Probability & Statistics for Engineers & Scientists

NINTH EDITION

Ronald E. Walpole
Roanoke College

Raymond H. Myers
Virginia Tech

Sharon L. Myers
Radford University

Keying Ye
University of Texas at San Antonio

Prentice Hall

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Chapter 9

One- and Two-Sample Estimation Problems

9.1 Introduction

In previous chapters, we emphasized sampling properties of the sample mean and variance. We also emphasized displays of data in various forms. The purpose of these presentations is to build a foundation that allows us to draw conclusions about the population parameters from experimental data. For example, the Central Limit Theorem provides information about the distribution of the sample mean $\bar{X}$. The distribution involves the population mean μ. Thus, any conclusions concerning μ drawn from an observed sample average must depend on knowledge of this sampling distribution. Similar comments apply to S^2 and σ^2. Clearly, any conclusions we draw about the variance of a normal distribution will likely involve the sampling distribution of S^2.

In this chapter, we begin by formally outlining the purpose of statistical inference. We follow this by discussing the problem of **estimation of population parameters**. We confine our formal developments of specific estimation procedures to problems involving one and two samples.

9.2 Statistical Inference

In Chapter 1, we discussed the general philosophy of formal statistical inference. **Statistical inference** consists of those methods by which one makes inferences or generalizations about a population. The trend today is to distinguish between the **classical method** of estