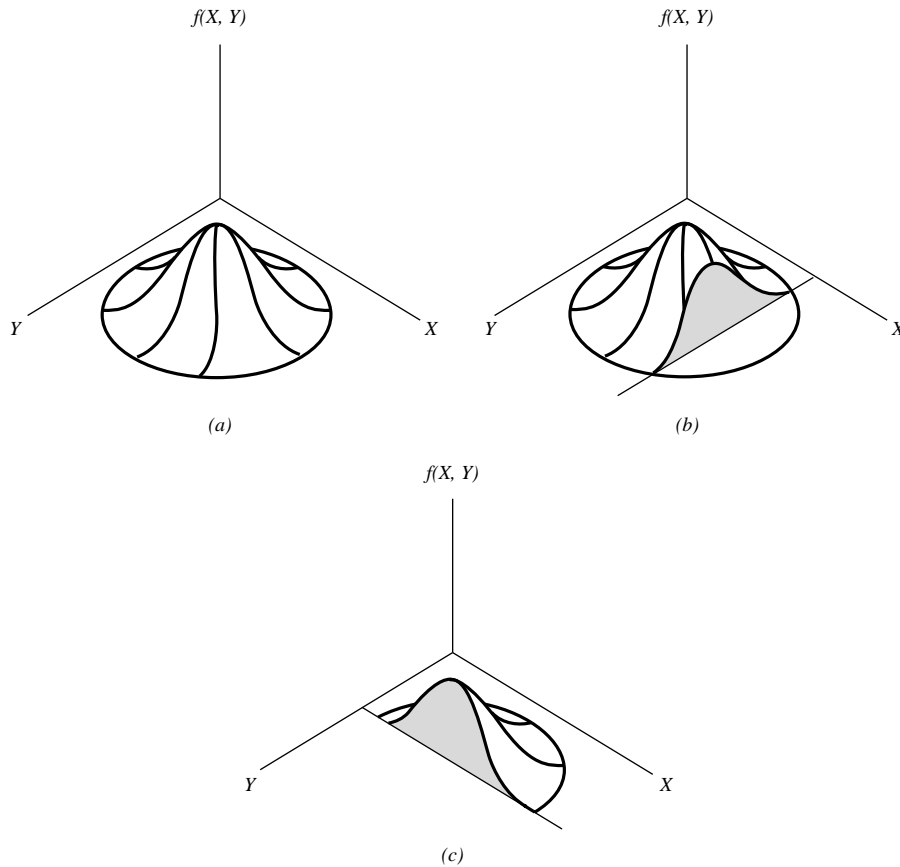


FIGURE 9.6.1 A bivariate normal distribution. (a) A bivariate normal distribution. (b) A cutaway showing normally distributed subpopulation of Y for given X . (c) A cutaway showing normally distributed subpopulation of X for given Y .

4. The subpopulations of Y values all have the same variance.
5. The subpopulations of X values all have the same variance.

The bivariate normal distribution is represented graphically in Figure 9.6.1. In this illustration we see that if we slice the mound parallel to Y at some value of X , the cutaway reveals the corresponding normal distribution of Y . Similarly, a slice through the mound parallel to X at some value of Y reveals the corresponding normally distributed subpopulation of X .

9.7 THE CORRELATION COEFFICIENT

The bivariate normal distribution discussed in Section 9.6 has five parameters, σ_x , σ_y , μ_x , μ_y , and ρ . The first four are, respectively, the standard deviations and means associated with the individual distributions. The other parameter, ρ , is called the population

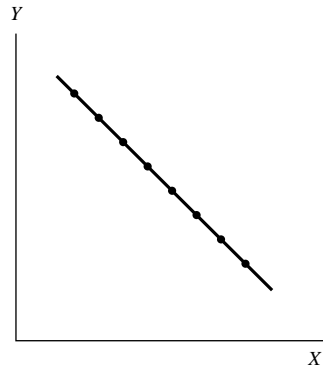


FIGURE 9.7.1 Scatter diagram for $r = -1$.

correlation coefficient and measures the strength of the linear relationship between X and Y .

The population correlation coefficient is the positive or negative square root of ρ^2 , the population coefficient of determination previously discussed, and since the coefficient of determination takes on values between 0 and 1 inclusive, ρ may assume any value between -1 and $+1$. If $\rho = 1$ there is a perfect direct linear correlation between the two variables, while $\rho = -1$ indicates perfect inverse linear correlation. If $\rho = 0$ the two variables are not linearly correlated. The sign of ρ will always be the same as the sign of β_1 , the slope of the population regression line for X and Y .

The sample correlation coefficient, r , describes the linear relationship between the sample observations on two variables in the same way that ρ describes the relationship in a population. The sample correlation coefficient is the square root of the sample coefficient of determination that was defined earlier.

Figures 9.4.5(d) and 9.4.5(c), respectively, show typical scatter diagrams where $r \rightarrow 0$ ($r^2 \rightarrow 0$) and $r = +1$ ($r^2 = 1$). Figure 9.7.1 shows a typical scatter diagram where $r = -1$.

We are usually interested in knowing if we may conclude that $\rho \neq 0$, that is, that X and Y are linearly correlated. Since ρ is usually unknown, we draw a random sample from the population of interest, compute r , the estimate of ρ , and test $H_0 : \rho = 0$ against the alternative $\rho \neq 0$. The procedure will be illustrated in the following example.

EXAMPLE 9.7.1

The purpose of a study by Kwast-Rabben et al. (A-7) was to analyze somatosensory evoked potentials (SEPs) and their interrelations following stimulation of digits I, III, and V in the hand. The researchers wanted to establish reference criteria in a control population. Thus, healthy volunteers were recruited for the study. In the future this information could be quite valuable as SEPs may provide a method to demonstrate functional disturbances in patients with suspected cervical root lesion who have pain and sensory symptoms. In the study, stimulation below-pain-level intensity was applied to the fingers. Recordings of spinal

responses were made with electrodes fixed by adhesive electrode cream to the subject's skin. One of the relationships of interest was the correlation between a subject's height (cm) and the peak spinal latency (Cv) of the SEP. The data for 155 measurements are shown in Table 9.7.1.

TABLE 9.7.1 Height and Spine SEP Measurements (Cv) from Stimulation of Digit I for 155 Subjects Described in Example 9.7.1

Height	Cv	Height	Cv	Height	Cv
149	14.4	168	16.3	181	15.8
149	13.4	168	15.3	181	18.8
155	13.5	168	16.0	181	18.6
155	13.5	168	16.6	182	18.0
156	13.0	168	15.7	182	17.9
156	13.6	168	16.3	182	17.5
157	14.3	168	16.6	182	17.4
157	14.9	168	15.4	182	17.0
158	14.0	170	16.6	182	17.5
158	14.0	170	16.0	182	17.8
160	15.4	170	17.0	184	18.4
160	14.7	170	16.4	184	18.5
161	15.5	171	16.5	184	17.7
161	15.7	171	16.3	184	17.7
161	15.8	171	16.4	184	17.4
161	16.0	171	16.5	184	18.4
161	14.6	172	17.6	185	19.0
161	15.2	172	16.8	185	19.6
162	15.2	172	17.0	187	19.1
162	16.5	172	17.6	187	19.2
162	17.0	173	17.3	187	17.8
162	14.7	173	16.8	187	19.3
163	16.0	174	15.5	188	17.5
163	15.8	174	15.5	188	18.0
163	17.0	175	17.0	189	18.0
163	15.1	175	15.6	189	18.8
163	14.6	175	16.8	190	18.3
163	15.6	175	17.4	190	18.6
163	14.6	175	17.6	190	18.8
164	17.0	175	16.5	190	19.2
164	16.3	175	16.6	191	18.5
164	16.0	175	17.0	191	18.5
164	16.0	176	18.0	191	19.0
165	15.7	176	17.0	191	18.5
165	16.3	176	17.4	194	19.8

(Continued)

Height	Cv	Height	Cv	Height	Cv
165	17.4	176	18.2	194	18.8
165	17.0	176	17.3	194	18.4
165	16.3	177	17.2	194	19.0
166	14.1	177	18.3	195	18.0
166	14.2	179	16.4	195	18.2
166	14.7	179	16.1	196	17.6
166	13.9	179	17.6	196	18.3
166	17.2	179	17.8	197	18.9
167	16.7	179	16.1	197	19.2
167	16.5	179	16.0	200	21.0
167	14.7	179	16.0	200	19.2
167	14.3	179	17.5	202	18.6
167	14.8	179	17.5	202	18.6
167	15.0	180	18.0	182	20.0
167	15.5	180	17.9	190	20.0
167	15.4	181	18.4	190	19.5
168	17.3	181	16.4		

Source: Data provided courtesy of Olga Kwast-Rabben, Ph.D.

Solution: The scatter diagram and least-squares regression line are shown in Figure 9.7.2.

Let us assume that the investigator wishes to obtain a regression equation to use for estimating and predicting purposes. In that case the sample correlation coefficient will be obtained by the methods discussed under the regression model.

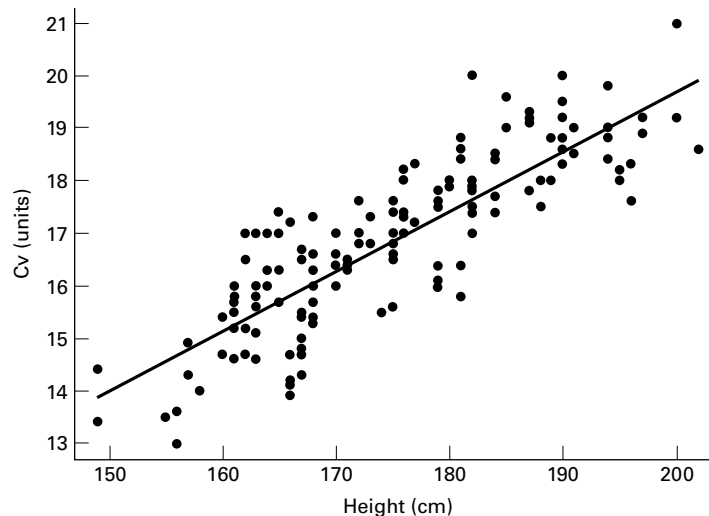


FIGURE 9.7.2 Height and cervical (spine) potentials in digit I stimulation for the data described in Example 9.7.1.

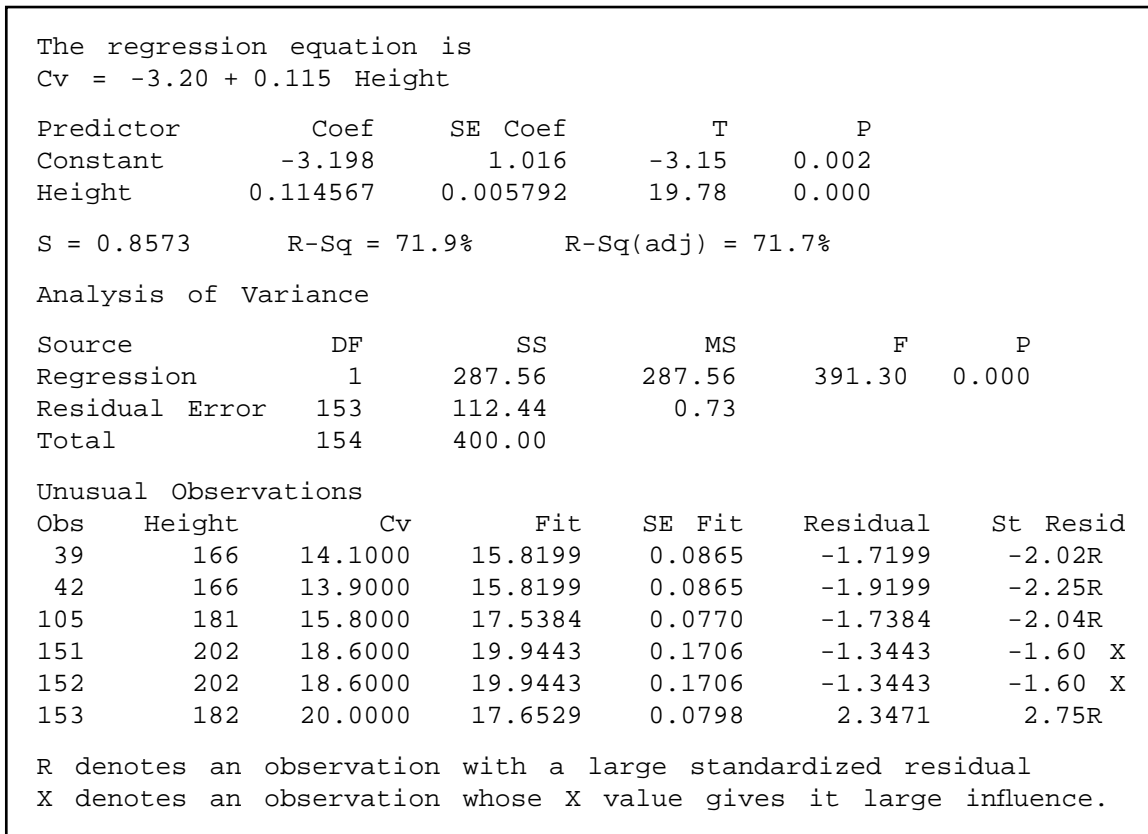


FIGURE 9.7.3 MINITAB output for Example 9.7.1 using the simple regression procedure.

The Regression Equation

Let us assume that we wish to predict Cv levels from knowledge of heights. In that case we treat height as the independent variable and Cv level as the dependent variable and obtain the regression equation and correlation coefficient with MINITAB as shown in Figure 9.7.3. For this example $r = \sqrt{.719} = .848$. We know that r is positive because the slope of the regression line is positive. We may also use the MINITAB correlation procedure to obtain r as shown in Figure 9.7.4.

The printout from the SAS[®] correlation procedure is shown in Figure 9.7.5. Note that the SAS[®] procedure gives descriptive measures for each variable as well as the p value for the correlation coefficient.

When a computer is not available for performing the calculations, r may be obtained by means of the following formulas:

$$r = \sqrt{\frac{\hat{\beta}_1^2 [\sum x_i^2 - (\sum x_i)^2/n]}{\sum y_i^2 - (\sum y_i)^2/n}} \quad (9.7.1)$$

Data:

C1: Height
C2: Cv

Dialog Box:

Stat > Basic Statistics > Correlation

Type *C1 C2* in **Variables**. Click **OK**.

Session command:

MTB > Correlation C1 C2.

OUTPUT:

Correlations: Height, Cv

Pearson correlation of Height and Cv = 0.848
P-Value = 0.000

FIGURE 9.7.4 MINITAB procedure for Example 9.7.1 using the correlation command.

The CORR Procedure
2 Variables: HEIGHT CV

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
HEIGHT	155	175.04516	11.92745	27132	149.00000	202.00000
CV	155	16.85613	1.61165	2613	13.00000	21.00000

Pearson Correlation Coefficients, N = 155
Prob > |r| under H0: Rho=0

	HEIGHT	CV
HEIGHT	1.00000	0.84788 <.0001
CV	0.84788 <.0001	1.00000

FIGURE 9.7.5 SAS® printout for Example 9.7.1.

An alternative formula for computing r is given by

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (9.7.2)$$

An advantage of this formula is that r may be computed without first computing b . This is the desirable procedure when it is not anticipated that the regression equation will be used.

Remember that the sample correlation coefficient, r , will always have the same sign as the sample slope, b . ■

EXAMPLE 9.7.2

Refer to Example 9.7.1. We wish to see if the sample value of $r = .848$ is of sufficient magnitude to indicate that, in the population, height and Cv SEP levels are correlated.

Solution: We conduct a hypothesis test as follows.

1. **Data.** See the initial discussion of Example 9.7.1.
2. **Assumptions.** We presume that the assumptions given in Section 9.6 are applicable.
3. **Hypotheses.**

$$H_0 : \rho = 0$$

$$H_A : \rho \neq 0$$

4. **Test statistic.** When $\rho = 0$, it can be shown that the appropriate test statistic is

$$t = r \sqrt{\frac{n-2}{1-r^2}} \quad (9.7.3)$$

5. **Distribution of test statistic.** When H_0 is true and the assumptions are met, the test statistic is distributed as Student's t distribution with $n - 2$ degrees of freedom.
6. **Decision rule.** If we let $\alpha = .05$, the critical values of t in the present example are ± 1.9754 (by interpolation). If, from our data, we compute a value of t that is either greater than or equal to $+1.9754$ or less than or equal to -1.9754 , we will reject the null hypothesis.
7. **Calculation of test statistic.** Our calculated value of t is

$$t = .848 \sqrt{\frac{153}{1-.719}} = 19.787$$

8. **Statistical decision.** Since the computed value of the test statistic does exceed the critical value of t , we reject the null hypothesis.

9. Conclusion. We conclude that, in the population, height and SEP levels in the spine are linearly correlated.

10. p value. Since $t = 19.787 > 2.6085$ (interpolated value of t for 153, .995), we have for this test, $p < .005$. ■

One may also notice that the test statistic for the correlation coefficient is equivalent to the test statistic for the slope of the regression line. Hence, squaring the t statistic in solution step 7 results in the F statistic provided in Figure 9.7.3. This may be useful when using a computer package that does not routinely provide the t statistic for the correlation coefficient (e.g., SPSS) and one does not wish to calculate the test statistic by hand.

A Test for Use When the Hypothesized ρ Is a Nonzero Value The use of the t statistic computed in the above test is appropriate only for testing $H_0 : \rho = 0$. If it is desired to test $H_0 : \rho = \rho_0$, where ρ_0 is some value other than zero, we must use another approach. Fisher (5) suggests that r be transformed to z_r as follows:

$$z_r = \frac{1}{2} \ln \frac{1+r}{1-r} \quad (9.7.4)$$

where \ln is a natural logarithm. It can be shown that z_r is approximately normally distributed with a mean of $z_\rho = \frac{1}{2} \ln \{(1+\rho)/(1-\rho)\}$ and estimated standard deviation of

$$\sigma_{z_\rho} = \frac{1}{\sqrt{n-3}} \quad (9.7.5)$$

To test the null hypothesis that ρ is equal to some value other than zero, the test statistic is

$$Z = \frac{z_r - z_\rho}{1/\sqrt{n-3}} \quad (9.7.6)$$

which follows approximately the standard normal distribution.

To determine z_r for an observed r and z_ρ for a hypothesized ρ , we consult Table I, thereby avoiding the direct use of natural logarithms.

Suppose in our present example we wish to test

$$H_0 : \rho = .80$$

against the alternative

$$H_A : \rho \neq .80$$

at the .05 level of significance. By consulting Table I (and interpolating), we find that for

$$r = .848, \quad z_r = 1.24726$$

and for

$$\rho = .80, \quad z_\rho = 1.09861$$

Our test statistic, then, is

$$Z = \frac{1.24726 - 1.09861}{1/\sqrt{155-3}} = 1.83$$

Since 1.83 is less than the critical value of $z = 1.96$, we are unable to reject H_0 . We conclude that the population correlation coefficient may be .80.

For sample sizes less than 25, Fisher's Z transformation should be used with caution, if at all. An alternative procedure from Hotelling (6) may be used for sample sizes equal to or greater than 10. In this procedure the following transformation of r is employed:

$$z^* = z_r - \frac{3z_r + r}{4n} \quad (9.7.7)$$

The standard deviation of z^* is

$$\sigma_{z^*} = \frac{1}{\sqrt{n-1}} \quad (9.7.8)$$

The test statistic is

$$Z^* = \frac{z^* - \zeta^*}{1/\sqrt{n-1}} = (z^* - \zeta^*)\sqrt{n-1} \quad (9.7.9)$$

where

$$\zeta^* (\text{pronounced zeta}) = z_\rho - \frac{(3z_\rho + \rho)}{4n}$$

Critical values for comparison purposes are obtained from the standard normal distribution.

In our present example, to test $H_0 : \rho = .80$ against $H_A : \rho \neq .80$ using the Hotelling transformation and $\alpha = .05$, we have

$$z^* = 1.24726 - \frac{3(1.24726) + .848}{4(155)} = 1.2339$$

$$\zeta^* = 1.09861 - \frac{3(1.09861) + .8}{4(155)} = 1.0920$$

$$Z^* = (1.2339 - 1.0920)\sqrt{155 - 1} = 1.7609$$

Since 1.7609 is less than 1.96, the null hypothesis is not rejected, and the same conclusion is reached as when the Fisher transformation is used.

Alternatives In some situations the data available for analysis do not meet the assumptions necessary for the valid use of the procedures discussed here for testing hypotheses about a population correlation coefficient. In such cases it may be more appropriate to use the Spearman rank correlation technique discussed in Chapter 13.

Confidence Interval for ρ Fisher's transformation may be used to construct $100(1 - \alpha)$ percent confidence intervals for ρ . The general formula for a confidence interval

$$\text{estimator} \pm (\text{reliability factor})(\text{standard error})$$

is employed. We first convert our estimator, r , to z_r , construct a confidence interval about z_ρ , and then reconvert the limits to obtain a $100(1 - \alpha)$ percent confidence interval about ρ . The general formula then becomes

$$z_r \pm z \left(1/\sqrt{n-3} \right) \quad (9.7.10)$$

For our present example the 95 percent confidence interval for z_ρ is given by

$$1.24726 \pm 1.96(1/\sqrt{155-3}) \\ (1.08828, 1.40624)$$

Converting these limits (by interpolation in Appendix Table I), which are values of z_r , into values of r gives

z_r	r
1.08828	.7962
1.40624	.8866

We are 95 percent confident, then, that ρ is contained in the interval .7962 to .88866. Because of the limited entries in the table, these limits must be considered as only approximate.

EXERCISES

In each of the following exercises:

- Prepare a scatter diagram.
- Compute the sample correlation coefficient.
- Test $H_0 : \rho = 0$ at the .05 level of significance and state your conclusions.
- Determine the p value for the test.
- Construct the 95 percent confidence interval for ρ .

- 9.7.1** The purpose of a study by Brown and Persley (A-8) was to characterize acute hepatitis A in patients more than 40 years old. They performed a retrospective chart review of 20 subjects who were diagnosed with acute hepatitis A, but were not hospitalized. Of interest was the use of age (years) to predict bilirubin levels (mg/dl). The following data were collected.

Age (Years)	Bilirubin (mg/dl)	Age (Years)	Bilirubin (mg/dl)
78	7.5	44	7.0
72	12.9	42	1.8
81	14.3	45	.8
59	8.0	78	3.8
64	14.1	47	3.5
48	10.9	50	5.1
46	12.3	57	16.5

(Continued)

Age (Years)	Bilirubin (mg/dl)	Age (Years)	Bilirubin (mg/dl)
42	1.0	52	3.5
58	5.2	58	5.6
52	5.1	45	1.9

Source: Data provided courtesy of Geri R. Brown, M.D.

9.7.2 Another variable of interest in the study by Reiss et al. (A-3) (see Exercise 9.3.4) was partial thromboplastin (aPTT), the standard test used to monitor heparin anticoagulation. Use the data in the following table to examine the correlation between aPTT levels as measured by the CoaguCheck point-of-care assay and standard laboratory hospital assay in 90 subjects receiving heparin alone, heparin with warfarin, and warfarin and exoenoxaparin.

Heparin		Warfarin		Warfarin and Exoenoxaparin	
CoaguCheck aPTT	Hospital aPTT	CoaguCheck aPTT	Hospital aPTT	CoaguCheck aPTT	Hospital aPTT
49.3	71.4	18.0	77.0	56.5	46.5
57.9	86.4	31.2	62.2	50.7	34.9
59.0	75.6	58.7	53.2	37.3	28.0
77.3	54.5	75.2	53.0	64.8	52.3
42.3	57.7	18.0	45.7	41.2	37.5
44.3	59.5	82.6	81.1	90.1	47.1
90.0	77.2	29.6	40.9	23.1	27.1
55.4	63.3	82.9	75.4	53.2	40.6
20.3	27.6	58.7	55.7	27.3	37.8
28.7	52.6	64.8	54.0	67.5	50.4
64.3	101.6	37.9	79.4	33.6	34.2
90.4	89.4	81.2	62.5	45.1	34.8
64.3	66.2	18.0	36.5	56.2	44.2
89.8	69.8	38.8	32.8	26.0	28.2
74.7	91.3	95.4	68.9	67.8	46.3
150.0	118.8	53.7	71.3	40.7	41.0
32.4	30.9	128.3	111.1	36.2	35.7
20.9	65.2	60.5	80.5	60.8	47.2
89.5	77.9	150.0	150.0	30.2	39.7
44.7	91.5	38.5	46.5	18.0	31.3
61.0	90.5	58.9	89.1	55.6	53.0
36.4	33.6	112.8	66.7	18.0	27.4
52.9	88.0	26.7	29.5	18.0	35.7
57.5	69.9	49.7	47.8	78.3	62.0
39.1	41.0	85.6	63.3	75.3	36.7
74.8	81.7	68.8	43.5	73.2	85.3
32.5	33.3	18.0	54.0	42.0	38.3
125.7	142.9	92.6	100.5	49.3	39.8
77.1	98.2	46.2	52.4	22.8	42.3
143.8	108.3	60.5	93.7	35.8	36.0

Source: Data provided courtesy of Curtis E. Haas, Pharm.D.

- 9.7.3** In the study by Parker et al. (A-4) (see Exercise 9.3.5), the authors also looked at the change in AUC (area under the curve of plasma concentration of digoxin) when comparing digoxin levels taken with and without grapefruit juice. The following table gives the AUC when digoxin was consumed with water (ng·hr/ml) and the change in AUC compared to the change in AUC when digoxin is taken with grapefruit juice (GFJ, %).

Water AUC Level (ng · hr/ml)	Change in AUC with GFJ (%)
6.96	17.4
5.59	24.5
5.31	8.5
8.22	20.8
11.91	-26.7
9.50	-29.3
11.28	-16.8

Source: Data provided courtesy of Robert B. Parker, Pharm.D.

- 9.7.4** An article by Tuzson et al. (A-9) in *Archives of Physical Medicine and Rehabilitation* reported the following data on peak knee velocity in walking (measured in degrees per second) at flexion and extension for 18 subjects with cerebral palsy.

Flexion (°/s)	Extension (°/s)
100	100
150	150
210	180
255	165
200	210
185	155
440	440
110	180
400	400
160	140
150	250
425	275
375	340
400	400
400	450
300	300
300	300
320	275

Source: Ann E. Tuzson, Kevin P. Granata, and Mark F. Abel, "Spastic Velocity Threshold Constrains Functional Performance in Cerebral Palsy," *Archives of Physical Medicine and Rehabilitation*, 84 (2003), 1363–1368.

9.7.5 Amyotrophic lateral sclerosis (ALS) is characterized by a progressive decline of motor function. The degenerative process affects the respiratory system. Butz et al. (A-10) investigated the longitudinal impact of nocturnal noninvasive positive-pressure ventilation on patients with ALS. Prior to treatment, they measured partial pressure of arterial oxygen (P_{aO_2}) and partial pressure of arterial carbon dioxide (P_{aCO_2}) in patients with the disease. The results were as follows:

P_{aCO_2}	P_{aO_2}
40.0	101.0
47.0	69.0
34.0	132.0
42.0	65.0
54.0	72.0
48.0	76.0
53.6	67.2
56.9	70.9
58.0	73.0
45.0	66.0
54.5	80.0
54.0	72.0
43.0	105.0
44.3	113.0
53.9	69.2
41.8	66.7
33.0	67.0
43.1	77.5
52.4	65.1
37.9	71.0
34.5	86.5
40.1	74.7
33.0	94.0
59.9	60.4
62.6	52.5
54.1	76.9
45.7	65.3
40.6	80.3
56.6	53.2
59.0	71.9

Source: M. Butz, K. H. Wollinsky, U. Widemuth-Catrinescu, A. Sperfeld, S. Winter, H. H. Mehrkens, A. C. Ludolph, and H. Schreiber, "Longitudinal Effects of Noninvasive Positive-Pressure Ventilation in Patients with Amyotrophic Lateral Sclerosis," *American Journal of Medical Rehabilitation*, 82 (2003) 597–604.

9.7.6 A simple random sample of 15 apparently healthy children between the ages of 6 months and 15 years yielded the following data on age, X , and liver volume per unit of body weight (ml/kg), Y :

X	Y	X	Y
.5	41	10.0	26
.7	55	10.1	35
2.5	41	10.9	25
4.1	39	11.5	31

(Continued)

X	Y	X	Y
5.9	50	12.1	31
6.1	32	14.1	29
7.0	41	15.0	23
8.2	42		

9.8 SOME PRECAUTIONS

Regression and correlation analysis are powerful statistical tools when properly employed. Their inappropriate use, however, can lead only to meaningless results. To aid in the proper use of these techniques, we make the following suggestions:

1. The assumptions underlying regression and correlation analysis should be reviewed carefully before the data are collected. Although it is rare to find that assumptions are met to perfection, practitioners should have some idea about the magnitude of the gap that exists between the data to be analyzed and the assumptions of the proposed model, so that they may decide whether they should choose another model; proceed with the analysis, but use caution in the interpretation of the results; or use the chosen model with confidence.
2. In simple linear regression and correlation analysis, the two variables of interest are measured on the same entity, called the *unit of association*. If we are interested in the relationship between height and weight, for example, these two measurements are taken on the same individual. It usually does not make sense to speak of the correlation, say, between the heights of one group of individuals and the weights of another group.
3. No matter how strong is the indication of a relationship between two variables, it should not be interpreted as one of cause and effect. If, for example, a significant sample correlation coefficient between two variables X and Y is observed, it can mean one of several things:
 - (a) X causes Y .
 - (b) Y causes X .
 - (c) Some third factor, either directly or indirectly, causes both X and Y .
 - (d) An unlikely event has occurred and a large sample correlation coefficient has been generated by chance from a population in which X and Y are, in fact, not correlated.
 - (e) The correlation is purely nonsensical, a situation that may arise when measurements of X and Y are not taken on a common unit of association.
4. The sample regression equation should not be used to predict or estimate outside the range of values of the independent variable represented in the sample. As illustrated in Section 9.5, this practice, called extrapolation, is risky. The true relationship between two variables, although linear over an interval of the independent variable, sometimes may be described at best as a curve outside this interval. If our sample by chance is drawn only from the interval where the relationship is linear, we have only a

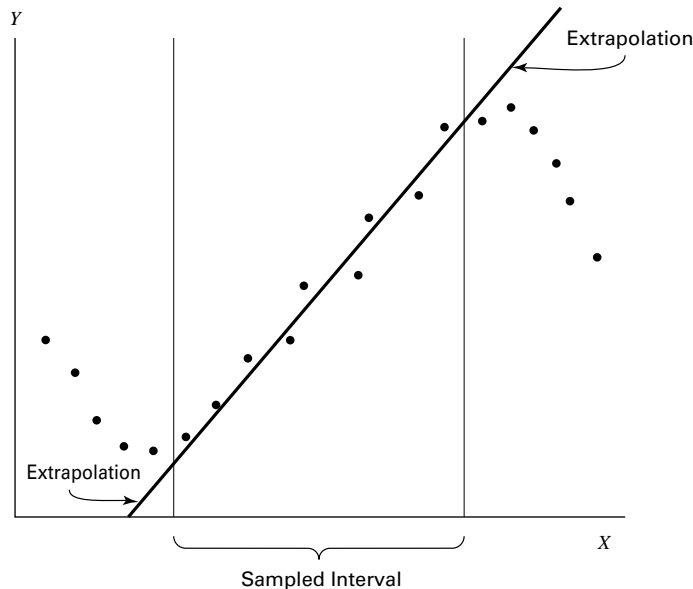


FIGURE 9.8.1 Example of extrapolation.

limited representation of the population, and to project the sample results beyond the interval represented by the sample may lead to false conclusions. Figure 9.8.1 illustrates the possible pitfalls of extrapolation.

9.9 SUMMARY

In this chapter, two important tools of statistical analysis, simple linear regression and correlation, are examined. The following outline for the application of these techniques has been suggested.

1. **Identify the model.** Practitioners must know whether the regression model or the correlation model is the appropriate one for answering their questions.
2. **Review assumptions.** It has been pointed out several times that the validity of the conclusions depends on how well the analyzed data fit the chosen model.
3. **Obtain the regression equation.** We have seen how the regression equation is obtained by the method of least squares. Although the computations, when done by hand, are rather lengthy, involved, and subject to error, this is not the problem today that it has been in the past. Computers are now in such widespread use that the researcher or statistician without access to one is the exception rather than the rule. No apology for lengthy computations is necessary to the researcher who has a computer available.
4. **Evaluate the equation.** We have seen that the usefulness of the regression equation for estimating and predicting purposes is determined by means of the analysis of

variance, which tests the significance of the regression mean square. The strength of the relationship between two variables under the correlation model is assessed by testing the null hypothesis that there is no correlation in the population. If this hypothesis can be rejected we may conclude, at the chosen level of significance, that the two variables are correlated.

5. Use the equation. Once it has been determined that it is likely that the regression equation provides a good description of the relationship between two variables, X and Y , it may be used for one of two purposes:

- (a) To predict what value Y is likely to assume, given a particular value of X , or
- (b) To estimate the mean of the subpopulation of Y values for a particular value of X .

This necessarily abridged treatment of simple linear regression and correlation may have raised more questions than it has answered. It may have occurred to the reader, for example, that a dependent variable can be more precisely predicted using two or more independent variables rather than one. Or, perhaps, he or she may feel that knowledge of the strength of the relationship among several variables might be of more interest than knowledge of the relationship between only two variables. The exploration of these possibilities is the subject of the next chapter, and the reader's curiosity along these lines should be at least partially relieved.

For those who would like to pursue further the topic of regression analysis a number of excellent references are available, including those by Dielman (7), Hocking (8), Mendenhall and Sincich (9), and Neter et al. (10).

SUMMARY OF FORMULAS FOR CHAPTER 9

Formula Number	Name	Formula
9.2.1	Assumption of linearity	$\mu_{y x} = \beta_0 + \beta_1 x$
9.2.2	Simple linear regression model	$y = \beta_0 + \beta_1 x + \epsilon$
9.2.3	Error (residual) term	$\epsilon = y - (\beta_0 + \beta_1 x) = y - \mu_{y x}$
9.3.1	Algebraic representation of a straight line	$y = a + bx$
9.3.2	Least square estimate of the slope of a regression line	$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$

(Continued)

9.3.3	Least square estimate of the intercept of a regression line	$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$
9.4.1	Deviation equation	$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$
9.4.2	Sum-of-squares equation	$\sum (y_i - \bar{y})^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2$
9.4.3	Estimated population coefficient of determination	$\hat{r}^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2 / (n - 2)}{\sum (y_i - \bar{y})^2 / (n - 1)}$
9.4.4–9.4.7	Means and variances of point estimators a and b	$\mu_{\hat{\beta}_0} = \beta_0$ $\sigma_{\hat{\beta}_0}^2 = \frac{\sigma_{y x}^2 \sum x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}$ $\mu_{\hat{\beta}_1} = \beta_1$ $\sigma_{\hat{\beta}_1}^2 = \frac{\sigma_{y x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$
9.4.8	z statistic for testing hypotheses about β	$z = \frac{\hat{\beta}_1 - (\beta_1)_0}{\sigma_{\hat{\beta}_1}}$
9.4.9	t statistic for testing hypotheses about β	$t = \frac{\hat{\beta}_1 - (\beta_1)_0}{s_{\hat{\beta}_1}}$
9.5.1	Prediction interval for Y for a given X	$\hat{y} \pm t_{(1-a/2)} s_{y x} \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$
9.5.2	Confidence interval for the mean of Y for a given X	$\hat{y} \pm t_{(1-a/2)} s_{y x} \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x_i - \bar{x})^2}}$
9.7.1–9.7.2	Correlation coefficient	$r = \sqrt{\frac{\hat{\beta}_1^2 [\sum x_i^2 - (\sum x_i)^2 / n]}{\sum y_i^2 - (\sum y_i)^2 / n}}$ $= \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$

9.7.3	t statistic for correlation coefficient	$t = r\sqrt{\frac{n-2}{1-r^2}}$
9.7.4	z statistic for correlation coefficient	$z_r = \frac{1}{2} \ln \frac{1+r}{1-r}$
9.7.5	Estimated standard deviation for z statistic	$\sigma_{z_p} = \frac{1}{\sqrt{n-3}}$
9.7.6	Z statistic for correlation coefficient	$Z = \frac{z_r - z_p}{1/\sqrt{n-3}}$
9.7.7	Z statistic for correlation coefficient when $n < 25$	$z^* = z_r - \frac{3z_r + r}{4n}$
9.7.8	Standard deviation for z^*	$\sigma_{z^*} = \frac{1}{\sqrt{n-3}}$
9.7.9	Z^* statistic for correlation coefficient	$Z^* = \frac{z^* - \xi^*}{1/\sqrt{n-1}} = (z^* - \xi^*)\sqrt{n-1}$, where $\xi^* = Z_p - \frac{(3z_p + \rho)}{4n}$
9.7.10	Confidence interval for ρ	$Z_r = Z(1/\sqrt{n-3})$
Symbol Key	<ul style="list-style-type: none"> • β_0 = regression intercept term • $\hat{\beta}_0$ = estimated regression intercept • α = probability of type I error or regression intercept • $\hat{\beta}_1$ = estimated regression slope • β_1 = regression slope • ϵ = error term • μ_x = population mean of statistic/variable x • n = sample size • σ_x^2 = population variance of statistic/variable x • ρ = population correlation coefficient • r = sample correlation coefficient • r^2 = sample coefficient of determination • t = t statistic • x_i = value of independent variable at i • \bar{x} = sample mean of independent variable • y_i = value of dependent variable at i • \bar{y} = sample mean of dependent variable • \hat{y} = estimated y • z = z statistic 	

REVIEW QUESTIONS AND EXERCISES

1. What are the assumptions underlying simple linear regression analysis when one of the objectives is to make inferences about the population from which the sample data were drawn?
2. Why is the regression equation called the least-squares equation?
3. Explain the meaning of $\hat{\beta}_0$ in the sample regression equation.
4. Explain the meaning of $\hat{\beta}_1$ in the sample regression equation.
5. Explain the following terms:
 - (a) Total sum of squares
 - (b) Explained sum of squares
 - (c) Unexplained sum of squares
6. Explain the meaning of and the method of computing the coefficient of determination.
7. What is the function of the analysis of variance in regression analysis?
8. Describe three ways in which one may test the null hypothesis that $\hat{\beta}_1 = 0$.
9. For what two purposes can a regression equation be used?
10. What are the assumptions underlying simple correlation analysis when inference is an objective?
11. What is meant by the unit of association in regression and correlation analysis?
12. What are the possible explanations for a significant sample correlation coefficient?
13. Explain why it is risky to use a sample regression equation to predict or to estimate outside the range of values of the independent variable represented in the sample.
14. Describe a situation in your particular area of interest where simple regression analysis would be useful. Use real or realistic data and do a complete regression analysis.
15. Describe a situation in your particular area of interest where simple correlation analysis would be useful. Use real or realistic data and do a complete correlation analysis.

In each of the following exercises, carry out the required analysis and test hypotheses at the indicated significance levels. Compute the p value for each test.

16. A study by Scrogin et al. (A-11) was designed to assess the effects of concurrent manipulations of dietary NaCl and calcium on blood pressure as well as blood pressure and catecholamine responses to stress. Subjects were salt-sensitive, spontaneously hypertensive male rats. Among the analyses performed by the investigators was a correlation between baseline blood pressure and plasma epinephrine concentration (E). The following data on these two variables were collected. Let $\alpha = .01$.

BP	PlasmaE	BP	PlasmaE
163.90	248.00	143.20	179.00
195.15	339.20	166.00	160.40
170.20	193.20	160.40	263.50
171.10	307.20	170.90	184.70

(Continued)

BP	PlasmaE	BP	PlasmaE
148.60	80.80	150.90	227.50
195.70	550.00	159.60	92.35
151.00	70.00	141.60	139.35
166.20	66.00	160.10	173.80
177.80	120.00	166.40	224.80
165.10	281.60	162.00	183.60
174.70	296.70	214.20	441.60
164.30	217.30	179.70	612.80
152.50	88.00	178.10	401.60
202.30	268.00	198.30	132.00
171.70	265.50		

Source: Data provided courtesy of Karie E. Scrogin.

17. Dean Parmalee (A-12) wished to know if the year-end grades assigned to Wright State University Medical School students are predictive of their second-year board scores. The following table shows, for 89 students, the year-end score (AVG, in percent of 100) and the score on the second-year medical board examination (BOARD).

AVG	BOARD	AVG	BOARD	AVG	BOARD
95.73	257	85.91	208	82.01	196
94.03	256	85.81	210	81.86	179
91.51	242	85.35	212	81.70	207
91.49	223	85.30	225	81.65	202
91.13	241	85.27	203	81.51	230
90.88	234	85.05	214	81.07	200
90.83	226	84.58	176	80.95	200
90.60	236	84.51	196	80.92	160
90.30	250	84.51	207	80.84	205
90.29	226	84.42	207	80.77	194
89.93	233	84.34	211	80.72	196
89.83	241	84.34	202	80.69	171
89.65	234	84.13	229	80.58	201
89.47	231	84.13	202	80.57	177
88.87	228	84.09	184	80.10	192
88.80	229	83.98	206	79.38	187
88.66	235	83.93	202	78.75	161
88.55	216	83.92	176	78.32	172
88.43	207	83.73	204	78.17	163
88.34	224	83.47	208	77.39	166
87.95	237	83.27	211	76.30	170
87.79	213	83.13	196	75.85	159
87.01	215	83.05	203	75.60	154
86.86	187	83.02	188	75.16	169
86.85	204	82.82	169	74.85	159
86.84	219	82.78	205	74.66	167

(Continued)

AVG	BOARD	AVG	BOARD	AVG	BOARD
86.30	228	82.57	183	74.58	154
86.13	210	82.56	181	74.16	148
86.10	216	82.45	173	70.34	159
85.92	212	82.24	185		

Source: Data provided courtesy of Dean Parmalee, M.D. and the Wright State University Statistical Consulting Center.

Perform a complete regression analysis with AVG as the independent variable. Let $\alpha = .05$ for all tests.

18. Maria Mathias (A-13) conducted a study of hyperactive children. She measured the children's attitude, hyperactivity, and social behavior before and after treatment. The following table shows for 31 subjects the age and improvement scores from pre-treatment to post-treatment for attitude (ATT), social behavior (SOC), and hyperactivity (HYP). A negative score for HYP indicates an improvement in hyperactivity; a positive score in ATT or SOC indicates improvement. Perform an analysis to determine if there is evidence to indicate that age (years) is correlated with any of the three outcome variables. Let $\alpha = .05$ for all tests.

Subject No.	AGE	ATT	HYP	SOC
1	9	-1.2	-1.2	0.0
2	9	0.0	0.0	1.0
3	13	-0.4	0.0	0.2
4	6	-0.4	-0.2	1.2
5	9	1.0	-0.8	0.2
6	8	0.8	0.2	0.4
7	8	-0.6	-0.2	0.6
8	9	-1.2	-0.8	-0.6
9	7	0.0	0.2	0.8
10	12	0.4	-0.8	0.4
11	9	-0.8	0.8	-0.2
12	10	1.0	-0.8	1.2
13	12	1.4	-1.6	0.6
14	9	1.0	-0.2	-0.2
15	12	0.8	-0.8	1.0
16	9	1.0	0.4	0.4
17	10	0.4	-0.2	0.6
18	7	0.0	-0.4	0.6
19	12	1.1	-0.6	0.8
20	9	0.2	-0.4	0.2
21	7	0.4	-0.2	0.6
22	6	0.0	-3.2	1.0
23	11	0.6	-0.4	0.0
24	11	0.4	-0.4	0.0
25	11	1.0	-0.7	-0.6
26	11	0.8	-0.8	0.0

(Continued)

Subject No.	AGE	ATT	HYP	SOC
27	11	1.2	0.6	1.0
28	11	0.2	0.0	-0.2
29	11	0.8	-1.2	0.3
30	8	0.0	0.0	-0.4
31	9	0.4	-0.2	0.2

Source: Data provided courtesy of Maria Mathias, M.D. and the Wright State University Statistical Consulting Center.

19. A study by Triller et al. (A-14) examined the length of time required for home health-care nurses to repackage a patient’s medications into various medication organizers (i.e., pill boxes). For the 19 patients in the study, researchers recorded the time required for repackaging of medications. They also recorded the number of problems encountered in the repackaging session.

Patient No.	No. of Problems	Repackaging Time (Minutes)	Patient No.	No. of Problems	Repackaging Time (Minutes)
1	9	38	11	1	10
2	2	25	12	2	15
3	0	5	13	1	17
4	6	18	14	0	18
5	5	15	15	0	23
6	3	25	16	10	29
7	3	10	17	0	5
8	1	5	18	1	22
9	2	10	19	1	20
10	0	15			

Source: Data provided courtesy of Darren M. Triller, Pharm.D.

Perform a complete regression analysis of these data using the number of problems to predict the time it took to complete a repackaging session. Let $\alpha = .05$ for all tests. What conclusions can be drawn from your analysis? How might your results be used by health-care providers?

20. The following are the pulmonary blood flow (PBF) and pulmonary blood volume (PBV) values recorded for 16 infants and children with congenital heart disease:

Y PBV (ml/sqM)	X PBF (L/min/sqM)
168	4.31
280	3.40
391	6.20
420	17.30
303	12.30
429	13.99
605	8.73
522	8.90
224	5.87
291	5.00

(Continued)

<i>Y</i> PBV (ml/sqM)	<i>X</i> PBF (L/min/sqM)
233	3.51
370	4.24
531	19.41
516	16.61
211	7.21
439	11.60

Find the regression equation describing the linear relationship between the two variables, compute r^2 , and test $H_0 : \beta_1 = 0$ by both the F test and the t test. Let $\alpha = .05$.

21. Fifteen specimens of human sera were tested comparatively for tuberculin antibody by two methods. The logarithms of the titers obtained by the two methods were as follows:

Method	
<i>A (X)</i>	<i>B (Y)</i>
3.31	4.09
2.41	3.84
2.72	3.65
2.41	3.20
2.11	2.97
2.11	3.22
3.01	3.96
2.13	2.76
2.41	3.42
2.10	3.38
2.41	3.28
2.09	2.93
3.00	3.54
2.08	3.14
2.11	2.76

Find the regression equation describing the relationship between the two variables, compute r^2 , and test $H_0 : \beta_1 = 0$ by both the F test and the t test.

22. The following table shows the methyl mercury intake and whole blood mercury values in 12 subjects exposed to methyl mercury through consumption of contaminated fish:

<i>X</i> Methyl Mercury Intake ($\mu\text{g Hg/day}$)	<i>Y</i> Mercury in whole blood (ng/g)
180	90
200	120
230	125
410	290

(Continued)

<i>X</i>	<i>Y</i>
Methyl Mercury Intake ($\mu\text{g Hg/day}$)	Mercury in Whole Blood (ng/g)
600	310
550	290
275	170
580	375
105	70
250	105
460	205
650	480

Find the regression equation describing the linear relationship between the two variables, compute r^2 , and test $H_0 : \beta_1 = 0$ by both the F and t tests.

23. The following are the weights (kg) and blood glucose levels (mg/100 ml) of 16 apparently healthy adult males:

Weight (<i>X</i>)	Glucose (<i>Y</i>)
64.0	108
75.3	109
73.0	104
82.1	102
76.2	105
95.7	121
59.4	79
93.4	107
82.1	101
78.9	85
76.7	99
82.1	100
83.9	108
73.0	104
64.4	102
77.6	87

Find the simple linear regression equation and test $H_0 : \beta_1 = 0$ using both ANOVA and the t test. Test $H_0 : \rho = 0$ and construct a 95 percent confidence interval for ρ . What is the predicted glucose level for a man who weighs 95 kg? Construct the 95 percent prediction interval for his weight. Let $\alpha = .05$ for all tests.

24. The following are the ages (years) and systolic blood pressures of 20 apparently healthy adults:

Age (<i>X</i>)	BP (<i>Y</i>)	Age (<i>X</i>)	BP (<i>Y</i>)
20	120	46	128
43	128	53	136
63	141	70	146

(Continued)

Age (X)	BP (Y)	Age (X)	BP (Y)
26	126	20	124
53	134	63	143
31	128	43	130
58	136	26	124
46	132	19	121
58	140	31	126
70	144	23	123

Find the simple linear regression equation and test $H_0 : \beta_1 = 0$ using both ANOVA and the t test. Test $H_0 : \rho = 0$ and construct a 95 percent confidence interval for ρ . Find the 95 percent prediction interval for the systolic blood pressure of a person who is 25 years old. Let $\alpha = .05$ for all tests.

25. The following data were collected during an experiment in which laboratory animals were inoculated with a pathogen. The variables are time in hours after inoculation and temperature in degrees Celsius.

Time	Temperature	Time	Temperature
24	38.8	44	41.1
28	39.5	48	41.4
32	40.3	52	41.6
36	40.7	56	41.8
40	41.0	60	41.9

Find the simple linear regression equation and test $H_0 : \beta_1 = 0$ using both ANOVA and the t test. Test $H_0 : \rho = 0$ and construct a 95 percent confidence interval for ρ . Construct the 95 percent prediction interval for the temperature at 50 hours after inoculation. Let $\alpha = .05$ for all tests.

For each of the studies described in Exercises 26 through 28, answer as many of the following questions as possible.

- Which is more relevant, regression analysis or correlation analysis, or are both techniques equally relevant?
 - Which is the independent variable?
 - Which is the dependent variable?
 - What are the appropriate null and alternative hypotheses?
 - Do you think the null hypothesis was rejected? Explain why or why not.
 - Which is the more relevant objective, prediction or estimation, or are the two equally relevant?
 - What is the sampled population?
 - What is the target population?
 - Are the variables directly or inversely related?
26. Lamarre-Cliche et al. (A-15) state, "The QT interval corrected for heart rate (QTc) is believed to reflect sympathovagal balance. It has also been established that β -blockers influence the autonomic nervous system." The researchers performed correlation analysis to measure the association between QTc interval, heart rate, heart rate change, and therapeutic blood pressure response for 73

hypertensive subjects taking β -blockers. The researchers found that QTc interval length, pretreatment heart rate, and heart rate change with therapy were not good predictors of blood pressure response to β 1-selective β -blockers in hypertensive subjects.

27. Skinner et al. (A-16) conducted a cross-sectional telephone survey to obtain 24-hour dietary recall of infants' and toddlers' food intakes, as reported by mothers or other primary caregivers. One finding of interest was that among 561 toddlers ages 15–24 months, the age in weeks of the child was negatively related to vitamin C density $\hat{\beta}_1 = -.43$, $p = .01$. When predicting calcium density, age in weeks of the child produced a slope coefficient of -1.47 with a p of $.09$.
28. Park et al. (A-17) studied 29 male subjects with clinically confirmed cirrhosis. Among other variables, they measured whole blood manganese levels (MnB), plasma manganese (MnP), urinary manganese (MnU), and pallidal index (PI), a measure of signal intensity in T1 weighted magnetic resonance imaging (MRI). They found a correlation coefficient of $.559$, $p < .01$, between MnB and PI. However, there were no significant correlations between MnP and Pi or MnU and Pi ($r = .353$, $p > .05$, $r = .252$, $p > .05$, respectively).

For the studies described in Exercises 29 through 46, do the following:

- (a) Perform a statistical analysis of the data (including hypothesis testing and confidence interval construction) that you think would yield useful information for the researchers.
- (b) Construct graphs that you think would be helpful in illustrating the relationships among variables.
- (c) Where you think appropriate, use techniques learned in other chapters, such as analysis of variance and hypothesis testing and interval estimation regarding means and proportions.
- (d) Determine p values for each computed test statistic.
- (e) State all assumptions that are necessary to validate your analysis.
- (f) Describe the population(s) about which you think inferences based on your analysis would be applicable.
- (g) If available, consult the cited reference and compare your analyses and results with those of the authors.
29. Moerloose et al. (A-18) conducted a study to evaluate the clinical usefulness of a new laboratory technique (method A) for use in the diagnosis of pulmonary embolism (PE). The performance of the new technique was compared with that of a standard technique (method B). Subjects consisted of patients with clinically suspected PE who were admitted to the emergency ward of a European university hospital. The following are the measurements obtained by the two techniques for 85 patients. The researchers performed two analyses: (1) on all 85 pairs of measurements and (2) on those pairs of measurements for which the value for method B was less than 1000.

B	A	B	A	B	A
9	119	703	599	2526	1830
84	115	725	610	2600	1880
86	108	727	3900	2770	2100
190	182	745	4050	3100	1780
208	294	752	785	3270	1870
218	226	884	914	3280	2480
251	311	920	1520	3410	1440

(Continued)

B	A	B	A	B	A
252	250	966	972	3530	2190
256	312	985	913	3900	2340
264	403	994	556	4260	3490
282	296	1050	1330	4300	4960
294	296	1110	1410	4560	7180
296	303	1170	484	4610	1390
311	336	1190	867	4810	1600
344	333	1250	1350	5070	3770
371	257	1280	1560	5470	2780
407	424	1330	1290	5576	2730
418	265	1340	1540	6230	1260
422	347	1400	1710	6260	2870
459	412	1530	1333	6370	2210
468	389	1560	1250	6430	2210
481	414	1840	764	6500	2380
529	667	1870	1680	7120	5220
540	486	2070	1310	7430	2650
562	720	2120	1360	7800	4910
574	343	2170	1770	8890	4080
646	518	2270	2240	9930	3840
664	801	2490	1910		
670	760	2520	2110		

Source: Data provided courtesy of Dr. Philippe de Moerloose.

30. Research by Huhtaniemi et al. (A-19) focused on the quality of serum luteinizing hormone (LH) during pubertal maturation in boys. Subjects, consisting of healthy boys entering puberty (ages 11 years 5 months to 12 years), were studied over a period of 18 months. The following are the concentrations (IU/L) of bioactive LH (B-LH) and immunoreactive LH (I-LH) in serum samples taken from the subjects. Only observations in which the subjects' B/I ratio was greater than 3.5 are reported here.

I-LH	B-LH	I-LH	B-LH
.104	.37	.97	3.63
.041	.28	.49	2.26
.124	.64	1	4.55
.808	2.32	1.17	5.06
.403	1.28	1.46	4.81
.27	.9	1.97	8.18
.49	2.45	.88	2.48
.66	2.8	1.24	4.8
.82	2.6	1.54	3.12
1.09	4.5	1.71	8.4
1.05	3.2	1.11	6
.83	3.65	1.35	7.2
.89	5.25	1.59	7.6
.75	2.9		

Source: Data provided courtesy of Dr. Ilpo T. Huhtaniemi.

31. Tsau et al. (A-20) studied urinary epidermal growth factor (EGF) excretion in normal children and those with acute renal failure (ARF). Random urine samples followed by 24-hour urine collection were obtained from 25 children. Subjects ranged in age from 1 month to 15 years. Urinary EGF excretion was expressed as a ratio of urinary EGF to urinary creatinine concentration (EGF/Cr). The authors conclude from their research results that it is reasonable to use random urine tests for monitoring EGF excretion. Following are the random (spot) and 24-hour urinary EGF/Cr concentrations (pmol/mmol) for the 25 subjects:

Subject	24-h Urine EGF/Cr (x)	Spot Urine EGF/Cr (y)	Subject	24-h Urine EGF/Cr (x)	Spot Urine EGF/Cr (y)
1	772	720	14	254	333
2	223	271	15 ^a	93	84
3	494	314	16	303	512
4	432	350	17	408	277
5 ^a	79	79	18	711	443
6 ^a	155	118	19	209	309
7	305	387	20	131	280
8	318	432	21	165	189
9 ^a	174	97	22	151	101
10	1318	1309	23	165	221
11	482	406	24	125	228
12	436	426	25	232	157
13	527	595			

^aSubjects with ARF.

Source: Data provided courtesy of Dr. Yong-Kwei Tsau.

32. One of the reasons for a study by Usaj and Starc (A-21) was an interest in the behavior of pH kinetics during conditions of long-term endurance and short-term endurance among healthy runners. The nine subjects participating in the study were marathon runners aged 26 ± 5 years. The authors report that they obtained a good correlation between pH kinetics and both short-term and long-term endurance. The following are the short- (V_{SE}) and long-term (V_{LE}) speeds and blood pH measurements for the participating subjects.

V_{LE}	V_{SE}	pH Range
5.4	5.6	.083
4.75	5.1	.1
4.6	4.6	.021
4.6	5	.065
4.55	4.9	.056
4.4	4.6	.01
4.4	4.9	.058
4.2	4.4	.013
4.2	4.5	.03

Source: Data provided courtesy of Anton Usaj, Ph.D.

33. Bean et al. (A-22) conducted a study to assess the performance of the isoelectric focusing/immunoblotting/laser densitometry (IEF/IB/LD) procedure to evaluate carbohydrate-deficient transferrin (CDT) derived from dry blood spots. The investigators evaluated paired serum (S) and dry

blood spot (DBS) specimens simultaneously for CDT. Assessment of CDT serves as a marker for alcohol abuse. The use of dry blood spots as a source of CDT for analysis by IEF/IB/LD results in simplified sampling, storage, and transportation of specimens. The following are the IEF/IB/LD values in densitometry units (DU) of CDT from 25 serum and dry blood spot specimens:

Specimen No.	S	DBS	Specimen No.	S	DBS
1	64	23	14	9	13
2	74	38	15	10	8
3	75	37	16	17	7
4	103	53	17	38	14
5	10	9	18	9	9
6	22	18	19	15	9
7	33	20	20	70	31
8	10	5	21	61	26
9	31	14	22	42	14
10	30	15	23	20	10
11	28	12	24	58	26
12	16	9	25	31	12
13	13	7			

Source: Data provided courtesy of Dr. Pamela Bean.

34. Kato et al. (A-23) measured the plasma concentration of adrenomedullin (AM) in patients with chronic congestive heart failure due to various cardiac diseases. AM is a hypotensive peptide, which, on the basis of other studies, the authors say, has an implied role as a circulating hormone in regulation of the cardiovascular system. Other data collected from the subjects included plasma concentrations of hormones known to affect the cardiovascular system. Following are the plasma AM (fmol/ml) and plasma renin activity (PRA) (ng/L $\cdot\cdot\cdot$ s) values for 19 heart failure patients:

Patient No.	Sex (1 = M, 2 = F)	Age (Years)	AM (fmol/ml)	PRA (ng/L $\cdot\cdot\cdot$ s)
1	1	70	12.11	.480594
2	1	44	7.306	.63894
3	1	72	6.906	1.219542
4	1	62	7.056	.450036
5	2	52	9.026	.19446
6	2	65	10.864	1.966824
7	2	64	7.324	.29169
8	1	71	9.316	1.775142
9	2	61	17.144	9.33408
10	1	68	6.954	.31947
11	1	63	7.488	1.594572
12	2	59	10.366	.963966
13	2	55	10.334	2.191842
14	2	57	13	3.97254
15	2	68	6.66	.52782
16	2	51	8.906	.350028
17	1	69	8.952	1.73625
18	1	71	8.034	.102786
19	1	46	13.41	1.13898

Source: Data provided courtesy of Dr. Johji Kato.

35. In a study reported on in *Archives of Disease in Childhood*, Golden et al. (A-24) tested the hypothesis that plasma calprotectin (PCal) (a neutrophil cytosolic protein released during neutrophil activation or death) concentration is an early and sensitive indicator of inflammation associated with bacterial infection in cystic fibrosis (CF). Subjects were children with confirmed CF and a control group of age- and sex-matched children without the disease. Among the data collected were the following plasma calprotectin ($\mu\text{g/L}$) and plasma copper (PCu) ($\mu\text{mol/L}$) measurements. Plasma copper is an index of acute phase response in cystic fibrosis. The authors reported a correlation coefficient of .48 between plasma calprotectin (\log_{10}) and plasma copper.

CF Subject No.	PCal	PCu	CF Subject No.	PCal	PCu	CF Subject No.	PCal	PCu
1	452	17.46	12	1548	15.31	22	674	18.11
2	590	14.84	13	708	17.00	23	3529	17.42
3	1958	27.42	14	8050	20.00	24	1467	17.42
4	2015	18.51	15	9942	25.00	25	1116	16.73
5	417	15.89	16	791	13.10	26	611	18.11
6	2884	17.99	17	6227	23.00	27	1083	21.56
7	1862	21.66	18	1473	16.70	28	1432	21.56
8	10471	19.03	19	8697	18.11	29	4422	22.60
9	25850	16.41	20	621	18.80	30	3198	18.91
10	5011	18.51	21	1832	17.08	31	544	14.37
11	5128	22.70						

Control Subject No.	PCal	PCu	Control Subject No.	PCal	PCu
1	674	16.73	17	368	16.73
2	368	16.73	18	674	16.73
3	321	16.39	19	815	19.82
4	1592	14.32	20	598	16.1
5	518	16.39	21	684	13.63
6	815	19.82	22	684	13.63
7	684	17.96	23	674	16.73
8	870	19.82	24	368	16.73
9	781	18.11	25	1148	24.15
10	727	18.11	26	1077	22.30
11	727	18.11	27	518	9.49
12	781	18.11	28	1657	16.10
13	674	16.73	29	815	19.82
14	1173	20.53	30	368	16.73
15	815	19.82	31	1077	22.30
16	727	18.11			

Source: Data provided courtesy of Dr. Barbara E. Golden.

36. Gelb et al. (A-25) conducted a study in which they explored the relationship between moderate to severe expiratory airflow limitation and the presence and extent of morphologic and CT scored emphysema in consecutively seen outpatients with chronic obstructive pulmonary disease. Among the data collected were the following measures of lung CT and pathology (PATH) for emphysema scoring:

CT Score	PATH	CT Score	PATH
5	15	45	50
90	70	45	40
50	20	85	75
10	25	7	0
12	25	80	85
35	10	15	5
40	35	45	40
45	30	37	35
5	5	75	45
25	50	5	5
60	60	5	20
70	60		

Source: Data provided courtesy of Dr. Arthur F. Gelb.

37. The objective of a study by Witteman et al. (A-26) was to investigate skin reactivity with purified major allergens and to assess the relation with serum levels of immunoglobulin E (IgE) antibodies and to determine which additional factors contribute to the skin test result. Subjects consisted of patients with allergic rhinitis, allergic asthma, or both, who were seen in a European medical center. As part of their study, the researchers collected, from 23 subjects, the following measurements on specific IgE (IU/ml) and skin test (ng/ml) in the presence of Lol p 5, a purified allergen from grass pollen. We wish to know the nature and strength of the relationship between the two variables. (*Note:* The authors converted the measurements to natural logarithms before investigating this relationship.)

IgE	Skin Test
24.87	.055
12.90	.041034
9.87	.050909
8.74	.046
6.88	.039032
5.90	.050909
4.85	.042142
3.53	.055
2.25	4.333333
2.14	.55
1.94	.050909
1.29	.446153
.94	.4
.91	.475

(Continued)

IgE	Skin Test
.55	4.461538
.30	4.103448
.14	7.428571
.11	4.461538
.10	6.625
.10	49.13043
.10	36.47058
.10	52.85714
.10	47.5

Source: Data provided courtesy of Dr. Jaring S. van der Zee.

38. Garland et al. (A-27) conducted a series of experiments to delineate the complex maternal-fetal pharmacokinetics and the effects of zidovudine (AZT) in the chronically instrumented maternal and fetal baboon (*Papio* species) during both steady-state intravenous infusion and oral bolus dosage regimens. Among the data collected were the following measurements on dosage (mg/kg/h) and steady-state maternal plasma AZT concentration (ng/ml):

Dosage	AZT Concentration	Dosage	AZT Concentration
2.5	832	2.0	771
2.5	672	1.8	757
2.5	904	0.9	213
2.5	554	0.6	394
2.5	996	0.9	391
1.9	878	1.3	430
2.1	815	1.1	440
1.9	805	1.4	352
1.9	592	1.1	337
0.9	391	0.8	181
1.5	710	0.7	174
1.4	591	1.0	470
1.4	660	1.1	426
1.5	694	0.8	170
1.8	668	1.0	360
1.8	601	0.9	320

Source: Data provided courtesy of Dr. Marianne Garland.

39. The purpose of a study by Halligan et al. (A-28) was to evaluate diurnal variation in blood pressure (BP) in women who were normotensive and those with pre-eclampsia. The subjects were similar in age, weight, and mean duration of gestation (35 weeks). The researchers collected the following BP readings. As part of their analysis they studied the relationship between mean day and night measurements and day/night differences for both diastolic and systolic BP in each group.

C1	C2	C3	C4	C5	C1	C2	C3	C4	C5
0	75	56	127	101	1	94	78	137	119
0	68	57	113	104	1	90	86	139	138
0	72	58	115	105	1	85	69	138	117
0	71	51	111	94	1	80	75	133	126
0	81	61	130	110	1	81	60	127	112
0	68	56	111	101	1	89	79	137	126
0	78	60	113	102	1	107	110	161	161
0	71	55	120	99	1	98	88	152	141
0	65	51	106	96	1	78	74	134	132
0	78	61	120	109	1	80	80	121	121
0	74	60	121	104	1	96	83	143	129
0	75	52	121	102	1	85	76	137	131
0	68	50	109	91	1	79	74	135	120
0	63	49	108	99	1	91	95	139	135
0	77	47	132	115	1	87	67	137	115
0	73	51	112	90	1	83	64	143	119
0	73	52	118	97	1	94	85	127	123
0	64	62	122	114	1	85	70	142	124
0	64	54	108	94	1	78	61	119	110
0	66	54	106	88	1	80	59	129	114
0	72	49	116	101	1	98	102	156	163
0	83	60	127	103	1	100	100	149	149
0	69	50	121	104	1	89	84	141	135
0	72	52	108	95	1	98	91	148	139

C1 = group (0 = normotensive, 1 = pre-eclamptic); C2 = day diastolic; C3 = night diastolic; C4 = day systolic; C5 = night systolic.

Source: Data provided courtesy of Dr. Aidan Halligan.

40. Marks et al. (A-29) conducted a study to determine the effects of rapid weight loss on contraction of the gallbladder and to evaluate the effects of ursodiol and ibuprofen on saturation, nucleation and growth, and contraction. Subjects were obese patients randomly assigned to receive ursodiol, ibuprofen, or placebo. Among the data collected were the following cholesterol saturation index values (CSI) and nucleation times (NT) in days of 13 (six male, seven female) placebo-treated subjects at the end of 6 weeks:

CSI	NT
1.20	4.00
1.42	6.00
1.18	14.00
.88	21.00
1.05	21.00
1.00	18.00
1.39	6.00
1.31	10.00
1.17	9.00

(Continued)

CSI	NT
1.36	14.00
1.06	21.00
1.30	8.00
1.71	2.00

Source: Data provided courtesy of Dr. Jay W. Marks.

41. The objective of a study by Peacock et al. (A-30) was to investigate whether spinal osteoarthritis is responsible for the fact that lumbar spine bone mineral density (BMD) is greater when measured in the anteroposterior plane than when measured in the lateral plane. Lateral spine radiographs were studied from women (age range 34 to 87 years) who attended a hospital outpatient department for bone density measurement and underwent lumbar spine radiography. Among the data collected were the following measurements on anteroposterior (A) and lateral (L) BMD (g/cm²):

ABMD	LBMD	ABMD	LBMD	ABMD	LBMD
.879	.577	1.098	.534	1.091	.836
.824	.622	.882	.570	.746	.433
.974	.643	.816	.558	1.127	.732
.909	.664	1.017	.675	1.411	.766
.872	.559	.669	.590	.751	.397
.930	.663	.857	.666	.786	.515
.912	.710	.571	.474	1.031	.574
.758	.592	1.134	.711	.622	.506
1.072	.702	.705	.492	.848	.657
.847	.655	.775	.348	.778	.537
1.000	.518	.968	.579	.784	.419
.565	.354	.963	.665	.659	.429
1.036	.839	.933	.626	.948	.485
.811	.572	.704	.194	.634	.544
.901	.612	.624	.429	.946	.550
1.052	.663	1.119	.707	1.107	.458
.731	.376	.686	.508	1.583	.975
.637	.488	.741	.484	1.026	.550
.951	.747	1.028	.787		
.822	.610	.649	.469		
.951	.710	1.166	.796		
1.026	.694	.954	.548		
1.022	.580	.666	.545		
		1.047	.706		
		.737	.526		

Source: Data provided courtesy of Dr. Cyrus Cooper.

42. Sloan et al. (A-31) note that cardiac sympathetic activation and parasympathetic withdrawal result in heart rate increases during psychological stress. As indicators of cardiac adrenergic activity, plasma epinephrine (E) and norepinephrine (NE) generally increase in response to psychological challenge. Power spectral analysis of heart period variability also provides estimates of cardiac autonomic

nervous system activity. The authors conducted a study to determine the relationship between neurohumoral and two different spectral estimates of cardiac sympathetic nervous system activity during a quiet resting baseline and in response to a psychologically challenging arithmetic task. Subjects were healthy, medication-free male and female volunteers with a mean age of 37.8 years. None had a history of cardiac, respiratory, or vascular disease. Among the data collected were the following measurements on E, NE, low-frequency (LF) and very-low-frequency (VLF) power spectral indices, and low-frequency/high frequency ratios (LH/HF). Measurements are given for three periods: baseline (B), a mental arithmetic task (MA), and change from baseline to task (DELTA).

Patient No.	E	NE	LF/HF	LF	Period	VLF
5	3.55535	6.28040	0.66706	7.71886	B	7.74600
5	0.05557	0.13960	-0.48115	-0.99826	DELTA	-2.23823
5	3.61092	6.41999	0.18591	6.72059	MA	5.50777
6	3.55535	6.24611	2.48308	7.33729	B	6.64353
6	0.10821	-0.05374	-2.03738	-0.77109	DELTA	-1.27196
6	3.66356	6.19236	0.44569	6.56620	MA	5.37157
7	3.29584	4.91998	-0.15473	7.86663	B	7.99450
7	0.59598	0.53106	0.14086	-0.81345	DELTA	-2.86401
7	3.89182	5.45104	-0.01387	7.05319	MA	5.13049
8	4.00733	5.97635	1.58951	8.18005	B	5.97126
8	0.29673	0.11947	-0.11771	-1.16584	DELTA	-0.39078
8	4.30407	6.09582	1.47180	7.01421	MA	5.58048
12	3.87120	5.35659	0.47942	6.56488	B	5.94960
12	*	*	0.19379	0.03415	DELTA	0.50134
12	*	*	0.67321	6.59903	MA	6.45094
13	3.97029	5.85507	0.13687	6.27444	B	5.58500
13	-0.20909	0.10851	1.05965	-0.49619	DELTA	-1.68911
13	3.76120	5.96358	1.19652	5.77825	MA	3.89589
14	3.63759	5.62040	0.88389	6.08877	B	6.12490
14	0.31366	0.07333	1.06100	1.37098	DELTA	-1.07633
14	3.95124	5.69373	1.94489	7.45975	MA	5.04857
18	4.44265	5.88053	0.99200	7.52268	B	7.19376
18	0.35314	0.62824	-0.10297	-0.57142	DELTA	-2.06150
18	4.79579	6.50877	0.88903	6.95126	MA	5.13226
19	*	5.03044	0.62446	6.90677	B	7.39854
19	*	0.69966	0.09578	0.94413	DELTA	-0.88309
19	2.94444	5.73010	0.72024	7.85090	MA	6.51545
20	3.91202	5.86363	1.11825	8.26341	B	6.89497
20	-0.02020	0.21401	-0.60117	-1.13100	DELTA	-1.12073
20	3.89182	6.07764	0.51708	7.13241	MA	5.77424
21	3.55535	6.21860	0.78632	8.74397	B	8.26111
21	0.31585	-0.52487	-1.92114	-2.38726	DELTA	-2.08151
21	3.87120	5.69373	-1.13483	6.35671	MA	6.17960
22	4.18965	5.76832	-0.02785	8.66907	B	7.51529
22	0.16705	-0.05459	0.93349	-0.89157	DELTA	-1.00414
22	4.35671	5.71373	0.90563	7.77751	MA	6.51115

(Continued)

Patient No.	E	NE	LF/HF	LF	Period	VLF
23	3.95124	5.52545	-0.24196	6.75330	B	6.93020
23	0.26826	0.16491	-0.00661	0.18354	DELTA	-1.18912
23	4.21951	5.69036	-0.24856	6.93684	MA	5.74108
24	3.78419	5.59842	-0.67478	6.26453	B	6.45268
24	0.32668	-0.17347	1.44970	0.52169	DELTA	0.39277
24	4.11087	5.42495	0.77493	6.78622	MA	6.84545
1	3.36730	6.13123	0.19077	6.75395	B	6.13708
1	0.54473	0.08538	0.79284	0.34637	DELTA	-0.56569
1	3.91202	6.21661	0.98361	7.10031	MA	5.57139
3	2.83321	5.92158	1.89472	7.92524	B	6.30664
3	1.15577	0.64930	-0.75686	-1.58481	DELTA	-1.95636
3	3.98898	6.57088	1.13786	6.34042	MA	4.35028
4	4.29046	5.73657	1.81816	7.02734	B	7.02882
4	0.14036	0.47000	-0.26089	-1.08028	DELTA	-1.43858
4	4.43082	6.20658	1.55727	5.94705	MA	5.59024
5	3.93183	5.62762	1.70262	6.76859	B	6.11102
5	0.80437	0.67865	-0.26531	-0.29394	DELTA	-0.94910
5	4.73620	6.30628	1.43731	6.47465	MA	5.16192
6	3.29584	5.47227	0.18852	6.49054	B	6.84279
6	-0.16034	0.27073	-0.16485	-1.12558	DELTA	-1.84288
6	3.13549	5.74300	0.02367	5.36496	MA	4.99991
8	3.25810	5.37064	-0.09631	7.23131	B	7.16371
8	0.40547	-0.13953	0.97906	-0.62894	DELTA	-2.15108
8	3.66356	5.23111	0.88274	6.60237	MA	5.01263
9	3.78419	5.94542	0.77839	5.86126	B	6.22910
9	0.64663	0.05847	-0.42774	-0.53530	DELTA	-2.18430
9	4.43082	6.00389	0.35066	5.32595	MA	4.04480
10	4.07754	5.87493	2.32137	6.71736	B	6.59769
10	0.23995	-0.00563	-0.25309	-0.00873	DELTA	-0.75357
10	4.31749	5.86930	2.06827	6.70863	MA	5.84412
11	4.33073	5.84064	2.89058	7.22570	B	5.76079
11	-3.63759	-0.01464	-1.22533	-1.33514	DELTA	-0.55240
11	0.69315	5.82600	1.66525	5.89056	MA	5.20839
12	3.55535	6.04501	1.92977	8.50684	B	7.15797
12	0.13353	0.12041	-0.15464	-0.84735	DELTA	0.13525
12	3.68888	6.16542	1.77513	7.65949	MA	7.29322
13	3.33220	4.63473	-0.11940	6.35464	B	6.76285
13	1.16761	1.05563	0.85621	0.63251	DELTA	-0.52121
13	4.49981	5.69036	0.73681	6.98716	MA	6.24164
14	3.25810	5.96358	1.10456	7.01270	B	7.49426
14	*	*	0.26353	-1.20066	DELTA	-3.15046
14	*	*	1.36809	5.81204	MA	4.34381
15	5.42935	6.34564	2.76361	9.48594	B	7.05730
15	*	*	-1.14662	-1.58468	DELTA	-0.08901
15	*	*	1.61699	7.90126	MA	6.96829
16	4.11087	6.59441	-0.23319	6.68269	B	6.76872
16	-0.06782	-0.54941	0.34755	-0.29398	DELTA	-1.80868

(Continued)

Patient No.	E	NE	LF/HF	LF	Period	VLF
16	4.04305	6.04501	0.11437	6.38871	MA	4.96004
17	*	6.28040	1.40992	6.09671	B	4.82671
17	*	-0.12766	-0.17490	-0.05945	DELTA	0.69993
17	*	6.15273	1.23501	6.03726	MA	5.52665
18	2.39790	6.03548	0.23183	6.39707	B	6.60421
18	1.06784	0.11299	0.27977	-0.38297	DELTA	-1.92672
18	3.46574	6.14847	0.51160	6.01410	MA	4.67749
19	4.21951	6.35784	1.08183	5.54214	B	5.69070
19	0.21131	-0.00347	0.12485	-0.54440	DELTA	-1.49802
19	4.43082	6.35437	1.20669	4.99774	MA	4.19268
20	4.14313	5.73334	0.89483	7.35045	B	6.93974
20	-0.11778	0.00000	0.17129	-0.58013	DELTA	-1.72916
20	4.02535	5.73334	1.06612	6.77032	MA	5.21058
21	3.66356	6.06843	-0.87315	5.09848	B	6.02972
21	0.20764	-0.10485	0.41178	-0.33378	DELTA	-2.00974
21	3.87120	5.96358	-0.46137	4.76470	MA	4.01998
22	3.29584	5.95324	2.38399	7.62877	B	7.54359
22	0.36772	0.68139	-0.75014	-0.89992	DELTA	-1.25555
22	3.66356	6.63463	1.63384	6.72884	MA	6.28804

* = missing data.

Source: Data provided courtesy of Dr. Richard P. Sloan.

43. The purpose of a study by Chati et al. (A-32) was to ascertain the role of physical deconditioning in skeletal muscle metabolic abnormalities in patients with chronic heart failure (CHF). Subjects included ambulatory CHF patients (12 males, two females) ages 35 to 74 years. Among the data collected were the following measurements, during exercise, of workload (WL) under controlled conditions, peak oxygen consumption (V_{O_2}), anaerobic ventilatory threshold (AT), both measured in ml/kg/min, and exercise total time (ET) in seconds.

WL	V_{O_2}	AT	ET	WL	V_{O_2}	AT	ET
7.557	32.800	13.280	933.000	3.930	22.500	18.500	720.000
3.973	8.170	6.770	255.000	3.195	17.020	8.520	375.000
5.311	16.530	11.200	480.000	2.418	15.040	12.250	480.000
5.355	15.500	10.000	420.000	0.864	7.800	4.200	240.000
6.909	24.470	11.550	960.000	2.703	12.170	8.900	513.000
1.382	7.390	5.240	346.000	1.727	15.110	6.300	540.000
8.636	19.000	10.400	600.000	7.773	21.100	12.500	1200.000

Source: Data provided courtesy of Dr. Zukaï Chati.

44. Czader et al. (A-33) investigated certain prognostic factors in patients with centroblastic-centrocytic non-Hodgkin's lymphomas (CB/CC NHL). Subjects consisted of men and women between the ages of 20 and 84 years at time of diagnosis. Among the data collected were the following measurements on two relevant factors, A and B. The authors reported a significant correlation between the two.

A	B	A	B	A	B
20.00	.154	22.34	.147	48.66	.569
36.00	.221	18.00	.132	20.00	.227
6.97	.129	18.00	.085	17.66	.125
13.67	.064	22.66	.577	14.34	.089
36.34	.402	45.34	.134	16.33	.051
39.66	.256	20.33	.246	18.34	.100
14.66	.188	16.00	.175	26.49	.202
27.00	.138	15.66	.105	13.33	.077
2.66	.078	23.00	.145	6.00	.206
22.00	.142	27.33	.129	15.67	.153
11.00	.086	6.27	.062	32.33	.549
20.00	.170			24.34	.147
22.66	.198			22.33	.769
7.34	.092			11.33	.130
29.67	.227			6.67	.099
11.66	.159				
8.05	.223				
22.66	.065				

Source: Data provided courtesy of Dr. Magdalena Czader and Dr. Anna Porwit-MacDonald.

45. Fleroxacin, a fluoroquinolone derivative with a broad antibacterial spectrum and potent activity in vitro against gram-negative and many gram-positive bacteria, was the subject of a study by Reigner and Welker (A-34). The objectives of their study were to estimate the typical values of clearance over systemic availability (CL/F) and the volume of distribution over systemic availability (V/F) after the administration of therapeutic doses of fleroxacin and to identify factors that influence the disposition of fleroxacin and to quantify the degree to which they do so. Subjects were 172 healthy male and female volunteers and uninfected patients representing a wide age range. Among the data analyzed were the following measurements (ml/min) of CL/F and creatinine clearance (CL_{cr}). According to the authors, previous studies have shown that there is a correlation between the two variables.

CL/F	CL _{cr}	CL/F	CL _{cr}	CL/F	CL _{cr}	CL/F	CL _{cr}
137.000	96.000	77.000	67.700	152.000	109.000	132.000	111.000
106.000	83.000	57.000	51.500	100.000	82.000	94.000	118.000
165.000	100.000	69.000	52.400	86.000	88.000	90.000	111.000
127.000	101.000	69.000	65.900	69.000	67.000	87.000	124.000
139.000	116.000	76.000	60.900	108.000	68.700	48.000	10.600
102.000	78.000	77.000	93.800	77.000	83.200	26.000	9.280
72.000	84.000	66.000	73.800	85.000	72.800	54.000	12.500
86.000	81.000	53.000	99.100	89.000	82.300	36.000	9.860
85.000	77.000	26.000	110.000	105.000	71.100	26.000	4.740
122.000	102.000	89.000	99.900	66.000	56.000	39.000	7.020
76.000	80.000	44.000	73.800	73.000	61.000	27.000	6.570
57.000	67.000	27.000	65.800	64.000	79.500	36.000	13.600
62.000	41.000	96.000	109.000	26.000	9.120	15.000	7.600
90.000	93.000	102.000	76.800	29.000	8.540	138.000	100.000

(Continued)

CL/F	CLer	CL/F	CLer	CL/F	CLer	CL/F	CLer
165.000	88.000	159.000	125.000	39.100	93.700	127.000	108.000
132.000	64.000	115.000	112.000	75.500	65.600	203.000	121.000
159.000	92.000	82.000	91.600	86.000	102.000	198.000	143.000
148.000	114.000	96.000	83.100	106.000	105.000	151.000	126.000
116.000	59.000	121.000	88.800	77.500	67.300	113.000	111.000
124.000	67.000	99.000	94.000	87.800	96.200	139.000	109.000
76.000	56.000	120.000	91.500	25.700	6.830	135.000	102.000
40.000	61.000	101.000	83.800	89.700	74.800	116.000	110.000
23.000	35.000	118.000	97.800	108.000	84.000	148.000	94.000
27.000	38.000	116.000	100.000	58.600	79.000	221.000	110.000
64.000	79.000	116.000	67.500	91.700	68.500	115.000	101.000
44.000	64.000	87.000	97.500	48.900	20.600	150.000	110.000
59.000	94.000	59.000	45.000	53.500	10.300	135.000	143.000
47.000	96.000	96.000	53.500	41.400	11.800	201.000	115.000
17.000	25.000	163.000	84.800	24.400	7.940	164.000	103.000
67.000	122.000	39.000	73.700	42.300	3.960	130.000	103.000
25.000	43.000	73.000	87.300	34.100	12.700	162.000	169.000
24.000	22.000	45.000	74.800	28.300	7.170	107.000	140.000
65.000	55.000	94.000	100.000	47.000	6.180	78.000	87.100
69.000	42.500	74.000	73.700	30.500	9.470	87.500	134.000
55.000	71.000	70.000	64.800	38.700	13.700	108.000	108.000
39.000	34.800	129.000	119.000	60.900	17.000	126.000	118.000
58.000	50.300	34.000	30.000	51.300	6.810	131.000	109.000
37.000	38.000	42.000	65.900	46.100	24.800	94.400	60.000
32.000	32.000	48.000	34.900	25.000	7.200	87.700	82.900
66.000	53.500	58.000	55.900	29.000	7.900	94.000	99.600
49.000	60.700	30.000	40.100	25.000	6.600	157.000	123.000
40.000	66.500	47.000	48.200	40.000	8.600		
34.000	22.600	35.000	14.800	28.000	5.500		
87.000	61.800	20.000	14.400				

Source: Data provided courtesy of Dr. Bruno Reigner.

46. Yasu et al. (A-35) used noninvasive magnetic resonance spectroscopy to determine the short- and long-term effects of percutaneous transvenous mitral commissurotomy (PTMC) on exercise capacity and metabolic responses of skeletal muscles during exercise. Data were collected on 11 patients (2 males, 9 females) with symptomatic mitral stenosis. Their mean age was 52 years with a standard deviation of 11. Among the data collected were the following measurements on changes in mitral valve area (d-MVA) and peak oxygen consumption (d-Vo₂) 3, 30, and 90 days post-PTMC:

Subject	Days		d-Vo ₂ (ml/kg/min)
	Post-PTMC	d-MVA (cm ²)	
1	3	0.64	0.3
2	3	0.76	-0.9
3	3	0.3	1.9
4	3	0.6	-3.1

Subject	Days		d- Vo_2 (ml/kg/min)
	Post-PTMC	d-MVA (cm^2)	
5	3	0.3	-0.5
6	3	0.4	-2.7
7	3	0.7	1.5
8	3	0.9	1.1
9	3	0.6	-7.4
10	3	0.4	-0.4
11	3	0.65	3.8
<hr/>			
1	30	0.53	1.6
2	30	0.6	3.3
3	30	0.4	2.6
4	30	0.5	*
5	30	0.3	3.6
6	30	0.3	0.2
7	30	0.67	4.2
8	30	0.75	3
9	30	0.7	2
10	30	0.4	0.8
11	30	0.55	4.2
<hr/>			
1	90	0.6	1.9
2	90	0.6	5.9
3	90	0.4	3.3
4	90	0.6	5
5	90	0.25	0.6
6	90	0.3	2.5
7	90	0.7	4.6
8	90	0.8	4
9	90	0.7	1
10	90	0.38	1.1
11	90	0.53	*

* = Missing data.

Source: Data provided courtesy of Dr. Takanori Yasu.

Exercises for Use with Large Data Sets Available on the Following Website:
www.wiley.com/college/daniel

1. Refer to the data for 1050 subjects with cerebral edema (CEREBRAL). Cerebral edema with consequent increased intracranial pressure frequently accompanies lesions resulting from head injury and other conditions that adversely affect the integrity of the brain. Available treatments for cerebral edema vary in effectiveness and undesirable side effects. One such treatment is glycerol, administered either orally or intravenously. Of interest to clinicians is the relationship between intracranial pressure and glycerol plasma concentration. Suppose you are a statistical consultant with a research team investigating the relationship between these two variables. Select a simple random sample from the population and perform the analysis that you think would be useful to the researchers. Present your findings and conclusions in narrative form and illustrate with graphs where appropriate. Compare your results with those of your classmates.

2. Refer to the data for 1050 subjects with essential hypertension (HYPERTEN). Suppose you are a statistical consultant to a medical research team interested in essential hypertension. Select a simple random sample from the population and perform the analyses that you think would be useful to the researchers. Present your findings and conclusions in narrative form and illustrate with graphs where appropriate. Compare your results with those of your classmates. Consult with your instructor regarding the size of sample you should select.
3. Refer to the data for 1200 patients with rheumatoid arthritis (CALCIUM). One hundred patients received the medicine at each dose level. Suppose you are a medical researchers wishing to gain insight into the nature of the relationship between dose level of prednisolone and total body calcium. Select a simple random sample of three patients from each dose level group and do the following.
 - (a) Use the total number of pairs of observations to obtain the least-squares equation describing the relationship between dose level (the independent variable) and total body calcium.
 - (b) Draw a scatter diagram of the data and plot the equation.
 - (c) Compute r and test for significance at the .05 level. Find the p value.
 - (d) Compare your results with those of your classmates.

REFERENCES

Methodology References

1. FREDERICK HARTWIG with BRAIN E. DEARING, *Exploratory Data Analysis*, Sage Publications, Beverly Hills, 1979.
2. IAIN M. JOHNSTONE and PAUL F. VELLEMAN, "The Resistant Line and Related Regression Methods," *Journal of the American Statistical Association*, 80 (1985), 1041–1054.
3. DONALD R. MCNEIL, *Interactive Data Analysis: A Practical Primer*, Wiley, New York, 1977.
4. PAUL F. VELLEMAN and DAVID C. HOAGLIN, *Applications, Basics, and Computing of Exploratory Data Analysis*, Duxbury, Belmont, CA, 1981.
5. R. A. FISHER, "On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample," *Metron*, 1 (1921), 3–21.
6. H. HOTELLING, "New Light on the Correlation Coefficient and Its Transforms," *Journal of the Royal Statistical Society, Series B*, 15 (1953), 193–232.
7. TERRY E. DIELMAN, *Applied Regression Analysis for Business and Economics*, Second Edition, Duxbury, Belmont, CA, 1996.
8. RONALD R. HOCKING, *Methods and Applications of Linear Models: Regression and the Analysis of Variance*, Wiley, New York, 1996.
9. WILLIAM MENDENHALL and TERRY SINCICH, *A Second Course in Statistics: Regression Analysis*, Fifth Edition, Prentice Hall, Upper Saddle River, NJ, 1996.
10. JOHN NETER, MICHAEL H. KUTNER, CHRISTOPHER J. NACHTSHEIM, and WILLIAM WASSERMAN, *Applied Linear Regression Models*, Third Edition, Irwin, Chicago, 1996.

Applications References

- A-1. JEAN-PIERRE DESPRÉS, DENIS PRUD'HOMME, MARIE-CHRISTINE POULIOT, ANGELO TREMBLAY, and CLAUDE BOUCHARD, "Estimation of Deep Abdominal Adipose-Tissue Accumulation from Simple Anthropometric Measurements in Men," *American Journal of Clinical Nutrition*, 54 (1991), 471–477.
- A-2. MORI J. KRANTZ, ILANA B. KUTINSKY, ALASTAIR D. ROBERSTON, and PHILIP S. MEHLER, "Dose-Related Effects of Methadone on QT Prolongation in a Series of Patients with Torsade de Pointes," *Pharmacotherapy*, 23 (2003), 802–805.
- A-3. ROBERT A. REISS, CURTIS E. HAAS, DEBORAH L. GRIFFIS, BERNADETTE PORTER, and MARY ANN TARA, "Point-of-Care Versus Laboratory Monitoring of Patients Receiving Different Anticoagulant Therapies," *Pharmacotherapy*, 22 (2002), 677–685.

- A-4. ROBERT B. PARKER, RYAN YATES, JUDITH E. SOBERMAN, and CASEY LAIZURE, "Effects of Grapefruit Juice on Intestinal P-glycoprotein: Evaluation Using Digoxin in Humans," *Pharmacotherapy*, 23 (2003), 979–987.
- A-5. R. B. EVANS, W. GORDON, and M. CONZEMIUS, "Effect of Velocity on Ground Reaction Forces in Dogs with Lameness Attributable to Tearing of the Cranial Cruciate Ligament," *American Journal of Veterinary Research*, 64 (2003), 1479–1481.
- A-6. DAVID KRIESER, ANDREW ROSENBERG, and GAD KAINER, "The Relationship Between Serum Creatinine, Serum Cystatin C and Glomerular Filtration Rate in Pediatric Renal Transplant Recipients: A Pilot Study," *Pediatric Transplantation*, 6 (2002), 392–395.
- A-7. OLGA KWAST-RABBEN, ROLF LIBELIUS, and HANNU HEIKKILÄ, "Somatosensory Evoked Potentials Following Stimulation of Digital Nerves," *Muscle and Nerve*, 26 (2002), 533–538.
- A-8. GERI R. BROWN and KIM PERSLEY, "Hepatitis A Epidemic in the Elderly," *Southern Medical Journal*, 95 (2002), 826–833.
- A-9. ANN E. TUZSON, KEVIN P. GRANATA, and MARK F. ABEL, "Spastic Velocity Threshold Constrains Functional Performance in Cerebral Palsy," *Archives of Physical Medicine and Rehabilitation*, 84 (2003), 1363–1368.
- A-10. M. BUTZ, K. H. WOLLINSKY, U. WIDEMUTH-CATRINESCU, A. SPERFELD, S. WINTER, H. H. MEHRKENS, A. C. LUDOLPH, and H. SCHREIBER, "Longitudinal Effects of Noninvasive Positive-Pressure Ventilation in Patients with Amyotrophic Lateral Sclerosis," *American Journal of Medical Rehabilitation*, 82 (2003), 597–604.
- A-11. KARIE E. SCROGIN, DANIEL C. HATTON, and DAVID A. MCCARRON, "The Interactive Effects of Dietary Sodium Chloride and Calcium on Cardiovascular Stress Responses," *American Journal of Physiology (Regulatory Integrative Comp. Physiol. 30)* 261 (1991), R945–R949.
- A-12. DEAN PARMALEE, Data analyzed by the Wright State University Statistical Consulting Center, Wright State University, Dayton, OH (2003).
- A-13. MARIA MATHIAS, Data analyzed by the Wright State University Statistical Consulting Center, Wright State University, Dayton, OH (2001).
- A-14. DARREN M. TRILLER, STEVEN L. CLAUSE, and CHRISTOPHER DOMAREW, "Analysis of Medication Management Activities for Home-Dwelling Patients," *American Journal of Health-System Pharmacy*, 59 (2002), 2356–2359.
- A-15. MAXIME LAMARRE-CLICHE, YVES LACOURCIÈRE, JACQUES DE CHAMPLAIN, LUC POIRIER, and PIERRE LAROCHELLE, "Does QTc Interval Predict the Response to Beta-Blockers and Calcium Channel Blockers in Hypertensives?" *Heart and Disease*, 5 (2003), 244–252.
- A-16. JEAN D. SKINNER, PAULA ZIEGLER, and MICHAEL PONZA, "Transitions in Infants' and Toddlers' Beverage Patterns," *Journal of the American Dietetic Association, Supplement*, 104 (2004), 45–50.
- A-17. NÈUNG HWA PARK, JI KANG PARK, YOUNGHEE CHOI, CHEOL-IN YOO, CHOONG RYEOL LEE, HUN LEE, HYU KYUNG KIM, SUNG-RYUL KIM, TAE-HEUM JEONG, JUNGSUN PARK, CHUNG SIK YOON, and YANGHO KIM, "Whole Blood Manganese Correlates with High Signal Intensities on T-Weighted MRI in Patients with Liver Cirrhosis," *NeuroToxicology*, 24 (2003), 909–915.
- A-18. PHILIPPE DE MOERLOOSE, SYLVIE DESMARAIS, HENRI BOUNAMEAUX, GUIDO REBER, ARNAUD PERRIER, GEORGES DUPUY, and JEAN-LOUIS PITTET, "Contribution of a New, Rapid, Individual and Quantitative Automated D-Dimer ELISA to Exclude Pulmonary Embolism," *Thrombosis and Haemostasis*, 75 (1996), 11–13.
- A-19. ILPO T. HUHTANIEMI, ANNE-MAARIT HAAVISTO, RAJA ANTTILA, MARTTI A. SIIMES, and LEO DUNKEL, "Sensitive Immunoassay and *In Vitro* Bioassay Demonstrate Constant Bioactive/Immunoreactive Ratio of Luteinizing Hormone in Healthy Boys During the Pubertal Maturation," *Pediatric Research*, 39 (1996), 180–184.
- A-20. YONG-KWEI TSAU, JI-NAN SHEU, CHIUNG-HUI CHEN, RU-JENG TENG, and HUI-CHI CHEN, "Decreased Urinary Epidermal Growth Factor in Children with Acute Renal Failure: Epidermal Growth Factor/Creatinine Ratio Not a Reliable Parameter for Urinary Epidermal Growth Factor Excretion," *Pediatric Research*, 39 (1996), 20–24.
- A-21. A. USAJ and V. STARC, "Blood pH and Lactate Kinetics in the Assessment of Running Endurance," *International Journal of Sports Medicine*, 17 (1996), 34–40.
- A-22. PAMELA BEAN, MARY SUSAN SUTPHIN, PATRICIA NECESSARY, MELKON S. AGOPIAN, KARSTEN LIEGMANN, CARL LUDVIGSEN, and JAMES B. PETER, "Carbohydrate-Deficient Transferrin Evaluation in Dry Blood Spots," *Alcoholism: Clinical and Experimental Research*, 20 (1996), 56–60.
- A-23. JOHJI KATO, KOHJI KOBAYASHI, TAKUMA ETOH, MIHO TANAKA, KAZUO KITAMURA, TAKUROH IMAMURA, YASUSHI KOIWAYA, KENJI KANGAWA, and TANENAO ETO, "Plasma Adrenomedullin Concentration in Patients with Heart Failure," *Journal of Clinical Endocrinology and Metabolism*, 81 (1996), 180–183.
- A-24. B. E. GOLDEN, P. A. CLOHESSY, G. RUSSELL, and M. K. FAGERHOL, "Calprotectin as a Marker of Inflammation in Cystic Fibrosis," *Archives of Disease in Childhood*, 74 (1996), 136–139.

- A-25. ARTHUR F. GELB, JAMES C. HOGG, NESTOR L. MÜLLER, MARK J. SCHEIN, JOSEPH KUEI, DONALD P. TASHKIN, JOEL D. EPSTEIN, JOZEF KOLLIN, ROBERT H. GREEN, NOE ZAMEL, W. MARK ELLIOTT, and LIDA HADJIAGHAI, "Contribution of Emphysema and Small Airways in COPD," *Chest*, 109 (1996), 353–359.
- A-26. AGNES M. WITTEMAN, STEVEN O. STAPEL, GERRARD J. PERDOK, DEMAN H. S. SJAMSOEDIN, HENK M. JANSEN, ROB C. AALBERSE, and JARING S. VAN DER ZEE, "The Relationship Between RAST and Skin Test Results in Patients with Asthma or Rhinitis: A Quantitative Study with Purified Major Allergens," *Journal of Allergy and Clinical Immunology*, 97 (1996), 16–25.
- A-27. MARIANNE GARLAND, HAZEL H. SZETO, SALHA S. DANIEL, PAMELA J. TROPPER, MICHAEL M. MYERS, and RAYMOND I. STARK, "Zidovudine Kinetics in the Pregnant Baboon," *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 11 (1996), 117–127.
- A-28. A. HALLIGAN, A. SHENNAN, P. C. LAMBERT, M. DE SWIET, and D. J. TAYLOR, "Diurnal Blood Pressure Difference in the Assessment of Preeclampsia," *Obstetrics & Gynecology*, 87 (1996), 205–208.
- A-29. JAY W. MARKS, GEORGE G. BONORRIS, and LESLIE J. SCHOENFIELD, "Effects of Ursodiol or Ibuprofen on Contraction of Gallbladder and Bile Among Obese Patients During Weight Loss," *Digestive Diseases and Sciences*, 41 (1996), 242–249.
- A-30. D. J. PEACOCK, P. EGGER, P. TAYLOR, M. I. D. CAWLEY, and C. COOPER, "Lateral Bone Density Measurements in Osteoarthritis of the Lumbar Spine," *Annals of the Rheumatic Diseases*, 55 (1996), 196–198.
- A-31. R. P. SLOAN, P. A. SHAPIRE, E. BAGIELLA, J. T. BIGGER, E. S. LO, and J. M. GORMAN, "Relationships Between Circulating Catecholamines and Low Frequency Heart Period Variability as Indices of Cardiac Sympathetic Activity During Mental Stress," *Psychosomatic Medicine*, 58 (1996) 25–31.
- A-32. ZUKĀI CHATI, FAÏEZ ZANNAD, CLAUDE JEANDEL, BRIGITTE LHERBIER, JEAN-MARIE ESCANYE, JACQUES ROBERT, and ETIENNE ALIOT, "Physical Deconditioning May Be a Mechanism for the Skeletal Muscle Energy Phosphate Metabolism Abnormalities in Chronic Heart Failure," *American Heart Journal*, 131 (1996), 560–566.
- A-33. MAGDALENA CZADER, JOANNA MAZUR, MIKAEL PETERSSON, JAN LILJEMARK, MATS STRÖMBERG, BIRGER CHRISTENSSON, BERNARD TRIBUKAIT, GERT AUER, ÅKE ÖST, and ANNA PORWIT, "Prognostic Significance of Proliferative and Apoptotic Fractions in Low Grade Follicle Center Cell-Derived Non-Hodgkin's Lymphomas," *Cancer*, 77 (1996), 1180–1188.
- A-34. B. G. REIGNER and H. A. WELKER, "Factors Influencing Elimination and Distribution of Fleroxacin: Metaanalysis of Individual Data from 10 Pharmacokinetic Studies," *Antimicrobial Agents and Chemotherapy*, 40 (1996), 575–580.
- A-35. TAKANORI YASU, TAKA'AKI KATSUKI, NOBUHIRO OHMURA, IKUKO NAKADA, MAFUMI OWA, MIKIHISA FUJII, AKIRA SAKAGUCHI, and MUNEYASU SAITO, "Delayed Improvement in Skeletal Muscle Metabolism and Exercise Capacity in Patients with Mitral Stenosis Following Immediate Hemodynamic Amelioration by Percutaneous Transvenous Mitral Commissurotomy," *American Journal of Cardiology*, 77 (1996), 492–497.

MULTIPLE REGRESSION AND CORRELATION

CHAPTER OVERVIEW

This chapter provides extensions of the simple linear regression and bivariate correlation models discussed in Chapter 9. The concepts and techniques discussed here are useful when the researcher wishes to consider simultaneously the relationships among more than two variables. Although the concepts, computations, and interpretations associated with analysis of multiple-variable data may seem complex, they are natural extensions of material explored in previous chapters.

TOPICS

- 10.1** INTRODUCTION
- 10.2** THE MULTIPLE LINEAR REGRESSION MODEL
- 10.3** OBTAINING THE MULTIPLE REGRESSION EQUATION
- 10.4** EVALUATING THE MULTIPLE REGRESSION EQUATION
- 10.5** USING THE MULTIPLE REGRESSION EQUATION
- 10.6** THE MULTIPLE CORRELATION MODEL
- 10.7** SUMMARY

LEARNING OUTCOMES

After studying this chapter, the student will

1. understand how to include more than one independent variable in a regression equation.
2. be able to obtain a multiple regression model and use it to make predictions.
3. be able to evaluate the multiple regression coefficients and the suitability of the regression model.
4. understand how to calculate and interpret multiple, bivariate, and partial correlation coefficients.

10.1 INTRODUCTION

In Chapter 9 we explored the concepts and techniques for analyzing and making use of the linear relationship between two variables. We saw that this analysis may lead to a linear equation that can be used to predict the value of some dependent variable given the value of an associated independent variable.

Intuition tells us that, in general, we ought to be able to improve our predicting ability by including more independent variables in such an equation. For example, a researcher may find that intelligence scores of individuals may be predicted from physical factors such as birth order, birth weight, and length of gestation along with certain hereditary and external environmental factors. Length of stay in a chronic disease hospital may be related to the patient's age, marital status, sex, and income, not to mention the obvious factor of diagnosis. The response of an experimental animal to some drug may depend on the size of the dose and the age and weight of the animal. A nursing supervisor may be interested in the strength of the relationship between a nurse's performance on the job, score on the state board examination, scholastic record, and score on some achievement or aptitude test. Or a hospital administrator studying admissions from various communities served by the hospital may be interested in determining what factors seem to be responsible for differences in admission rates.

The concepts and techniques for analyzing the associations among several variables are natural extensions of those explored in the previous chapters. The computations, as one would expect, are more complex and tedious. However, as is pointed out in Chapter 9, this presents no real problem when a computer is available. It is not unusual to find researchers investigating the relationships among a dozen or more variables. For those who have access to a computer, the decision as to how many variables to include in an analysis is based not on the complexity and length of the computations but on such considerations as their meaningfulness, the cost of their inclusion, and the importance of their contribution.

In this chapter we follow closely the sequence of the previous chapter. The regression model is considered first, followed by a discussion of the correlation model. In considering the regression model, the following points are covered: a description of the model, methods for obtaining the regression equation, evaluation of the equation, and the uses that may be made of the equation. In both models the possible inferential procedures and their underlying assumptions are discussed.

10.2 THE MULTIPLE LINEAR REGRESSION MODEL

In the multiple regression model we assume that a linear relationship exists between some variable Y , which we call the dependent variable, and k independent variables, X_1, X_2, \dots, X_k . The independent variables are sometimes referred to as *explanatory variables*, because of their use in explaining the variation in Y . They are also called *predictor variables*, because of their use in predicting Y .

Assumptions The assumptions underlying multiple regression analysis are as follows.

1. The X_i are nonrandom (fixed) variables. This assumption distinguishes the multiple regression model from the multiple correlation model, which will be presented in Section 10.6. This condition indicates that any inferences that are drawn from sample data apply only to the set of X values observed and not to some larger collection of X 's. Under the regression model, correlation analysis is not meaningful. Under the correlation model to be presented later, the regression techniques that follow may be applied.
2. For each set of X_i values there is a subpopulation of Y values. To construct certain confidence intervals and test hypotheses, it must be known, or the researcher must be willing to assume, that these subpopulations of Y values are normally distributed. Since we will want to demonstrate these inferential procedures, the assumption of normality will be made in the examples and exercises in this chapter.
3. The variances of the subpopulations of Y are all equal.
4. The Y values are independent. That is, the values of Y selected for one set of X values do not depend on the values of Y selected at another set of X values.

The Model Equation The assumptions for multiple regression analysis may be stated in more compact fashion as

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj} + \epsilon_j \quad (10.2.1)$$

where y_j is a typical value from one of the subpopulations of Y values; the β_i are called the regression coefficients; $x_{1j}, x_{2j}, \dots, x_{kj}$ are, respectively, particular values of the independent variables X_1, X_2, \dots, X_k ; and ϵ_j is a random variable with mean 0 and variance σ^2 , the common variance of the subpopulations of Y values. To construct confidence intervals for and test hypotheses about the regression coefficients, we assume that the ϵ_j are normally and independently distributed. The statements regarding ϵ_j are a consequence of the assumptions regarding the distributions of Y values. We will refer to Equation 10.2.1 as the *multiple linear regression model*.

When Equation 10.2.1 consists of one dependent variable and two independent variables, that is, when the model is written

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \epsilon_j \quad (10.2.2)$$

a *plane* in three-dimensional space may be fitted to the data points as illustrated in Figure 10.2.1. When the model contains more than two independent variables, it is described geometrically as a *hyperplane*.

In Figure 10.2.1 the observer should visualize some of the points as being located above the plane and some as being located below the plane. The deviation of a point from the plane is represented by

$$\epsilon_j = y_j - \beta_0 - \beta_1 x_{1j} - \beta_2 x_{2j} \quad (10.2.3)$$

In Equation 10.2.2, β_0 represents the point where the plane cuts the Y -axis; that is, it represents the Y -intercept of the plane. β_1 measures the average change in Y for a unit

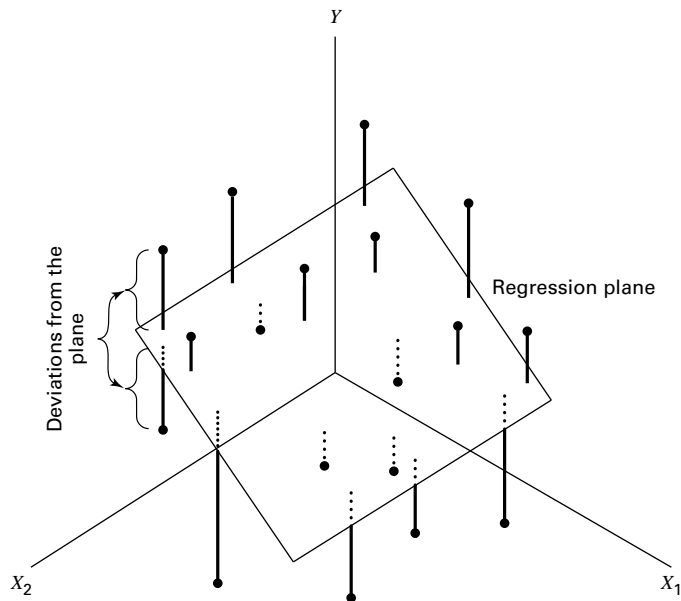


FIGURE 10.2.1 Multiple regression plane and scatter of points.

change in X_1 when X_2 remains unchanged, and β_2 measures the average change in Y for a unit change in X_2 when X_1 remains unchanged. For this reason β_1 and β_2 are referred to as *partial regression coefficients*.

10.3 OBTAINING THE MULTIPLE REGRESSION EQUATION

Unbiased estimates of the parameters $\beta_0, \beta_1, \dots, \beta_k$ of the model specified in Equation 10.2.1 are obtained by the method of least squares. This means that the sum of the squared deviations of the observed values of Y from the resulting regression surface is minimized. In the three-variable case, as illustrated in Figure 10.2.1, the sum of the squared deviations of the observations from the plane are a minimum when β_0, β_1 , and β_2 are estimated by the method of least squares. In other words, by the method of least squares, sample estimates of $\beta_0, \beta_1, \dots, \beta_k$ are selected in such a way that the quantity

$$\sum \epsilon_j^2 = \sum (y_j - \beta_0 - \beta_1 x_{1j} - \beta_2 x_{2j} - \dots - \beta_k x_{kj})^2$$

is minimized. This quantity, referred to as the sum of squares of the residuals, may also be written as

$$\sum \epsilon_j^2 = \sum (y_j - \hat{y}_j)^2 \quad (10.3.1)$$

indicating the fact that the sum of squares of deviations of the observed values of Y from the values of Y calculated from the estimated equation is minimized.

Estimates of the multiple regression parameters may be obtained by means of arithmetic calculations performed on a handheld calculator. This method of obtaining the estimates is tedious, time-consuming, subject to errors, and a waste of time when a computer is available. Those interested in examining or using the arithmetic approach may consult earlier editions of this text or those by Snedecor and Cochran (1) and Steel and Torrie (2), who give numerical examples for four variables, and Anderson and Bancroft (3), who illustrate the calculations involved when there are five variables. In the following example we use SPSS software to illustrate an interesting graphical summary of sample data collected on three variables. We then use MINITAB and SAS to illustrate the application of multiple regression analysis.

EXAMPLE 10.3.1

Researchers Jansen and Keller (A-1) used age and education level to predict the capacity to direct attention (CDA) in elderly subjects. CDA refers to neural inhibitory mechanisms that focus the mind on what is meaningful while blocking out distractions. The study collected information on 71 community-dwelling older women with normal mental status. The CDA measurement was calculated from results on standard visual and auditory measures requiring the inhibition of competing and distracting stimuli. In this study, CDA scores ranged from -7.65 to 9.61 with higher scores corresponding with better attentional functioning. The measurements on CDA, age in years, and education level (years of schooling) for 71 subjects are shown in Table 10.3.1. We wish to obtain the sample multiple regression equation.

TABLE 10.3.1 CDA Scores, Age, and Education Level for 71 Subjects Described in Example 10.3.1

Age	Ed-Level	CDA	Age	Ed-Level	CDA
72	20	4.57	79	12	3.17
68	12	-3.04	87	12	-1.19
65	13	1.39	71	14	0.99
85	14	-3.55	81	16	-2.94
84	13	-2.56	66	16	-2.21
90	15	-4.66	81	16	-0.75
79	12	-2.70	80	13	5.07
74	10	0.30	82	12	-5.86
69	12	-4.46	65	13	5.00
87	15	-6.29	73	16	0.63
84	12	-4.43	85	16	2.62
79	12	0.18	83	17	1.77
71	12	-1.37	83	8	-3.79
76	14	3.26	76	20	1.44
73	14	-1.12	77	12	-5.77
86	12	-0.77	83	12	-5.77
69	17	3.73	79	14	-4.62
66	11	-5.92	69	12	-2.03

(Continued)

Age	Ed-Level	CDA	Age	Ed-Level	CDA
65	16	5.74	66	14	-2.22
71	14	2.83	75	12	0.80
80	18	-2.40	77	16	-0.75
81	11	-0.29	78	12	-4.60
66	14	4.44	83	20	2.68
76	17	3.35	85	10	-3.69
70	12	-3.13	76	18	4.85
76	12	-2.14	75	14	-0.08
67	12	9.61	70	16	0.63
72	20	7.57	79	16	5.92
68	18	2.21	75	18	3.63
102	12	-2.30	94	8	-7.07
67	12	1.73	76	18	6.39
66	14	6.03	84	18	-0.08
75	18	-0.02	79	17	1.07
91	13	-7.65	78	16	5.31
74	15	4.17	79	12	0.30
90	15	-0.68			

Source: Data provided courtesy of Debra A. Jansen, Ph.D., R.N.

Prior to analyzing the data using multiple regression techniques, it is useful to construct plots of the relationships among the variables. This is accomplished by making separate plots of each pair of variables, (X_1, X_2) , (X_1, Y) , and (X_2, Y) . A software package such as SPSS displays each combination simultaneously in a matrix format as shown in Figure 10.3.1. From this figure it is apparent that we should expect a negative relationship

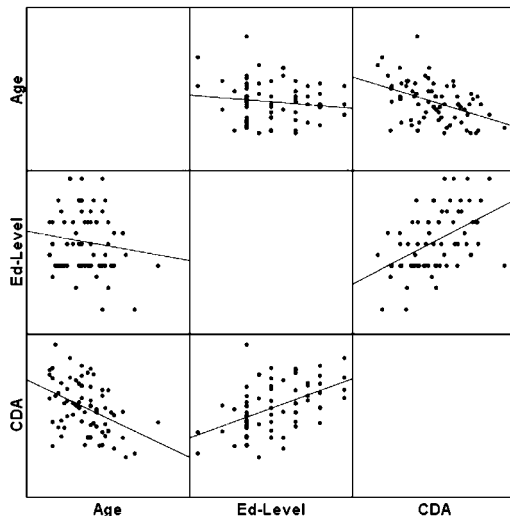


FIGURE 10.3.1 SPSS matrix scatter plot of the data in Table 10.3.1.

Dialog box:**Stat > Regression > Regression**Type **Y** in **Response** and **X1 X2**in **Predictors**.Check **Residuals**.Check **Standard resids**.Check **OK**.**Session command:**

```

MTB > Name C4 = `SRES1'
      C5 = `FITS1' C6 = `RESI1'
MTB > Regress `y' 2 `x1' `x2';
SUBC> SResiduals `SRES1';
SUBC> Fits `FITS1';
SUBC> Constant;
SUBC> Residuals `RESI1'.

```

Output:**Regression Analysis: Y versus X1, X2**

The regression equation is

$$Y = 5.49 - 0.184 X1 + 0.611 X2$$

Predictor	Coef	SE Coef	T	P
Constant	5.494	4.443	1.24	0.220
X1	-0.18412	0.04851	-3.80	0.000
X2	0.6108	0.1357	4.50	0.000

S = 3.134 R-Sq = 37.1% R-Sq (adj) = 35.2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	393.39	196.69	20.02	0.000
Residual Error	68	667.97	9.82		
Total	70	1061.36			

Source	DF	Seq SS
X1	1	194.24
X2	1	199.15

Unusual Observations

Obs	X1	Y	Fit	SE Fit	Residual	St Resid
28	67	9.610	0.487	0.707	9.123	2.99R
31	102	-2.300	-5.957	1.268	3.657	1.28X
44	80	5.070	-1.296	0.425	6.366	2.05R
67	94	-7.070	-6.927	1.159	-0.143	-0.05X

R denotes an observation with a large standardized residual.

X denotes an observation whose X value gives it large influence.

FIGURE 10.3.2 MINITAB procedure and output for Example 10.3.1.

The REG Procedure					
Model: MODEL1					
Dependent Variable: CDA					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	393.38832	196.69416	20.02	<.0001
Error	68	667.97084	9.82310		
Corrected Total	70	1061.35915			
Root MSE		3.13418	R-Square	0.3706	
Dependent Mean		0.00676	Adj R-Sq	0.3521	
Coeff Var		46360			
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	5.49407	4.44297	1.24	0.2205
AGE	1	-0.18412	0.04851	-3.80	0.0003
EDUC	1	0.61078	0.13565	4.50	<.0001

FIGURE 10.3.3 SAS® output for Example 10.3.1.

between CDA and Age and a positive relationship between CDA and Ed-Level. We shall see that this is indeed the case when we use MINITAB to analyze the data.

Solution: We enter the observations on age, education level, and CDA in c1 through c3 and name them X1, X2, and Y, respectively. The MINITAB dialog box and session command, as well as the output, are shown in Figure 10.3.2. We see from the output that the sample multiple regression equation, in the notation of Section 10.2, is

$$\hat{y}_j = 5.49 - .184x_{1j} + .611x_{2j}$$

Other output entries will be discussed in the sections that follow.

The SAS output for Example 10.3.1 is shown in Figure 10.3.3. ■

After the multiple regression equation has been obtained, the next step involves its evaluation and interpretation. We cover this facet of the analysis in the next section.

EXERCISES

Obtain the regression equation for each of the following data sets.

- 10.3.1** Machiel Naeije (A-2) studied the relationship between maximum mouth opening and measurements of the lower jaw (mandible). He measured the dependent variable, maximum mouth opening (MMO, measured in mm), as well as predictor variables, mandibular length (ML, measured in mm) and angle of rotation of the mandible (RA, measured in degrees) of 35 subjects.

MMO (Y)	ML (X_1)	RA (X_2)	MMO (Y)	ML (X_1)	RA (X_2)
52.34	100.85	32.08	50.82	90.65	38.33
51.90	93.08	39.21	40.48	92.99	25.93
52.80	98.43	33.74	59.68	108.97	36.78
50.29	102.95	34.19	54.35	91.85	42.02
57.79	108.24	35.13	47.00	104.30	27.20
49.41	98.34	30.92	47.23	93.16	31.37
53.28	95.57	37.71	41.19	94.18	27.87
59.71	98.85	44.71	42.76	89.56	28.69
53.32	98.32	33.17	51.88	105.85	31.04
48.53	92.70	31.74	42.77	89.29	32.78
51.59	88.89	37.07	52.34	92.58	37.82
58.52	104.06	38.71	50.45	98.64	33.36
62.93	98.18	43.89	43.18	83.70	31.93
57.62	91.01	41.06	41.99	88.46	28.32
65.64	96.98	41.92	39.45	94.93	24.82
52.85	97.85	35.25	38.91	96.81	23.88
64.43	96.89	45.11	49.10	93.13	36.17
57.25	98.35	39.44			

Source: Data provided courtesy of M. Naeije, D.D.S.

- 10.3.2** Family caregiving of older adults is more common in Korea than in the United States. Son et al. (A-3) studied 100 caregivers of older adults with dementia in Seoul, South Korea. The dependent variable was caregiver burden as measured by the Korean Burden Inventory (KBI). Scores ranged from 28 to 140, with higher scores indicating higher burden. Explanatory variables were indexes that measured the following:

ADL: total activities of daily living (low scores indicate that the elderly perform activities independently).

MEM: memory and behavioral problems (higher scores indicate more problems).

COG: cognitive impairment (lower scores indicate a greater degree of cognitive impairment).

The reported data are as follows:

KBI (Y)	ADL (X_1)	MEM (X_2)	COG (X_3)	KBI (Y)	ADL (X_1)	MEM (X_2)	COG (X_3)
28	39	4	18	88	76	50	5
68	52	33	9	54	79	44	11
59	89	17	3	73	48	57	9
91	57	31	7	87	90	33	6

(Continued)

KBI (Y)	ADL (X_1)	MEM (X_2)	COG (X_3)	KBI (Y)	ADL (X_1)	MEM (X_2)	COG (X_3)
70	28	35	19	47	55	11	20
38	34	3	25	60	83	24	11
46	42	16	17	65	50	21	25
57	52	6	26	57	44	31	18
89	88	41	13	85	79	30	20
48	90	24	3	28	24	5	22
74	38	22	13	40	40	20	17
78	83	41	11	87	35	15	27
43	30	9	24	80	55	9	21
76	45	33	14	49	45	28	17
72	47	36	18	57	46	19	17
61	90	17	0	32	37	4	21
63	63	14	16	52	47	29	3
77	34	35	22	42	28	23	21
85	76	33	23	49	61	8	7
31	26	13	18	63	35	31	26
79	68	34	26	89	68	65	6
92	85	28	10	67	80	29	10
76	22	12	16	43	43	8	13
91	82	57	3	47	53	14	18
78	80	51	3	70	60	30	16
103	80	20	18	99	63	22	18
99	81	20	1	53	28	9	27
73	30	7	17	78	35	18	14
88	27	27	27	112	37	33	17
64	72	9	0	52	82	25	13
52	46	15	22	68	88	16	0
71	63	52	13	63	52	15	0
41	45	26	18	49	30	16	18
85	77	57	0	42	69	49	12
52	42	10	19	56	52	17	20
68	60	34	11	46	59	38	17
57	33	14	14	72	53	22	21
84	49	30	15	95	65	56	2
91	89	64	0	57	90	12	0
83	72	31	3	88	88	42	6
73	45	24	19	81	66	12	23
57	73	13	3	104	60	21	7
69	58	16	15	88	48	14	13
81	33	17	21	115	82	41	13
71	34	13	18	66	88	24	14
91	90	42	6	92	63	49	5
48	48	7	23	97	79	34	3
94	47	17	18	69	71	38	17
57	32	13	15	112	66	48	13
49	63	32	15	88	81	66	1

Source: Data provided courtesy of Gwi-Ryung Son, R.N., Ph.D.

- 10.3.3** In a study of factors thought to be related to patterns of admission to a large general hospital, an administrator obtained these data on 10 communities in the hospital's catchment area:

Community	Persons per 1000 Population Admitted During Study Period (Y)	Index of Availability of Other Health Services (X_1)	Index of Indigency (X_2)
1	61.6	6.0	6.3
2	53.2	4.4	5.5
3	65.5	9.1	3.6
4	64.9	8.1	5.8
5	72.7	9.7	6.8
6	52.2	4.8	7.9
7	50.2	7.6	4.2
8	44.0	4.4	6.0
9	53.8	9.1	2.8
10	53.5	6.7	6.7
Total	571.6	69.9	55.6

- 10.3.4** The administrator of a general hospital obtained the following data on 20 surgery patients during a study to determine what factors appear to be related to length of stay:

Postoperative Length of Stay in Days (Y)	Number of Current Medical Problems (X_1)	Preoperative Length of Stay in Days (X_2)
6	1	1
6	2	1
11	2	2
9	1	3
16	3	3
16	1	5
4	1	1
8	3	1
11	2	2
13	3	2
13	1	4
9	1	2
17	3	3
17	2	4
12	4	1
6	1	1
5	1	1

(Continued)

Postoperative Length of Stay in Days (Y)	Number of Current Medical Problems (X ₁)	Preoperative Length of Stay in Days (X ₂)
12	3	2
8	1	2
9	2	2
Total 208	38	43

- 10.3.5** A random sample of 25 nurses selected from a state registry yielded the following information on each nurse's score on the state board examination and his or her final score in school. Both scores relate to the nurse's area of affiliation. Additional information on the score made by each nurse on an aptitude test, taken at the time of entering nursing school, was made available to the researcher. The complete data are as follows:

State Board Score (Y)	Final Score (X ₁)	Aptitude Test Score (X ₂)
440	87	92
480	87	79
535	87	99
460	88	91
525	88	84
480	89	71
510	89	78
530	89	78
545	89	71
600	89	76
495	90	89
545	90	90
575	90	73
525	91	71
575	91	81
600	91	84
490	92	70
510	92	85
575	92	71
540	93	76
595	93	90
525	94	94
545	94	94
600	94	93
625	94	73
Total 13,425	2263	2053

10.3.6 The following data were collected on a simple random sample of 20 patients with hypertension. The variables are

Y = mean arterial blood pressure(mm Hg)

X_1 = age(years)

X_2 = weight(kg)

X_3 = body surface area(sq m)

X_4 = duration of hypertension(years)

X_5 = basal pulse(beatsth/min)

X_6 = measure of stress

Patient	Y	X_1	X_2	X_3	X_4	X_5	X_6
1	105	47	85.4	1.75	5.1	63	33
2	115	49	94.2	2.10	3.8	70	14
3	116	49	95.3	1.98	8.2	72	10
4	117	50	94.7	2.01	5.8	73	99
5	112	51	89.4	1.89	7.0	72	95
6	121	48	99.5	2.25	9.3	71	10
7	121	49	99.8	2.25	2.5	69	42
8	110	47	90.9	1.90	6.2	66	8
9	110	49	89.2	1.83	7.1	69	62
10	114	48	92.7	2.07	5.6	64	35
11	114	47	94.4	2.07	5.3	74	90
12	115	49	94.1	1.98	5.6	71	21
13	114	50	91.6	2.05	10.2	68	47
14	106	45	87.1	1.92	5.6	67	80
15	125	52	101.3	2.19	10.0	76	98
16	114	46	94.5	1.98	7.4	69	95
17	106	46	87.0	1.87	3.6	62	18
18	113	46	94.5	1.90	4.3	70	12
19	110	48	90.5	1.88	9.0	71	99
20	122	56	95.7	2.09	7.0	75	99

10.4 EVALUATING THE MULTIPLE REGRESSION EQUATION

Before one uses a multiple regression equation to predict and estimate, it is desirable to determine first whether it is, in fact, worth using. In our study of simple linear regression we have learned that the usefulness of a regression equation may be evaluated by a consideration of the sample coefficient of determination and estimated slope. In evaluating a multiple regression equation we focus our attention on the *coefficient of multiple determination* and the partial regression coefficients.

The Coefficient of Multiple Determination In Chapter 9 the coefficient of determination is discussed in considerable detail. The concept extends logically

to the multiple regression case. The total variation present in the Y values may be partitioned into two components—the explained variation, which measures the amount of the total variation that is explained by the fitted regression surface, and the unexplained variation, which is that part of the total variation not explained by fitting the regression surface. The measure of variation in each case is a sum of squared deviations. The total variation is the sum of squared deviations of each observation of Y from the mean of the observations and is designated by $\sum (y_j - \bar{y})^2$ or SST . The explained variation, designated $\sum (\hat{y}_j - \bar{y})^2$ or SSR , is the sum of squared deviations of the calculated values from the mean of the observed Y values. This sum of squared deviations is called the *sum of squares due to regression* (SSR). The unexplained variation, written as $\sum (y_j - \hat{y}_j)^2$, is the sum of squared deviations of the original observations from the calculated values. This quantity is referred to as the *sum of squares about regression* or the *error sum of squares* (SSE). We may summarize the relationship among the three sums of squares with the following equation:

$$\begin{aligned} \sum (y_j - \bar{y})^2 &= \sum (\hat{y}_j - \bar{y})^2 + \sum (y_j - \hat{y}_j)^2 \\ SST &= SSR + SSE \\ \text{total sum of squares} &= \text{explained(regression)sum of squares} \\ &\quad + \text{unexplained(error)sum of squares} \end{aligned} \tag{10.4.1}$$

The coefficient of multiple determination, $R_{y.12\dots k}^2$ is obtained by dividing the explained sum of squares by the total sum of squares. That is,

$$R_{y.12\dots k}^2 = \frac{\sum (\hat{y}_j - \bar{y})^2}{\sum (y_j - \bar{y})^2} = \frac{SSR}{SST} \tag{10.4.2}$$

The subscript $y.12\dots k$ indicates that in the analysis Y is treated as the dependent variable and the X variables from X_1 through X_k are treated as the independent variables. The value of $R_{y.12\dots k}^2$ indicates what proportion of the total variation in the observed Y values is explained by the regression of Y on X_1, X_2, \dots, X_k . In other words, we may say that $R_{y.12\dots k}^2$ is a measure of the goodness of fit of the regression surface. This quantity is analogous to r^2 , which was computed in Chapter 9.

EXAMPLE 10.4.1

Refer to Example 10.3.1. Compute $R_{y.12}^2$.

Solution: For our illustrative example we have in Figure 10.3.1

$$\begin{aligned} SST &= 1061.36 \\ SSR &= 393.39 \\ SSE &= 667.97 \\ R_{y.12}^2 &= \frac{393.39}{1061.36} = .3706 \approx .371 \end{aligned}$$

We say that about 37.1 percent of the total variation in the Y values is explained by the fitted regression plane, that is, by the linear relationship with age and education level. ■

Testing the Regression Hypothesis To determine whether the overall regression is significant (that is, to determine whether $R_{y,12}^2$ is significant), we may perform a hypothesis test as follows.

1. **Data.** The research situation and the data generated by the research are examined to determine if multiple regression is an appropriate technique for analysis.
2. **Assumptions.** We assume that the multiple regression model and its underlying assumptions as presented in Section 10.2 are applicable.
3. **Hypotheses.** In general, the null hypothesis is $H_0: \beta_1 = \beta_2 = \beta_3 = \cdots = \beta_k = 0$ and the alternative is H_A : not all $\beta_i = 0$. In words, the null hypothesis states that all the independent variables are of no value in explaining the variation in the Y values.
4. **Test statistic.** The appropriate test statistic is V.R., which is computed as part of an analysis of variance. The general ANOVA table is shown as Table 10.4.1. In Table 10.4.1, MSR stands for mean square due to regression and MSE stands for mean square about regression or, as it is sometimes called, the error mean square.
5. **Distribution of test statistic.** When H_0 is true and the assumptions are met, V.R. is distributed as F with k and $n - k - 1$ degrees of freedom.
6. **Decision rule.** Reject H_0 if the computed value of V.R. is equal to or greater than the critical value of F .
7. **Calculation of test statistic.** See Table 10.4.1.
8. **Statistical decision.** Reject or fail to reject H_0 in accordance with the decision rule.
9. **Conclusion.** If we reject H_0 we conclude that, in the population from which the sample was drawn, the dependent variable is linearly related to the independent variables as a group. If we fail to reject H_0 , we conclude that, in the population from which our sample was drawn, there may be no linear relationship between the dependent variable and the independent variables as a group.
10. **p value.** We obtain the p value from the table of the F distribution.

We illustrate the hypothesis testing procedure by means of the following example.

TABLE 10.4.1 ANOVA Table for Multiple Regression

Source	SS	d.f.	MS	V.R.
Due to regression	SSR	k	$MSR = SSR/k$	MSR/MSE
About regression	SSE	$n - k - 1$	$MSE = SSE/(n - k - 1)$	
Total	SST	$n - 1$		

EXAMPLE 10.4.2

We wish to test the null hypothesis of no linear relationship among the three variables discussed in Example 10.3.1: CDA score, age, and education level.

Solution:

- Data.** See the description of the data given in Example 10.3.1.
- Assumptions.** We assume that the assumptions discussed in Section 10.2 are met.
- Hypotheses.**

$$\begin{aligned} H_0: &= \beta_1 = \beta_2 = 0 \\ H_A: &= \text{not all } \beta_i = 0 \end{aligned}$$

- Test statistic.** The test statistic is V.R.
- Distribution of test statistic.** If H_0 is true and the assumptions are met, the test statistic is distributed as F with 2 numerator and 68 denominator degrees of freedom.
- Decision rule.** Let us use a significance level of $\alpha = .01$. The decision rule, then, is reject H_0 if the computed value of V.R. is equal to or greater than 4.95 (obtained by interpolation).
- Calculation of test statistic.** The ANOVA for the example is shown in Figure 10.3.1, where we see that the computed value of V.R. is 20.02.
- Statistical decision.** Since 20.02 is greater than 4.95, we reject H_0 .
- Conclusion.** We conclude that, in the population from which the sample came, there is a linear relationship among the three variables.
- p value.** Since 20.02 is greater than 5.76, the p value for the test is less than .005. ■

Inferences Regarding Individual β 's Frequently, we wish to evaluate the strength of the linear relationship between Y and the independent variables individually. That is, we may want to test the null hypothesis that $\beta_i = 0$ against the alternative $\beta_i \neq 0$ ($i = 1, 2, \dots, k$). The validity of this procedure rests on the assumptions stated earlier: that for each combination of X_i values there is a normally distributed subpopulation of Y values with variance σ^2 .

Hypothesis Tests for the β_i To test the null hypothesis that β_i is equal to some particular value, say, β_{i0} , the following t statistic may be computed:

$$t = \frac{\hat{\beta}_i - \beta_{i0}}{s_{\hat{\beta}_i}} \quad (10.4.3)$$

where the degrees of freedom are equal to $n - k - 1$, and $s_{\hat{\beta}_i}$ is the standard deviation of the $\hat{\beta}_i$.

The standard deviations of the $\hat{\beta}_i$ are given as part of the output from most computer software packages that do regression analysis.

EXAMPLE 10.4.3

Let us refer to Example 10.3.1 and test the null hypothesis that age (years) is irrelevant in predicting the capacity to direct attention (CDA).

Solution:

1. **Data.** See Example 10.3.1.
2. **Assumptions.** See Section 10.2.
3. **Hypotheses.**

$$H_0: \beta_1 = 0$$

$$H_A: \beta_1 \neq 0$$

$$\text{Let } \alpha = .05$$

4. **Test statistic.** See Equation 10.4.3.
5. **Distribution of test statistic.** When H_0 is true and the assumptions are met, the test statistic is distributed as Student's t with 68 degrees of freedom.
6. **Decision rule.** Reject H_0 if the computed t is either greater than or equal to 1.9957 (obtained by interpolation) or less than or equal to -1.9957 .
7. **Calculation of test statistic.** By Equation 10.4.3 and data from Figure 10.3.2 we compute

$$t = \frac{\hat{\beta}_1 - 0}{s_{\hat{\beta}_1}} = \frac{-.18412}{.04851} = -3.80$$

8. **Statistical decision.** The null hypothesis is rejected since the computed value of t , -3.80 , is less than -1.9957 .
9. **Conclusion.** We conclude, then, that there is a linear relationship between age and CDA in the presence of education level.
10. **p value.** For this test, $p < 2(.005) = .01$ because $-3.80 < -2.6505$ (obtained by interpolation). As shown in Figure 10.3.2, the p -value is $<.001$ for this test.

Now, let us perform a similar test for the second partial regression coefficient, β_2 :

$$H_0: \beta_2 = 0$$

$$H_A: \beta_2 \neq 0$$

$$\alpha = .05$$

$$t = \frac{\hat{\beta}_2 - 0}{s_{\hat{\beta}_2}} = \frac{.6108}{.1357} = 4.50$$

In this case also the null hypothesis is rejected, since 4.50 is greater than 1.9957. We conclude that there is a linear relationship between education level and CDA in the presence age, and that education level, used in this manner, is a useful variable for predicting CDA. [For this test, $p < 2(.005) = .01$.

Confidence Intervals for the β_i When the researcher has been led to conclude that a partial regression coefficient is not 0, he or she may be interested in obtaining a confidence interval for this β_i . Confidence intervals for the β_i may be constructed in the usual way by using a value from the t distribution for the reliability factor and standard errors given above.

A $100(1 - \alpha)$ percent confidence interval for β_i is given by

$$\hat{\beta}_i \pm t_{1-(\alpha/2), n-k-1} s_{\hat{\beta}_i}$$

For our illustrative example we may compute the following 95 percent confidence intervals for β_1 and β_2 .

The 95 percent confidence interval for β_1 is

$$\begin{aligned} &-.18412 \pm 1.9957(.04851) \\ &-.18412 \pm .0968 \\ &(-.28092, -.08732) \end{aligned}$$

The 95 percent confidence interval for β_2 is

$$\begin{aligned} &.6108 \pm (1.9957)(.1357) \\ &.6108 \pm .2708 \\ &(.3400, .8816) \end{aligned}$$

We may give these intervals the usual probabilistic and practical interpretations. We are 95 percent confident, for example, that β_2 is contained in the interval from .3400 to .8816 since, in repeated sampling, 95 percent of the intervals that may be constructed in this manner will include the true parameter.

Some Precautions One should be aware of the problems involved in carrying out multiple hypothesis tests and constructing multiple confidence intervals from the same sample data. The effect on α of performing multiple hypothesis tests from the same data is discussed in Section 8.2. A similar problem arises when one wishes to construct confidence intervals for two or more partial regression coefficients. The intervals will not be independent, so that the tabulated confidence coefficient does not, in general, apply. In other words, all such intervals would not be $100(1 - \alpha)$ percent confidence intervals.

In order to maintain approximate $100(1 - \alpha)$ confidence intervals for partial regression coefficients, adjustments must be made to the calculation of errors in the previous equations. These adjustments are sometimes called *family-wise error rates*, and can be found in many computer software packages. The topic is discussed in detail by Kutner, et al. (4).

Another problem sometimes encountered in the application of multiple regression is an apparent incompatibility in the results of the various tests of significance that one may perform. In a given problem for a given level of significance, one or the other of the following situations may be observed.

1. R^2 and all $\hat{\beta}_i$ significant
2. R^2 and some but not all $\hat{\beta}_i$ significant
3. R^2 significant but none of the $\hat{\beta}_i$ significant
4. All $\hat{\beta}_i$ significant but not R^2
5. Some $\hat{\beta}_i$ significant, but not all nor R^2
6. Neither R^2 nor any $\hat{\beta}_i$ significant

Notice that situation 1 exists in our illustrative example, where we have a significant R^2 and two significant regression coefficients. This situation does not occur in all cases. In fact, situation 2 is very common, especially when a large number of independent variables have been included in the regression equation.

EXERCISES

- 10.4.1** Refer to Exercise 10.3.1. (a) Calculate the coefficient of multiple determination; (b) perform an analysis of variance; (c) test the significance of each $\hat{\beta}_i (i > 0)$. Let $\alpha = .05$ for all tests of significance and determine the p value for all tests; (d) construct a 95 percent confidence interval for each significant sample slope.
- 10.4.2** Refer to Exercise 10.3.2. Do the analysis suggested in Exercise 10.4.1.
- 10.4.3** Refer to Exercise 10.3.3. Do the analysis suggested in Exercise 10.4.1.
- 10.4.4** Refer to Exercise 10.3.4. Do the analysis suggested in Exercise 10.4.1.
- 10.4.5** Refer to Exercise 10.3.5. Do the analysis suggested in Exercise 10.4.1.
- 10.4.6** Refer to Exercise 10.3.6. Do the analysis suggested in Exercise 10.4.1.

10.5 USING THE MULTIPLE REGRESSION EQUATION

As we learned in the previous chapter, a regression equation may be used to obtain a computed value of Y , \hat{y} , when a particular value of X is given. Similarly, we may use our multiple regression equation to obtain a \hat{y} value when we are given particular values of the two or more X variables present in the equation.

Just as was the case in simple linear regression, we may, in multiple regression, interpret a \hat{y} value in one of two ways. First we may interpret \hat{y} as an estimate of the mean of the subpopulation of Y values assumed to exist for particular combinations of X_i values. Under this interpretation \hat{y} is called an *estimate*, and when it is used for this purpose, the equation is thought of as an *estimating equation*. The second interpretation of \hat{y} is that it is the value Y is most likely to assume for given values of the X_i . In this case \hat{y}

is called the *predicted value* of Y , and the equation is called a *prediction equation*. In both cases, intervals may be constructed about the \hat{y} value when the normality assumption of Section 10.2 holds true. When \hat{y} is interpreted as an estimate of a population mean, the interval is called a *confidence interval*, and when \hat{y} is interpreted as a predicted value of Y , the interval is called a *prediction interval*. Now let us see how each of these intervals is constructed.

The Confidence Interval for the Mean of a Subpopulation of Y Values Given Particular Values of the X_i

We have seen that a $100(1 - \alpha)$ percent confidence interval for a parameter may be constructed by the general procedure of adding to and subtracting from the estimator a quantity equal to the reliability factor corresponding to $1 - \alpha$ multiplied by the standard error of the estimator. We have also seen that in multiple regression the estimator is

$$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_{1j} + \hat{\beta}_2 x_{2j} + \cdots + \hat{\beta}_k x_{kj} \quad (10.5.1)$$

If we designate the standard error of this estimator by $s_{\hat{y}_j}$, the $100(1 - \alpha)$ percent confidence interval for the mean of Y , given specified X_i is as follows:

$$\hat{y}_j \pm t_{(1-\alpha/2), n-k-1} s_{\hat{y}_j} \quad (10.5.2)$$

The Prediction Interval for a Particular Value of Y Given Particular Values of the X_i

When we interpret \hat{y} as the value Y is most likely to assume when particular values of the X_i are observed, we may construct a prediction interval in the same way in which the confidence interval was constructed. The only difference in the two is the standard error. The standard error of the prediction is slightly larger than the standard error of the estimate, which causes the prediction interval to be wider than the confidence interval.

If we designate the standard error of the prediction by $s'_{\hat{y}_j}$, the $100(1 - \alpha)$ percent prediction interval is

$$\hat{y}_j \pm t_{(1-\alpha/2), n-k-1} s'_{\hat{y}_j} \quad (10.5.3)$$

The calculations of $s_{\hat{y}_j}$ and $s'_{\hat{y}_j}$ in the multiple regression case are complicated and will not be covered in this text. The reader who wishes to see how these statistics are calculated may consult the book by Anderson and Bancroft (3), other references listed at the end of this chapter and Chapter 9, and previous editions of this text. The following example illustrates how MINITAB may be used to obtain confidence intervals for the mean of Y and prediction intervals for a particular value of Y .

EXAMPLE 10.5.1

We refer to Example 10.3.1. First, we wish to construct a 95 percent confidence interval for the mean CDA score (Y) in a population of 68-year-old subjects (X_1) who completed 12 years of education (X_2). Second, suppose we have a subject who is 68 years of age

and has an education level of 12 years. What do we predict to be this subject's CDA score?

Solution: The point estimate of the mean CDA score is

$$\hat{y} = 5.494 - .18412(68) + .6108(12) = .3034$$

The point prediction, which is the same as the point estimate obtained previously, also is

$$\hat{y} = 5.494 - .18412(68) + .6108(12) = .3034$$

To obtain the confidence interval and the prediction interval for the parameters for which we have just computed a point estimate and a point prediction, we use MINITAB as follows. After entering the information for a regression analysis of our data as shown in Figure 10.3.2, we click on Options in the dialog box. In the box labeled "Prediction intervals for new observations," we type 68 and 12 and click OK twice. In addition to the regression analysis, we obtain the following output:

New Obs	Fit	SE Fit	95.0% CI	95.0% PI
1	0.303	0.672	(-1.038, 1.644)	(-6.093, 6.699)

We interpret these intervals in the usual ways. We look first at the confidence interval. We are 95 percent confident that the interval from -1.038 to 1.644 includes the mean of the subpopulation of Y values for the specified combination of X_i values, since this parameter would be included in about 95 percent of the intervals that can be constructed in the manner shown.

Now consider the subject who is 68 years old and has 12 years of education. We are 95 percent confident that this subject would have a CDA score somewhere between -6.093 and 6.699 . The fact that the P.I. is wider than the C.I. should not be surprising. After all, it is easier to estimate the mean response than it is estimate an individual observation. ■

EXERCISES

For each of the following exercises compute the y value and construct (a) 95 percent confidence and (b) 95 percent prediction intervals for the specified values of X_i .

- 10.5.1** Refer to Exercise 10.3.1 and let $x_{1j} = 95$ and $x_{2j} = 35$.
10.5.2 Refer to Exercise 10.3.2 and let $x_{1j} = 50$, $x_{2j} = 20$, and $x_{3j} = 22$.
10.5.3 Refer to Exercise 10.3.3 and let $x_{1j} = 5$ and $x_{2j} = 6$.
10.5.4 Refer to Exercise 10.3.4 and let $x_{1j} = 1$ and $x_{2j} = 2$.

10.5.5 Refer to Exercise 10.3.5 and let $x_{1j} = 90$ and $x_{2j} = 80$.

10.5.6 Refer to Exercise 10.3.6 and let $x_{1j} = 50$, $x_{2j} = 95.0$, $x_{3j} = 2.00$, $x_{4j} = 6.00$, $x_{5j} = 75$, and $x_{6j} = 70$.

10.6 THE MULTIPLE CORRELATION MODEL

We pointed out in the preceding chapter that while regression analysis is concerned with the form of the relationship between variables, the objective of correlation analysis is to gain insight into the strength of the relationship. This is also true in the multivariable case, and in this section we investigate methods for measuring the strength of the relationship among several variables. First, however, let us define the model and assumptions on which our analysis rests.

The Model Equation We may write the correlation model as

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj} + \epsilon_j \quad (10.6.1)$$

where y_j is a typical value from the population of values of the variable Y , the β 's are the regression coefficients defined in Section 10.2, and the x_{ij} are particular (known) values of the random variables X_i . This model is similar to the multiple regression model, but there is one important distinction. In the multiple regression model, given in Equation 10.2.1, the X_i are nonrandom variables, but in the multiple correlation model the X_i are random variables. In other words, in the correlation model there is a joint distribution of Y and the X_i that we call a *multivariate distribution*. Under this model, the variables are no longer thought of as being dependent or independent, since logically they are interchangeable and either of the X_i may play the role of Y .

Typically, random samples of units of association are drawn from a population of interest, and measurements of Y and the X_i are made.

A least-squares plane or hyperplane is fitted to the sample data by methods described in Section 10.3, and the same uses may be made of the resulting equation. Inferences may be made about the population from which the sample was drawn if it can be assumed that the underlying distribution is normal, that is, if it can be assumed that the joint distribution of Y and X_i is a *multivariate normal distribution*. In addition, sample measures of the degree of the relationship among the variables may be computed and, under the assumption that sampling is from a multivariate normal distribution, the corresponding parameters may be estimated by means of confidence intervals, and hypothesis tests may be carried out. Specifically, we may compute an estimate of the *multiple correlation coefficient* that measures the dependence between Y and the X_i . This is a straightforward extension of the concept of correlation between two variables that we discuss in Chapter 9. We may also compute *partial correlation coefficients* that measure the intensity of the relationship between any two variables when the influence of all other variables has been removed.

The Multiple Correlation Coefficient As a first step in analyzing the relationships among the variables, we look at the multiple correlation coefficient.

The multiple correlation coefficient is the square root of the coefficient of multiple determination and, consequently, the sample value may be computed by taking the square root of Equation 10.4.2. That is,

$$R_{y.12\dots k} = \sqrt{R_{y.12\dots k}^2} = \sqrt{\frac{\sum (\hat{y}_j - \bar{y})^2}{\sum (y_j - \bar{y})^2}} = \sqrt{\frac{SSR}{SST}} \quad (10.6.2)$$

To illustrate the concepts and techniques of multiple correlation analysis, let us consider an example.

EXAMPLE 10.6.1

Wang et al. (A-4), using cadaveric human femurs from subjects ages 16 to 19 years, investigated toughness properties of the bone and measures of the collagen network within the bone. Two variables measuring the collagen network are porosity (P , expressed as a percent) and a measure of collagen network tensile strength (S). The measure of toughness (W , Newtons), is the force required for bone fracture. The 29 cadaveric femurs used in the study were free from bone-related pathologies. We wish to analyze the nature and strength of the relationship among the three variables. The measurements are shown in the following table.

TABLE 10.6.1 Bone Toughness and Collagen Network Properties for 29 Femurs

W	P	S
193.6	6.24	30.1
137.5	8.03	22.2
145.4	11.62	25.7
117.0	7.68	28.9
105.4	10.72	27.3
99.9	9.28	33.4
74.0	6.23	26.4
74.4	8.67	17.2
112.8	6.91	15.9
125.4	7.51	12.2
126.5	10.01	30.0
115.9	8.70	24.0
98.8	5.87	22.6
94.3	7.96	18.2
99.9	12.27	11.5
83.3	7.33	23.9
72.8	11.17	11.2
83.5	6.03	15.6
59.0	7.90	10.6

(Continued)

W	P	S
87.2	8.27	24.7
84.4	11.05	25.6
78.1	7.61	18.4
51.9	6.21	13.5
57.1	7.24	12.2
54.7	8.11	14.8
78.6	10.05	8.9
53.7	8.79	14.9
96.0	10.40	10.3
89.0	11.72	15.4

Source: Data provided courtesy of Xiaodu Wang, Ph.D.

Solution: We use MINITAB to perform the analysis of our data. Readers interested in the derivation of the underlying formulas and the arithmetic procedures involved may consult the texts listed at the end of this chapter and Chapter 9, as well as previous editions of this text. If a least-squares prediction equation and multiple correlation coefficient are desired as part of the analysis, we may obtain them by using the previously described MINITAB multiple regression procedure. When we do this with the sample values of Y , X_1 , and X_2 , stored in Columns 1 through 3, respectively, we obtain the output shown in Figure 10.6.1.

The least-squares equation, then, is

$$\hat{y}_j = 35.61 + 1.451x_{1j} + 2.3960x_{2j}$$

The regression equation is
 $Y = 35.6 + 1.45 X1 + 2.40 X2$

Predictor	Coef	SE Coef	T	P
Constant	35.61	29.13	1.22	0.232
X1	1.451	2.763	0.53	0.604
X2	2.3960	0.7301	3.28	0.003

S = 27.42 R-Sq = 29.4% R-Sq(adj) = 24.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	8151.1	4075.6	5.42	0.011
Residual Error	26	19553.5	752.1		
Total	28	27704.6			

FIGURE 10.6.1 Output from MINITAB multiple regression procedure for the data in Table 10.6.1.

This equation may be used for estimation and prediction purposes and may be evaluated by the methods discussed in Section 10.4.

As we see in Figure 10.6.1, the multiple regression output also gives us the coefficient of multiple determination, which, in our present example, is

$$R_{y,12}^2 = .294$$

The multiple correlation coefficient, therefore, is

$$R_{y,12} = \sqrt{.294} = .542$$

Interpretation of $R_{y,12}$

We interpret $R_{y,12}$ as a measure of the correlation among the variables force required to fracture, porosity, and collagen network strength in the sample of 29 femur bones from subjects ages 16 to 19. If our data constitute a random sample from the population of such persons, we may use $R_{y,12}$ as an estimate of $\rho_{y,12}$, the true population multiple correlation coefficient. We may also interpret $R_{y,12}$ as the simple correlation coefficient between y_j and \hat{y} , the observed and calculated values, respectively, of the “dependent” variable. Perfect correspondence between the observed and calculated values of Y will result in a correlation coefficient of 1, while a complete lack of a linear relationship between observed and calculated values yields a correlation coefficient of 0. The multiple correlation coefficient is always given a positive sign.

We may test the null hypothesis that $\rho_{y,12\dots k} = 0$ by computing

$$F = \frac{R_{y,12\dots k}^2}{1 - R_{y,12\dots k}^2} \cdot \frac{n - k - 1}{k} \quad (10.6.3)$$

The numerical value obtained from Equation 10.6.3 is compared with the tabulated value of F with k and $n - k - 1$ degrees of freedom. The reader will recall that this is identical to the test of $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$ described in Section 10.4.

For our present example let us test the null hypothesis that $\rho_{y,12} = 0$ against the alternative that $\rho_{y,12} \neq 0$. We compute

$$F = \frac{.294}{1 - .294} \cdot \frac{29 - 2 - 1}{2} = 5.41$$

Since 5.41 is greater than 4.27, $p < .025$, so that we may reject the null hypothesis at the .025 level of significance and conclude that the force required for fracture is correlated with porosity and the measure of collagen network strength in the sampled population.

The computed value of F for testing H_0 that the population multiple correlation coefficient is equal to zero is given in the analysis of variance table in Figure 10.6.1 and is 5.42. The two computed values of F differ as a result of differences in rounding in the intermediate calculations. ■

Partial Correlation The researcher may wish to have a measure of the strength of the linear relationship between two variables when the effect of the remaining variables has been removed. Such a measure is provided by the *partial correlation* coefficient.

For example, the partial sample correlation coefficient $r_{y.12}$ is a measure of the correlation between Y and X_1 after controlling for the effect of X_2 .

The partial correlation coefficients may be computed from the *simple correlation coefficients*. The simple correlation coefficients measure the correlation between two variables when no effort has been made to control other variables. In other words, they are the coefficients for any pair of variables that would be obtained by the methods of simple correlation discussed in Chapter 9.

Suppose we have three variables, Y , X_1 , and X_2 . The sample partial correlation coefficient measuring the correlation between Y and X_1 after controlling for X_2 , for example, is written $r_{y1.2}$. In the subscript, the symbol to the right of the decimal point indicates the variable whose effect is being controlled, while the two symbols to the left of the decimal point indicate which variables are being correlated. For the three-variable case, there are two other sample partial correlation coefficients that we may compute. They are $r_{y2.1}$ and $r_{12.y}$.

The Coefficient of Partial Determination The square of the partial correlation coefficient is called the coefficient of partial determination. It provides useful information about the interrelationships among variables. Consider $r_{y1.2}$, for example. Its square, $r_{y1.2}^2$ tells us what proportion of the remaining variability in Y is explained by X_1 after X_2 has explained as much of the total variability in Y as it can.

Calculating the Partial Correlation Coefficients For three variables the following simple correlation coefficients may be calculated:

- r_{y1} , the simple correlation between Y and X_1
- r_{y2} , the simple correlation between Y and X_2
- r_{12} , the simple correlation between X_1 and X_2

The MINITAB correlation procedure may be used to compute these simple correlation coefficients as shown in Figure 10.6.2. As noted earlier, the sample observations are stored in Columns 1 through 3. From the output in Figure 10.6.2 we see that $r_{12} = -.08$, $r_{y1} = .043$, and $r_{y2} = .535$.

The sample partial correlation coefficients that may be computed from the simple correlation coefficients in the three-variable case are:

1. The partial correlation between Y and X_1 after controlling for the effect of X_2 :

$$r_{y1.2} = (r_{y1} - r_{y2}r_{12}) / \sqrt{(1-r_{y2}^2)(1-r_{12}^2)} \quad (10.6.4)$$

2. The partial correlation between Y and X_2 after controlling for the effect of X_1 :

$$r_{y2.1} = (r_{y2} - r_{y1}r_{12}) / \sqrt{(1-r_{y1}^2)(1-r_{12}^2)} \quad (10.6.5)$$

3. The partial correlation between X_1 and X_2 after controlling for the effect of Y :

$$r_{12.y} = (r_{12} - r_{y1}r_{y2}) / \sqrt{(1-r_{y1}^2)(1-r_{y2}^2)} \quad (10.6.6)$$

<p>Dialog box:</p> <p>Stat > Basic Statistics > Correlation</p> <p>Type <i>C1-C3</i> in Variables. Click OK.</p> <p>Output:</p> <table> <thead> <tr> <th></th> <th>Y</th> <th>X1</th> </tr> </thead> <tbody> <tr> <td>X1</td> <td>0.043 0.823</td> <td></td> </tr> <tr> <td>X2</td> <td>0.535 0.003</td> <td>-0.080 0.679</td> </tr> </tbody> </table> <p>Cell Contents: Pearson correlation P-Value</p>		Y	X1	X1	0.043 0.823		X2	0.535 0.003	-0.080 0.679	<p>Session Command:</p> <p>MTB> CORRELATION C1-C3</p>
	Y	X1								
X1	0.043 0.823									
X2	0.535 0.003	-0.080 0.679								

FIGURE 10.6.2 MINITAB procedure for calculating the simple correlation coefficients for the data in Table 10.6.1.

EXAMPLE 10.6.2

To illustrate the calculation of sample partial correlation coefficients, let us refer to Example 10.6.1, and calculate the partial correlation coefficients among the variables force to fracture (Y), porosity (X_1), and collagen network strength (X_2).

Solution: Instead of computing the partial correlation coefficients from the simple correlation coefficients by Equations 10.6.4 through 10.6.6, we use MINITAB to obtain them.

The MINITAB procedure for computing partial correlation coefficients is based on the fact that a given partial correlation coefficient is itself the simple correlation between two sets of residuals. A set of residuals is obtained as follows. Suppose we have measurements on two variables, X (independent) and Y (dependent). We obtain the least-squares prediction equation, $\hat{y} = \hat{\beta}_0 + \hat{\beta}_x X$. For each value of X we compute a residual, which is equal to $(y_i - \hat{y}_i)$, the difference between the observed value of Y and the predicted value of Y associated with the X .

Now, suppose we have three variables, X_1 , X_2 , and Y . We want to compute the partial correlation coefficient between X_1 and Y while holding X_2 constant. We regress X_1 on X_2 and compute the residuals, which we may call residual set A . We regress Y on X_2 and compute the residuals, which we may call residual set B . The simple correlation coefficient measuring the strength of the relationship between residual set A and residual set B is the partial correlation coefficient between X_1 and Y after controlling for the effect of X_2 .

When using MINITAB we store each set of residuals in a different column for future use in calculating the simple correlation coefficients between them.

We use session commands rather than a dialog box to calculate the partial correlation coefficients when we use MINITAB. With the observations on X_1 , X_2 , and Y stored in Columns 1 through 3, respectively, the procedure for the data of Table 10.6.1 is shown in Figure 10.6.3. The output shows that $r_{y1.2} = .102$, $r_{12.y} = -.122$, and $r_{y2.1} = .541$.

Partial correlations can be calculated directly using SPSS software as seen in Figure 10.6.5. This software displays, in a succinct table, both the partial correlation coefficient and the p value associated with each partial correlation. ■

Testing Hypotheses About Partial Correlation Coefficients We may test the null hypothesis that any one of the population partial correlation coefficients is 0 by means of the t test. For example, to test $H_0: \rho_{y1.2\dots k} = 0$, we compute

$$t = r_{y1.2\dots k} \sqrt{\frac{n - k - 1}{1 - r_{y1.2\dots k}^2}} \quad (10.6.7)$$

which is distributed as Student's t with $n - k - 1$ degrees of freedom.

Let us illustrate the procedure for our current example by testing $H_0: \rho_{y1.2} = 0$ against the alternative, $H_A: \rho_{y1.2} \neq 0$. The computed t is

$$t = .102 \sqrt{\frac{29 - 2 - 1}{1 - (.102)^2}} = .523$$

Since the computed t of .523 is smaller than the tabulated t of 2.0555 for 26 degrees of freedom and $\alpha = .05$ (two-sided test), we fail to reject H_0 at the .05 level of significance and conclude that there may be no correlation between force required for fracture and porosity after controlling for the effect of collagen network strength. Significance tests for the other two partial correlation coefficients will be left as an exercise for the reader. Note that p values for these tests are calculated by MINITAB as shown in Figure 10.6.3.

The SPSS statistical software package for the PC provides a convenient procedure for obtaining partial correlation coefficients. To use this feature choose "Analyze" from the menu bar, then "Correlate," and, finally, "Partial." Following this sequence of choices the Partial Correlations dialog box appears on the screen. In the box labeled "Variables;" enter the names of the variables for which partial correlations are desired. In the box labeled "Controlling for;" enter the names of the variable(s) for which you wish to control. Select either a two-tailed or one-tailed level of significance. Unless the option is deselected, actual significance levels will be displayed. For Example 10.6.2, Figure 10.6.4 shows the SPSS computed partial correlation coefficients between the other two variables when controlling, successively, for X_1 (porosity), X_2 (collagen network strength), and Y (force required for fracture).

```
MTB > regress C1 1 C2;  
SUBC> residuals C4.
```

```
MTB > regress C3 1 C2;  
SUBC> residuals C5.
```

```
MTB > regress C1 1 C3;  
SUBC> residuals C6.
```

```
MTB > regress C2 1 C3;  
SUBC> residuals C7.
```

```
MTB > regress C2 1 C1;  
SUBC> residuals C8.
```

```
MTB > regress C3 1 C1;  
SUBC> residuals C9.
```

```
MTB > corr C4 C5
```

Correlations: C4, C5

```
Pearson correlation of C4 and C5 = 0.102  
P-Value = 0.597
```

```
MTB > corr C6 C7
```

Correlations: C6, C7

```
Pearson correlation of C6 and C7 = -0.122  
P-Value = 0.527
```

```
MTB > corr C8 C9
```

Correlations: C8, C9

```
Pearson correlation of C8 and C9 = 0.541  
P-Value = 0.002
```

FIGURE 10.6.3 MINITAB procedure for computing partial correlation coefficients from the data of Table 10.6.1.

```

Controlling for: X1

           X2           Y

X2      1.0000      .5412
      (    0)      (   26)
      P= .          P= .003

Y        .5412      1.0000
      (   26)      (    0)
      P= .003      P= .

Controlling for: X2

           Y           X1

Y        1.0000      .1024
      (    0)      (   26)
      P= .          P= .604

X1       .1024      1.0000
      (   26)      (    0)
      P= .604      P= .

Controlling for: Y

           X1           X2

X1       1.0000      -.1225
      (    0)      (   26)
      P= .          P= .535

X2      -.1225      1.0000
      (   26)      (    0)
      P= .535      P= .

(Coefficient / (D.F.) / 2-tailed Significance)
"." is printed if a coefficient cannot be computed

```

FIGURE 10.6.4 Partial coefficients obtained with SPSS for Windows, Example 10.6.2.

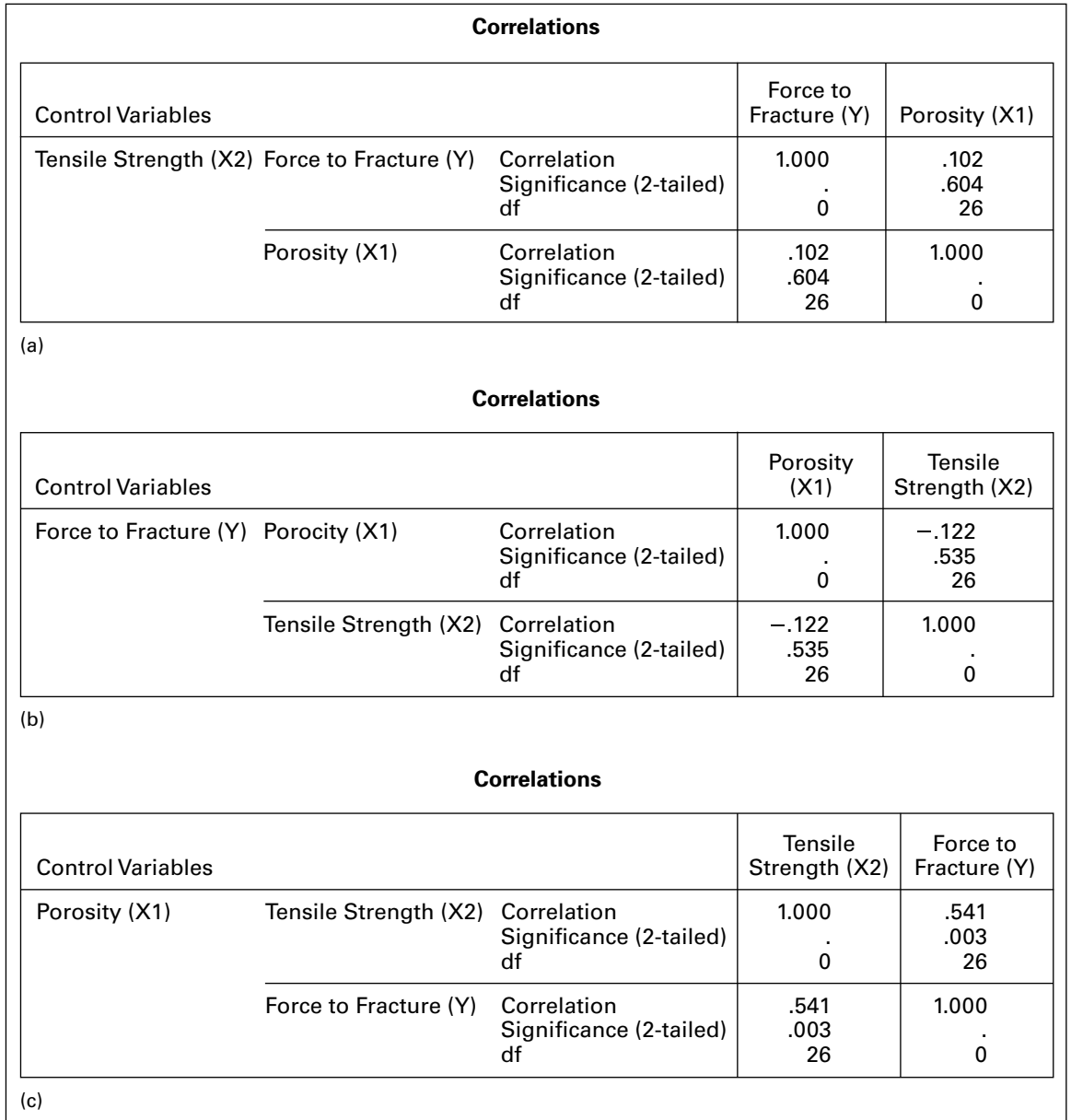


FIGURE 10.6.5 Partial correlation coefficients for the data in Example 10.6.1. (a) $r_{Y1.2}$, (b) $r_{12.Y}$ and (c) $r_{Y2.1}$.

Although our illustration of correlation analysis is limited to the three-variable case, the concepts and techniques extend logically to the case of four or more variables. The number and complexity of the calculations increase rapidly as the number of variables increases.

EXERCISES

- 10.6.1** The objective of a study by Anton et al. (A-5) was to investigate the correlation structure of multiple measures of HIV burden in blood and tissue samples. They measured HIV burden four ways. Two measurements were derived from blood samples, and two measurements were made on rectal tissue. The two blood measures were based on HIV DNA assays and a second co-culture assay that was a modification of the first measure. The third and fourth measurements were quantitations of HIV-1 DNA and RNA from rectal biopsy tissue. The table below gives data on HIV levels from these measurements for 34 subjects.

HIV DNA Blood (Y)	HIV Co-Culture Blood (X_1)	HIV DNA Rectal Tissue (X_2)	HIV RNA Rectal Tissue (X_3)
115	.38	899	56
86	1.65	167	158
19	.16	73	152
6	.08	146	35
23	.02	82	60
147	1.98	2483	1993
27	.15	404	30
140	.25	2438	72
345	.55	780	12
92	.22	517	5
85	.09	346	5
24	.17	82	12
109	.41	1285	5
5	.02	380	5
95	.84	628	32
46	.02	451	5
25	.64	159	5
187	.20	1335	121
5	.04	30	5
47	.02	13	30
118	.24	5	5
112	.72	625	83
79	.45	719	70
52	.23	309	167
52	.06	27	29
7	.37	199	5
13	.13	510	42
80	.24	271	15
86	.96	273	45
26	.29	534	71
53	.25	473	264
185	.28	2932	108
30	.19	658	33
9	.03	103	5
76	.21	2339	5

(Continued)

HIV DNA Blood (Y)	HIV Co-Culture Blood (X_1)	HIV DNA Rectal Tissue (X_2)	HIV RNA Rectal Tissue (X_3)
51	.09	31	36
73	.06	158	5
47	.08	773	5
48	.12	545	67
16	.03	5	5

Source: Data provided courtesy of Peter A. Anton, M.D.

- Compute the simple correlation coefficients between all possible pairs of variables.
- Compute the multiple correlation coefficient among the four variables. Test the overall correlation for significance.
- Calculate the partial correlations between HIV DNA blood and each one of the other variables while controlling for the other two. (These are called second-order partial correlation coefficients.)
- Calculate the partial correlation between HIV co-culture blood and HIV DNA, controlling for the other two variables.
- Calculate the partial correlation between HIV co-culture blood and HIV RNA, controlling for the other two variables.
- Calculate the partial correlations between HIV DNA and HIV RNA, controlling for the other two variables.

10.6.2 The following data were obtained on 12 males between the ages of 12 and 18 years (all measurements are in centimeters):

Height (Y)	Radius Length (X_1)	Femur Length (X_2)
149.0	21.00	42.50
152.0	21.79	43.70
155.7	22.40	44.75
159.0	23.00	46.00
163.3	23.70	47.00
166.0	24.30	47.90
169.0	24.92	48.95
172.0	25.50	49.90
174.5	25.80	50.30
176.1	26.01	50.90
176.5	26.15	50.85
179.0	26.30	51.10
Total	1992.1	290.87
		573.85

- Find the sample multiple correlation coefficient and test the null hypothesis that $\rho_{y,12} = 0$.
- Find each of the partial correlation coefficients and test each for significance. Let $\alpha = .05$ for all tests.
- Determine the p value for each test.
- State your conclusions.

10.6.3 The following data were collected on 15 obese girls:

Weight in Kilograms (Y)	Lean Body Weight (X_1)	Mean Daily Caloric Intake (X_2)	
79.2	54.3	2670	
64.0	44.3	820	
67.0	47.8	1210	
78.4	53.9	2678	
66.0	47.5	1205	
63.0	43.0	815	
65.9	47.1	1200	
63.1	44.0	1180	
73.2	44.1	1850	
66.5	48.3	1260	
61.9	43.5	1170	
72.5	43.3	1852	
101.1	66.4	1790	
66.2	47.5	1250	
99.9	66.1	1789	
Total	1087.9	741.1	22739

- (a) Find the multiple correlation coefficient and test it for significance.
 (b) Find each of the partial correlation coefficients and test each for significance. Let $\alpha = .05$ for all tests.
 (c) Determine the p value for each test.
 (d) State your conclusions.
- 10.6.4 A research project was conducted to study the relationships among intelligence, aphasia, and apraxia. The subjects were patients with focal left hemisphere damage. Scores on the following variables were obtained through the application of standard tests.

Y = intelligence

X_1 = ideomotor apraxia

X_2 = constructive apraxia

X_3 = lesion volume(pixels)

X_4 = severity of aphasia

The results are shown in the following table. Find the multiple correlation coefficient and test for significance. Let $\alpha = .05$ and find the p value.

Subject	Y	X_1	X_2	X_3	X_4
1	66	7.6	7.4	2296.87	2
2	78	13.2	11.9	2975.82	8
3	79	13.0	12.4	2839.38	11
4	84	14.2	13.3	3136.58	15

(Continued)

Subject	Y	X ₁	X ₂	X ₃	X ₄
5	77	11.4	11.2	2470.50	5
6	82	14.4	13.1	3136.58	9
7	82	13.3	12.8	2799.55	8
8	75	12.4	11.9	2565.50	6
9	81	10.7	11.5	2429.49	11
10	71	7.6	7.8	2369.37	6
11	77	11.2	10.8	2644.62	7
12	74	9.7	9.7	2647.45	9
13	77	10.2	10.0	2672.92	7
14	74	10.1	9.7	2640.25	8
15	68	6.1	7.2	1926.60	5

10.7 SUMMARY

In this chapter we examine how the concepts and techniques of simple linear regression and correlation analysis are extended to the multiple-variable case. The least-squares method of obtaining the regression equation is presented and illustrated. This chapter also is concerned with the calculation of descriptive measures, tests of significance, and the uses to be made of the multiple regression equation. In addition, the methods and concepts of correlation analysis, including partial correlation, are discussed.

When the assumptions underlying the methods of regression and correlation presented in this and the previous chapter are not met, the researcher must resort to alternative techniques such as those discussed in Chapter 13.

SUMMARY OF FORMULAS FOR CHAPTER 10

Formula Number	Name	Formula
10.2.1	Representation of the multiple linear regression equation	$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj} + \epsilon_j$
10.2.2	Representation of the multiple linear regression equation with two independent variables	$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \epsilon_j$

(Continued)

10.2.3	Random deviation of a point from a plane when there are two independent variables	$\epsilon_j = y_j - \beta_0 - \beta_1 x_{1j} - \beta_2 x_{2j}$
10.3.1	Sum-of-squared residuals	$\sum \epsilon_j^2 = \sum (y_j - \hat{y}_j)^2$
10.4.1	Sum-of-squares equation	$\sum (y_j - \bar{y})^2 = \sum (\hat{y}_j - \bar{y})^2 + \sum (y_j - \hat{y}_j)^2$ $SST = SSR + SSE$
10.4.2	Coefficient of multiple determination	$R_{y.12\dots k}^2 = \frac{\sum (\hat{y}_j - \bar{y})^2}{\sum (y_j - \bar{y})^2} = \frac{SSR}{SST}$
10.4.3	t statistic for testing hypotheses about β_i	$t = \frac{\hat{\beta}_i - \beta_{i0}}{s_{\hat{\beta}_i}}$
10.5.1	Estimation equation for multiple linear regression	$\hat{y}_j = \hat{\beta}_0 + \hat{\beta}_1 x_{1j} + \hat{\beta}_2 x_{2j} + \cdots + \hat{\beta}_k x_{kj}$
10.5.2	Confidence interval for the mean of Y for a given X	$\hat{y}_j \pm t_{(1-\alpha/2), n-k-1} s_{\hat{y}_j}$
10.5.3	Prediction interval for Y for a given X	$\hat{y}_j \pm t_{(1-\alpha/2), n-k-1} s'_{\hat{y}_j}$
10.6.1	Multiple correlation model	$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj} + \epsilon_j$
10.6.2	Multiple correlation coefficient	$R_{y.12\dots k} = \sqrt{R_{y.12\dots k}^2} = \sqrt{\frac{\sum (\hat{y}_j - \bar{y})^2}{\sum (y_j - \bar{y})^2}} = \sqrt{\frac{SSR}{SST}}$
10.6.3	F statistic for testing the multiple correlation coefficient	$F = \frac{R_{y.12\dots k}^2}{1 - R_{y.12\dots k}^2} \cdot \frac{n - k - 1}{k}$
10.6.4–10.6.6	Partial correlation between two variables (1 and 2) after controlling for a third (3)	$r_{12.3} = (r_{12} - r_{13}r_{23}) / \sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}$

10.6.7	<i>t</i> statistic for testing hypotheses about partial correlation coefficients	$t = r_{y1.2\dots k} \sqrt{\frac{n - k - 1}{1 - r_{y1.2\dots k}^2}}$
Symbol Key	<ul style="list-style-type: none"> • $\hat{\beta}_x$ = estimated regression/correlation coefficient <i>x</i> • β_x = regression/correlation coefficient • ϵ = model error term • <i>k</i> = number of independent variables • <i>n</i> = sample size • $r_{12.3}$ = sample partial correlation coefficient between 1 and 2 after controlling for 3 • <i>R</i> = sample correlation coefficient • R^2 = multiple coefficient of determination • <i>t</i> = <i>t</i> statistic • x_i = value of independent variable at <i>i</i> • \bar{x} = sample mean of independent variable • y_i = value of dependent variable at <i>i</i> • \bar{y} = sample mean of dependent variable • \hat{y} = estimated <i>y</i> • <i>z</i> = <i>z</i> statistic 	

REVIEW QUESTIONS AND EXERCISES

1. What are the assumptions underlying multiple regression analysis when one wishes to infer about the population from which the sample data have been drawn?
2. What are the assumptions underlying the correlation model when inference is an objective?
3. Explain fully the following terms:

(a) Coefficient of multiple determination	(b) Multiple correlation coefficient
(c) Simple correlation coefficient	(d) Partial correlation coefficient
4. Describe a situation in your particular area of interest where multiple regression analysis would be useful. Use real or realistic data and do a complete regression analysis.
5. Describe a situation in your particular area of interest where multiple correlation analysis would be useful. Use real or realistic data and do a complete correlation analysis.

In Exercises 6 through 11 carry out the indicated analysis and test hypotheses at the indicated significance levels. Compute the *p* value for each test.

6. We learned in Example 9.7.1 that the purpose of a study by Kwast-Rabben et al. (A-6) was to analyze somatosensory evoked potentials (SEPs) and their interrelations following stimulation of digits I, III, and V in the hand. Healthy volunteers were recruited for the study. Researchers applied stimulation below-pain-level intensity to the fingers. Recordings of spinal responses were made with electrodes fixed by adhesive electrode cream to the subject's skin. Results are shown in the following table for

114 subjects. Use multiple regression to see how well you can predict the peak spinal latency (Cv) of the SEP for digit I when age (years) and arm length (cm) are the predictor variables. Evaluate the usefulness of your prediction equation.

Age	Arm Length	Cv Dig.I	Age	Arm Length	Cv Dig.I	Age	Arm Length	Cv Dig.I
35.07	76.5	13.50	32.00	82.0	16.30	42.08	94.0	17.70
35.07	76.5	13.50	32.00	82.0	15.40	40.09	94.0	17.70
21.01	77.0	13.00	38.09	86.5	16.60	40.09	94.0	17.40
21.01	77.0	13.60	38.09	86.5	16.00	42.09	92.5	18.40
47.06	75.5	14.30	58.07	85.0	17.00	20.08	95.0	19.00
47.06	75.5	14.90	58.07	85.0	16.40	50.08	94.5	19.10
26.00	80.0	15.40	54.02	88.0	17.60	50.08	94.5	19.20
26.00	80.0	14.70	48.10	92.0	16.80	47.11	97.5	17.80
53.04	82.0	15.70	48.10	92.0	17.00	47.11	97.5	19.30
53.04	82.0	15.80	54.02	88.0	17.60	26.05	96.0	17.50
43.07	79.0	15.20	45.03	91.5	17.30	26.05	96.0	18.00
39.08	83.5	16.50	45.03	91.5	16.80	43.02	98.0	18.00
39.08	83.5	17.00	35.11	94.0	17.00	43.02	98.0	18.80
43.07	79.0	14.70	26.04	88.0	15.60	32.06	98.5	18.30
29.06	81.0	16.00	51.07	87.0	16.80	32.06	98.5	18.60
29.06	81.0	15.80	51.07	87.0	17.40	33.09	97.0	18.80
50.02	86.0	15.10	26.04	88.0	16.50	33.09	97.0	19.20
25.07	81.5	14.60	35.11	94.0	16.60	35.02	100.0	18.50
25.07	81.5	15.60	52.00	88.5	18.00	35.02	100.0	18.50
25.10	82.5	14.60	44.02	90.0	17.40	26.05	96.0	19.00
47.04	86.0	17.00	44.02	90.0	17.30	26.05	96.0	18.50
47.04	86.0	16.30	24.05	91.0	16.40	25.08	100.5	19.80
37.00	83.0	16.00	24.00	87.0	16.10	25.06	100.0	18.80
37.00	83.0	16.00	24.00	87.0	16.10	25.06	100.0	18.40
34.10	84.0	16.30	24.00	87.0	16.00	25.08	100.5	19.00
47.01	87.5	17.40	24.00	87.0	16.00	30.05	101.0	18.00
47.01	87.5	17.00	53.05	90.0	17.50	30.05	101.0	18.20
30.04	81.0	14.10	53.05	90.0	17.50	36.07	104.5	18.90
23.06	81.5	14.20	52.06	90.0	18.00	36.07	104.5	19.20
23.06	81.5	14.70	52.06	90.0	17.90	35.09	102.0	21.00
30.04	81.0	13.90	53.04	93.0	18.40	35.09	102.0	19.20
78.00	81.0	17.20	22.04	90.0	16.40	21.01	101.5	18.60
41.02	83.5	16.70	22.04	90.0	15.80	21.01	101.5	18.60
41.02	83.5	16.50	46.07	95.5	18.80	40.00	95.5	20.00
28.07	78.0	14.80	46.07	95.5	18.60	42.09	92.5	18.40
28.07	78.0	15.00	47.00	93.5	18.00	42.08	94.0	18.50
36.05	88.0	17.30	47.00	93.5	17.90	35.04	86.0	16.00
35.04	86.0	15.30	39.05	94.5	17.40	36.05	88.0	16.60

Source: Data provided courtesy of Olga Kwast-Rabben, Ph.D.

7. The following table shows the weight and total cholesterol and triglyceride levels in 15 patients with primary type II hyperlipoproteinemia just prior to initiation of treatment:

Y Weight (kg)	X_1 Total Cholesterol (mg/100 ml)	X_2 Triglyceride (mg/100 ml)
76	302	139
97	336	101
83	220	57
52	300	56
70	382	113
67	379	42
75	331	84
78	332	186
70	426	164
99	399	205
75	279	230
78	332	186
70	410	160
77	389	153
76	302	139

Compute the multiple correlation coefficient and test for significance at the .05 level.

8. In a study of the relationship between creatinine excretion, height, and weight, the data shown in the following table were collected on 20 infant males:

Infant	Creatinine Excretion (mg/day) Y	Weight (kg) X_1	Height (cm) X_2
1	100	9	72
2	115	10	76
3	52	6	59
4	85	8	68
5	135	10	60
6	58	5	58
7	90	8	70
8	60	7	65
9	45	4	54
10	125	11	83
11	86	7	64
12	80	7	66
13	65	6	61
14	95	8	66
15	25	5	57
16	125	11	81
17	40	5	59

(Continued)

Infant	Creatinine Excretion (mg/day)	Weight (kg)	Height (cm)
	Y	X_1	X_2
18	95	9	71
19	70	6	62
20	120	10	75

- (a) Find the multiple regression equation describing the relationship among these variables.
 (b) Compute R^2 and do an analysis of variance.
 (c) Let $X_1 = 10$ and $X_2 = 60$ and find the predicted value of Y .
9. A study was conducted to examine those variables thought to be related to the job satisfaction of nonprofessional hospital employees. A random sample of 15 employees gave the following results:

Score on Job Satisfaction Test (Y)	Coded Intelligence Score (X_1)	Index of Personal Adjustment (X_2)
54	15	8
37	13	1
30	15	1
48	15	7
37	10	4
37	14	2
31	8	3
49	12	7
43	1	9
12	3	1
30	15	1
37	14	2
61	14	10
31	9	1
31	4	5

- (a) Find the multiple regression equation describing the relationship among these variables.
 (b) Compute the coefficient of multiple determination and do an analysis of variance.
 (c) Let $X_1 = 10$ and $X_2 = 5$ and find the predicted value of Y .
10. A medical research team obtained the index of adiposity, basal insulin, and basal glucose values on 21 normal subjects. The results are shown in the following table. The researchers wished to investigate the strength of the association among these variables.

Index of Adiposity Y	Basal Insulin ($\mu\text{U/ml}$) X_1	Basal Glucose ($\text{mg}/100\text{ ml}$) X_2
90	12	98
112	10	103
127	14	101
137	11	102
103	10	90
140	38	108
105	9	100
92	6	101
92	8	92
96	6	91
114	9	95
108	9	95
160	41	117
91	7	101
115	9	86
167	40	106
108	9	84
156	43	117
167	17	99
165	40	104
168	22	85

Compute the multiple correlation coefficient and test for significance at the .05 level.

11. As part of a study to investigate the relationship between stress and certain other variables, the following data were collected on a simple random sample of 15 corporate executives.
- Find the least-squares regression equation for these data.
 - Construct the analysis of variance table and test the null hypothesis of no relationship among the five variables.
 - Test the null hypothesis that each slope in the regression model is equal to zero.
 - Find the multiple coefficient of determination and the multiple correlation coefficient. Let $\alpha = .05$ and find the p value for each test.

Measure of Stress (Y)	Measure of Firm Size (X_1)	Number of Years in Present Position (X_2)	Annual Salary ($\times 1000$) (X_3)	Age (X_4)
101	812	15	\$30	38
60	334	8	20	52
10	377	5	20	27
27	303	10	54	36

(Continued)

Measure of Stress (Y)	Measure of Firm Size (X_1)	Number of Years in Present Position (X_2)	Annual Salary ($\times 1000$) (X_3)	Age (X_4)
89	505	13	52	34
60	401	4	27	45
16	177	6	26	50
184	598	9	52	60
34	412	16	34	44
17	127	2	28	39
78	601	8	42	41
141	297	11	84	58
11	205	4	31	51
104	603	5	38	63
76	484	8	41	30

For each of the studies described in Exercises 12 through 16, answer as many of the following questions as possible:

- (a) Which is more relevant, regression analysis or correlation analysis, or are both techniques equally relevant?
 - (b) Which is the dependent variable?
 - (c) What are the independent variables?
 - (d) What are the appropriate null and alternative hypotheses?
 - (e) Which null hypotheses do you think were rejected? Why?
 - (f) Which is the more relevant objective, prediction or estimation, or are the two equally relevant? Explain your answer.
 - (g) What is the sampled population?
 - (h) What is the target population?
 - (i) Which variables are related to which other variables? Are the relationships direct or inverse?
 - (j) Write out the regression equation using appropriate numbers for parameter estimates.
 - (k) What is the numerical value of the coefficient of multiple determination?
 - (l) Give numerical values for any correlation coefficients that you can.
12. Hashimoto et al. (A-7) developed a multiple regression model to predict the number of visits to emergency rooms at Jikei University hospitals in Tokyo for children having an asthma attack. The researchers found that the number of visits per night increased significantly when climate conditions showed a rapid decrease from higher barometric pressure, from higher air temperature, and from higher humidity, as well as lower wind speed. The final model demonstrated that 22 percent of the variation in the number of visits was explained by variation in the predictor variables mentioned above with eight other significant climate variables.
 13. Correlation was one of many procedures discussed in a study reported by Stenvinkel et al. (A-8). In a cohort of 204 subjects with end-stage renal disease, they found no significant correlations between log plasma adiponectin levels and age and no significant correlation between log plasma adiponectin and glomerular filtration rate.

14. Van Schuylenbergh et al. (A-9) used physiological and anthropometric measurements as independent variables to predict triathlon performance (expressed in minutes). Ten triathletes underwent extensive physiological testing in swimming, cycling, and running. Within 2 weeks after the last laboratory test, all subjects competed in the National University Triathlon Championship. The final regression model was

$$TP = 130 - 9.2MLSSR - 25.9MLSSS + 1.4BLCR$$

in which TP = triathlon performance in minutes, MLSSR = the running speed at MLSS (m/s), MLSSS = the swimming speed at MLSS, and BLCR = blood lactate concentration at running MLSS (mmol/L). MLSS refers to maximal lactate steady state and is generally acknowledged to be a good marker of functional aerobic power during prolonged exercise. It also differs for each physical activity. For the above model $R^2 = .98$.

15. Maximal static inspiratory ($P_{I_{max}}$) mouth pressure is a simple measurement of respiratory muscle strength. A study by Tomalak et al. (A-10) examined correlations among the variables with $P_{I_{max}}$ (measured sitting), forced expiratory volume (FEV), peak expiratory flow (PEF), and maximal inspiratory flow (PIF) in 144 boys and 152 girls ages 7–14. The researchers found $P_{I_{max}}$ was correlated with FEV, PEF, and PIF in boys ($p = .001$, $p = .0055$, and $p = .002$, respectively) and for girls the correlations were also significant ($p < .001$, $p < .001$, and $p < .001$, respectively).
16. Di Monaco et al. (A-11) used multiple regression to predict bone mineral density of the femoral neck (among other locations). Among 124 Caucasian, healthy postmenopausal women, they found that weight ($p < .001$), age ($p < .01$), and total lymphocyte count ($p < .001$) were each useful in predicting bone mineral density. In addition, $R^2 = .40$.

For each of the data sets given in Exercises 17 through 19, do as many of the following as you think appropriate:

- Obtain the least-squares multiple regression equation.
 - Compute the sample coefficient of multiple determination.
 - Compute the sample coefficient of multiple correlation.
 - Compute simple coefficients of determination and correlation.
 - Compute partial correlation coefficients.
 - Construct graphs.
 - Formulate relevant hypotheses, perform the appropriate tests, and find p values.
 - State the statistical decisions and clinical conclusions that the results of your hypothesis tests justify.
 - Use your regression equations to make predictions and estimates about the dependent variable for your selected values of the independent variables.
 - Construct confidence intervals for relevant population parameters.
 - Describe the population(s) to which you think your inferences are applicable.
17. Pellegrino et al. (A-12) hypothesized that maximal bronchoconstriction can be predicted from the bronchomotor effect of deep inhalation and the degree of airway sensitivity to methacholine (MCh). One group of participants consisted of 26 healthy or mildly asthmatic subjects (22 males, 4 females) who had limited bronchoconstriction to inhaled MCh. The mean age of the patients was 31 years with a standard deviation of 8. There was one smoker in the group. Among the data collected on each subject were the following observations on various lung function measurement variables:

(X ₁) FEV ₁	(X ₂) FEV ₁ , % Pred	(X ₃) FEV ₁ / FVC, %	(X ₄) V̇m ₅₀	(X ₅) V̇p ₅₀	(X ₆) M/P Ratio	(X ₇) MP Slope	(X ₈) PD ₁₅ FEV ₁ (In mg)	(X ₉) PD ₄₀ V̇m ₅₀ (In mg)	(X ₁₀) PD ₄₀ V̇p ₅₀ (In mg)	(X ₁₁) FEV ₁ Max decr (%)	(X ₁₂) V̇m ₅₀ Max decr(%)	(X ₁₃) V̇p ₅₀ Max decr(%)
5.22	108.75	83.92	5.30	3.90	1.36	0.75	8.44	8.24	6.34	21.40	55.40	74.40
5.38	123.96	78.54	6.00	3.70	1.62	0.56	7.76	7.00	6.18	15.80	50.80	85.14
3.62	111.04	86.19	3.10	2.85	1.10	0.69	6.92	6.61	5.56	30.40	54.36	83.07
3.94	94.26	85.28	4.10	2.70	1.52	0.44	6.79	8.52	6.38	16.40	29.10	58.50
4.48	104.43	76.58	3.21	3.00	1.07	0.63	8.79	9.74	6.68	27.80	46.30	76.70
5.28	117.33	81.99	5.65	5.55	1.02	0.83	8.98	8.97	8.19	32.60	70.80	90.00
3.80	93.37	76.61	3.75	4.70	0.80	0.50	10.52	10.60	10.04	15.80	35.30	64.90
3.14	104.67	82.63	3.20	3.20	1.00	0.70	6.18	6.58	6.02	37.60	64.10	87.50
5.26	120.09	84.84	6.30	7.40	0.89	0.55	11.85	11.85	11.85	11.70	29.10	41.20
4.87	121.14	89.69	5.50	5.50	1.00	0.56	11.85	11.85	11.85	10.30	16.40	29.70
5.35	124.71	84.65	5.60	7.00	0.80	0.40	11.98	11.98	11.29	0.00	18.00	47.20
4.30	95.98	80.37	5.78	4.90	1.18	0.59	6.48	6.19	5.11	17.00	48.20	79.60
3.75	87.82	65.79	2.26	1.65	1.37	0.53	6.25	7.02	5.03	27.10	39.53	81.80
4.41	112.21	69.78	3.19	2.95	1.08	0.57	7.66	8.08	5.51	24.70	48.80	85.90
4.66	108.37	78.72	5.00	5.90	0.85	0.49	7.79	9.77	6.10	15.00	35.00	70.30
5.19	99.05	73.62	4.20	1.50	2.80	0.63	5.15	5.78	4.72	31.40	61.90	86.70
4.32	122.38	75.13	4.39	3.30	1.33	0.74	6.20	6.34	5.10	28.25	60.30	78.00
4.05	95.97	84.38	3.40	2.50	1.30	0.59	5.64	8.52	5.61	18.20	29.50	46.00
3.23	88.25	87.30	4.00	4.00	1.00	0.71	3.47	3.43	2.77	21.60	64.50	86.00
3.99	105.56	86.74	5.30	2.70	1.96	0.76	6.40	5.20	6.17	22.50	63.00	77.80
4.37	102.34	80.18	3.20	1.80	1.77	0.85	5.05	4.97	5.42	35.30	57.00	78.00
2.67	68.11	65.12	1.70	1.30	1.38	0.91	3.97	3.95	4.11	32.40	58.80	82.40
4.75	103.71	73.08	4.60	3.60	1.21	0.71	6.34	5.29	6.04	18.85	47.50	72.20
3.19	88.12	85.07	3.20	1.80	1.77	0.76	5.08	4.85	5.16	36.20	83.40	93.00
3.29	102.17	92.68	3.80	2.40	1.58	0.50	8.21	6.90	10.60	21.60	28.10	66.70
2.87	95.03	95.67	3.00	3.00	1.00	0.75	6.24	5.99	7.50	27.00	46.70	68.30

$\dot{V}_{m_{50}}$ and $\dot{V}_{p_{50}}$ = maximal and partial forced expiratory flows at 50 percent of control FVC; M/P ratio = ratio of $\dot{V}_{m_{50}}$ to $\dot{V}_{p_{50}}$ at control; MP slope = slope of the regression of percent decrements of $\dot{V}_{m_{50}}$ and $\dot{V}_{p_{50}}$ recorded during the MCh inhalation challenge; PD₁₅FEV₁ = dose of MCh that decreased FEV₁ by 15 percent of control; PD₄₀ $\dot{V}_{m_{50}}$ and PD₄₀ $\dot{V}_{p_{50}}$ = doses of MCh that decreased $\dot{V}_{m_{50}}$ and $\dot{V}_{p_{50}}$ by 40 percent of control respectively; % max decr = percent maximal decrement at plateau. Source: Data provided courtesy of Dr. Riccardo Pellegrino.

- The purpose of a study by O'Brien et al. (A-13) was to assess hypothalamic-pituitary-adrenal (HPA) axis function (known to be altered in depression) in patients with Alzheimer's disease (AD) by means of the adrenocorticotrophic hormone (ACTH) test, which assesses adrenal function by measuring cortisol production by the adrenal gland in response to an injection of ACTH. AD subjects (mean age 69.9 years with standard deviation of 9.8) were recruited from referrals to a hospital memory clinic. Normal control subjects consisted of spouses of patients and residents of a retirement hostel (mean age 73.8 with standard deviation of 11.6). There were eight males and eight females in the AD group and 10 males and eight females in the control group. Among the data collected were the following observations on age (C1), age at onset for AD subjects (C2), length of history of disease in months (C3), cognitive examination score (C4), peak cortisol level (C5), and total hormone response (C6):

Alzheimer's Disease Subjects						Controls					
C1	C2	C3	C4	C5	C6	C1	C2	C3	C4	C5	C6
73	69	48	75	400.00	44610	70	•	•	97	419.00	53175
87	83	48	39	565.00	63855	81	•	•	93	470.00	54285
60	54	72	67	307.00	31110	82	•	•	93	417.00	47160
62	57	60	64	335.00	36000	57	•	•	101	215.00	27120
75	70	48	51	352.00	44760	87	•	•	91	244.00	23895
63	60	24	79	426.00	47250	88	•	•	88	355.00	33565
81	77	48	51	413.00	51825	87	•	•	91	392.00	42810
66	64	24	61	402.00	41745	70	•	•	100	354.00	45105
78	73	60	32	518.00	66030	63	•	•	103	457.00	48765
72	64	72	61	505.00	49905	87	•	•	81	323.00	39360
69	65	48	73	427.00	55350	73	•	•	94	386.00	48150
76	73	36	63	409.00	51960	87	•	•	91	244.00	25830
46	41	60	73	333.00	33030	58	•	•	103	353.00	42060
77	75	18	63	591.00	73125	85	•	•	93	335.00	37425
64	61	36	59	559.00	60750	58	•	•	99	470.00	55140
72	69	30	47	511.00	54945	67	•	•	100	346.00	50745
						68	•	•	100	262.00	28440
						62	•	•	93	271.00	23595

• = Not applicable.

Source: Data provided courtesy of Dr. John T. O'Brien.

19. Johnson et al. (A-14) note that the ability to identify the source of remembered information is a fundamental cognitive function. They conducted an experiment to explore the relative contribution of perceptual cues and cognitive operations information to age-related deficits in discriminating memories from different external sources (external source monitoring). Subjects for the experiment included 96 graduate and undergraduate students (41 males and 55 females) ranging in ages from 18 to 27 years. Among the data collected were the following performance recognition scores on source monitoring conditions (C1, C2, C3) and scores on the Benton Facial Recognition Test (C4), the Wechsler Adult Intelligence Scale—Revised (WAIS-R), WAIS-R Block Design subscale (C5), WAIS-R Vocabulary subscale (C6), the Benton Verbal Fluency Test (C7), and the Wisconsin Card Sorting Test (C8):

C1	C2	C3	C4	C5	C6	C7	C8
0.783	2.63	0.808	25	38	62	67	6
0.909	3.36	0.846	*	*	50	*	*
0.920	2.14	0.616	23	25	53	47	6
0.727	3.36	0.846	25	40	49	58	6
0.737	2.93	0.731	*	*	59	*	*
0.600	4.07	0.962	19	50	51	35	6
0.840	3.15	0.885	*	*	57	*	*
0.850	3.06	0.769	*	*	55	*	*
0.875	3.72	0.923	24	23	52	35	6

(Continued)

C1	C2	C3	C4	C5	C6	C7	C8
0.792	3.15	0.884	*	*	50	*	*
0.680	4.07	0.962	*	*	56	*	*
0.731	4.64	1.000	23	30	59	47	3
0.826	1.84	0.616	*	*	52	*	*
0.609	2.98	0.846	*	*	56	*	*
0.923	4.64	1.000	*	*	53	*	*
0.773	3.36	0.846	*	*	60	*	*
0.714	1.62	0.577	23	43	53	42	6
0.667	3.72	0.923	20	32	59	28	6
0.769	1.40	0.423	*	*	51	*	*
0.565	3.55	0.885	*	*	45	*	*
0.824	1.78	0.577	*	*	45	*	*
0.458	1.90	0.615	21	46	50	47	6
0.840	4.07	0.962	*	*	59	*	*
0.720	4.07	0.962	*	*	53	*	*
0.917	3.72	0.923	24	31	43	37	6
0.560	4.07	0.926	*	*	62	*	*
0.840	4.07	0.962	26	22	50	40	6
0.720	4.07	0.962	*	*	52	*	*
0.783	1.74	0.577	*	*	54	*	*
0.696	1.62	0.539	*	*	57	*	*
0.625	3.72	0.923	22	37	55	40	6
0.737	1.12	0.423	*	*	47	*	*
0.900	1.92	0.654	22	40	46	42	6
0.565	3.55	0.885	22	43	56	64	6
0.680	4.07	0.962	*	*	54	*	*
0.760	4.07	0.962	*	*	58	*	*
0.958	1.90	0.615	24	36	46	43	6
0.652	2.98	0.846	*	*	54	*	*
0.560	4.07	0.962	*	*	56	*	*
0.500	1.92	0.654	24	42	45	46	6
0.826	2.63	0.808	*	*	60	*	*
0.783	2.58	0.808	*	*	60	*	*
0.783	2.63	0.808	*	*	49	*	*
0.750	2.14	0.692	22	37	62	58	6
0.913	2.11	0.693	*	*	46	*	*
0.952	1.49	0.539	26	32	48	36	6
0.800	4.07	0.962	*	*	59	*	*
0.870	3.55	0.885	*	*	48	*	*
0.652	1.97	0.654	*	*	59	*	*
0.640	4.07	0.962	25	36	56	54	6
0.692	4.64	1.000	23	23	58	25	6
0.917	3.72	0.923	*	*	55	*	*
0.760	4.07	0.962	22	35	52	33	6
0.739	3.55	0.885	24	43	58	43	6

(Continued)

C1	C2	C3	C4	C5	C6	C7	C8
0.857	3.20	0.808	*	*	59	*	*
0.727	3.36	0.846	*	*	61	*	*
0.833	2.80	0.846	*	*	56	*	*
0.840	4.07	0.962	21	11	49	58	3
0.478	2.27	0.731	*	*	60	*	*
0.920	4.07	0.962	24	40	64	50	6
0.731	4.64	1.000	20	40	51	50	6
0.920	4.07	0.962	23	50	61	53	6
0.720	4.07	0.962	*	*	57	*	*
1.000	2.79	0.807	25	47	56	30	6
0.708	3.72	0.923	24	16	57	42	6
1.000	4.64	1.000	25	48	55	54	6
0.739	3.55	0.885	23	27	57	38	6
0.600	4.20	0.962	22	38	57	33	6
0.962	4.64	1.000	25	37	63	31	6
0.772	2.22	0.731	24	48	51	41	6
0.800	2.92	0.847	24	28	47	45	6
0.923	4.64	1.000	25	45	54	48	6
0.870	3.50	0.885	24	44	54	48	5
0.808	4.64	1.000	24	43	57	58	6
1.000	4.07	0.962	25	30	59	49	6
0.870	3.55	0.885	26	44	61	35	6
0.923	4.64	1.000	*	*	52	*	*
0.958	2.58	0.808	27	32	52	33	6
0.826	3.50	0.885	21	31	61	44	6
0.962	3.72	0.923	23	31	57	38	6
0.783	3.50	0.885	23	46	60	36	6
0.905	3.20	0.808	23	34	55	37	4
1.000	4.64	1.000	23	33	57	33	6
0.875	3.72	0.923	21	34	55	29	6
0.885	4.07	0.962	*	*	52	*	*
0.913	2.92	0.846	23	44	57	47	6
0.962	4.07	0.961	24	36	54	43	6
0.682	3.36	0.846	20	41	61	34	1
0.810	2.63	0.769	20	40	57	43	6
0.720	2.79	0.808	25	23	64	43	3
0.875	2.80	0.846	24	43	59	43	2
0.923	3.72	0.924	25	40	58	33	6
0.909	3.36	0.846	24	43	56	41	6
0.920	4.07	0.962	24	50	52	28	6
1.000	3.72	0.923	21	45	64	46	6
0.609	3.50	0.885	22	25	49	35	6

* = Missing data.

Source: Data provided courtesy of Dr. Doreen M. De Leonardis.

Exercises for Use with the Large Data Sets Available on the Following Website:
www.wiley.com/college/daniel

1. Winters et al. (A-15) conducted a study involving 248 high-school students enrolled in introductory physical education courses. The researchers wanted to know if social cognitive theory constructs were correlated with discretionary, “leisure-time” physical exercise. The main outcome variable is STREN, which is the number of days in a week that a high-school student engaged in strenuous physical activity (operationally defined as exercise that results in sweating, labored breathing, and rapid heart rate). Students in the study filled out lengthy questionnaires from which the following variables were derived:

SELFR100—measures personal regulation of goal-directed behavior (higher values indicate more goal oriented).

SS100—measures social support, social encouragement, and social expectation that are provided by friends and family for physical exercise (higher values indicate more support).

SSE100—measures perceived ability to overcome barriers to exercise (higher values indicate higher ability).

OEVNORM—measures outcome expectations and their associated expectancies for physical exercise (higher values indicate stronger perceived links to desired outcomes from exercise).

With these data (LTEXER),

- (a) Calculate the bivariate correlation for each pair of variables and interpret the meaning of each.
- (b) Using STREN as the dependent variable, compute the multiple correlation coefficient.
- (c) Using STREN as the dependent variable, calculate the partial correlation coefficient for STREN and SELFR100 after controlling for SS100.
- (d) Using STREN as the dependent variable, calculate the partial correlation coefficient for STREN and SSE100 after controlling for OEVNORM.

Note that there many missing values in this data set.

2. With data obtained from a national database on childbirth, Matulavich et al. (A-16) examined the number of courses of prescribed steroids a mother took during pregnancy (STERIODS). The size of the baby was measured by length (cm), weight (grams), and head circumference (cm). Calculate the correlation of the number of courses of steroids with each of the three outcome variables. What are the hypotheses for your tests? What are the p -values? What are your conclusions? (The name of the data set is STERLENGTH.)
3. Refer to the data on cardiovascular risk factors (RISKFACT). The subjects are 1000 males engaged in sedentary occupations. You wish to study the relationships among risk factors in this population. The variables are

Y = oxygen consumption

X_1 = systolic blood pressure (mm Hg)

X_2 = total cholesterol (mg/dl)

X_3 = HDL cholesterol (mg/dl)

X_4 = triglycerides (mg/dl)

Select a simple random sample from this population and carry out an appropriate statistical analysis. Prepare a narrative report of your findings and compare them with those of your classmates. Consult with your instructor regarding the size of the sample.

4. Refer to the data on 500 patients who have sought treatment for the relief of respiratory disease symptoms (RESPDIS). A medical research team is conducting a study to determine what factors may be related to respiratory disease. The dependent variable Y is a measure of the severity of the disease. A larger value indicates a more serious condition. The independent variables are as follows:

X_1 = education (highest grade completed)

X_2 = measure of crowding of living quarters

X_3 = measure of air quality at place of residence (a larger number indicates poorer quality)

X_4 = nutritional status (a large number indicates a higher level of nutrition)

X_5 = smoking status (0 = smoker, 1 = nonsmoker)

Select a simple random sample of subjects from this population and conduct a statistical analysis that you think would be of value to the research team. Prepare a narrative report of your results and conclusions. Use graphic illustrations where appropriate. Compare your results with those of your classmates. Consult your instructor regarding the size of sample you should select.

REFERENCES

Methodology References

1. GEORGE W. SNEDECOR and WILLIAM G. COCHRAN, *Statistical Methods*, Sixth Edition, Iowa State University Press, Ames, 1967.
2. ROBERT G. D. STEEL and JAMES H. TORRIE, *Principles and Procedures of Statistics*, McGraw-Hill, New York, 1960.
3. R. L. ANDERSON and T. A. BANCROFT, *Statistical Theory in Research*, McGraw-Hill, New York, 1952.
4. M. H. KUTNER, C. J. NACHTSHEIM, J. NETER, and W. LI, *Applied Linear Statistical Models*, Fifth Edition, McGraw-Hill/Irwin, New York, 2005.

Applications References

- A-1. DEBRA A. JANSEN and MARY L. KELLER, "Cognitive Function in Community-Dwelling Elderly Women," *Journal of Gerontological Nursing*, 29 (2003), 34–43.
- A-2. M. NAEIJE, "Local Kinematic and Anthropometric Factors Related to the Maximum Mouth Opening in Healthy Individuals," *Journal of Oral Rehabilitation*, 29 (2002), 534–539.
- A-3. GWI-RYUNG SON, MAY L. WYKLE, and JACLENE A. ZAUSZNIIEWSKI, "Korean Adult Child Caregivers of Older Adults with Dementia," *Journal of Gerontological Nursing*, 29 (2003), 19–28.
- A-4. X. WANG, X. SHEN, X. LI, and C. MAULI AGRAWAL, "Age-Related Changes in the Collagen Network and Toughness of Bone," *Bone*, 31 (2002), 1–7.
- A-5. PETER A. ANTON, RONALD T. MITSUYASU, STEVEN G. DEEKS, DAVID T. SCADDEN, BRIDGET WAGNER, CHRISTINE HUANG, CATHERINE MACKEN, DOUGLAS D. RICHMAN, CINDY CHRISTOPHERSON, FLAVIA BORELLINI, RICHARD LAZAR, and KRISTEN M. HEGE, "Multiple Measures of HIV Burden in Blood and Tissue Are Correlated with Each Other but Not with Clinical Parameter in Aviremic Subjects," *AIDS*, 17 (2003), 53–63.
- A-6. OLGA KWAST-RABBEN, ROLF LIBELIUS, and HANNU HEIKKILÄ, "Somatosensory Evoked Potentials Following Stimulation of Digital Nerves," *Muscle and Nerve*, 26 (2002), 533–538.
- A-7. MITSUO HASHIMOTO, TAIKI FUKUDA, TETSUYA SHIMIZU, SHOU WATANABE, SATOSHI WATANUKI, YOSHIKATSU ETO, and MITSUYOSHI URASHIMA, "Influence of Climate Factors on Emergency Visits for Childhood Asthma Attack," *Pediatrics International*, 46 (2004), 48–52.
- A-8. PETER STENVINKEL, ALICIA MARCHELEWSKA, ROBERTO PECOITS-FILHO, OLOF HEIMBÜRGER, ZHENGZHONG ZHANG, CATHERINE HOFF, CLIFF HOLMES, JOANAS AXELSSON, SIVONNE ARVIDSSON, MARTIN SCHALLING, PETER BARANY, BENGT LINKHOLM, and LOUISE NORDFORS, "Adiponectin in Renal Disease: Relationship to Phenotype and Genetic Variation in the Genetic Encoding Adiponectin," *Kidney International*, 65 (2004), 274–281.

- A-9. R. VAN SCHUYLENBERGH, B. VANDEN EYNDE, and P. HESPEL, "Prediction of Sprint Triathlon Performance from Laboratory Tests," *European Journal of Applied Physiology*, 91 (2004), 94–99.
- A-10. WALDEMAR TOMALAK, ANDRZEJ POGORZELSKI, and JAROSLAW PRUSAK, "Normal Values for Maximal Static Inspiratory and Expiratory Pressures in Healthy Children," *Pediatric Pulmonology*, 34 (2002), 42–46.
- A-11. MARCO DI MONACO, FULVIA VALLERO, ROBERTO DI MONACO, FULVIO MAUTINO, and ALBERTO CAVANNA, "Total Lymphocyte Count and Femoral Bone Mineral Density in Postmenopausal Women," *Journal of Bone Mineral Metabolism*, 22 (2004), 58–63.
- A-12. RICCARDO PELLEGRINO, BENEDETTO VIOLANTE, and VITO BRUSASCO, "Maximal Bronchoconstriction in Humans: Relationship to Deep Inhalation and Airway Sensitivity," *American Journal of Respiratory and Critical Care Medicine*, 153 (1996), 115–121.
- A-13. J. T. O'BRIEN, D. AMES, I. SCHWEITZER, M. MASTWYK, and P. COLMAN, "Enhanced Adrenal Sensitivity to Adrenocorticotrophic Hormone (ACTH) Is Evidence of HPA Axis Hyperactivity in Alzheimer's Disease," *Psychological Medicine*, 26 (1996), 7–14.
- A-14. MARCIA K. JOHNSON, DOREEN M. DE LEONARDIS, SHAHIN HARSHTRUDI, and SUSAN A. FERGUSON, "Aging and Single Versus Multiple Cues in Source Monitoring," *Psychology and Aging*, 10 (1995), 507–517.
- A-15. ERIC R. WINTERS, RICK L. PETOSA, and THOMAS E. CHARLETON, "Using Social Cognitive Theory to Explain Discretionary, 'Leisure-Time' Physical Exercise Among High School Students," *Journal of Adolescent Health*, 32 (2003), 436–442.
- A-16. ANGELA MATULAVICH, DONNA MILES-CURRY, BARBARA WARNER, BOBBE GRAY, and the National Institute of Child and Health Development. Data analyzed at the Wright State University Statistical Consulting Center.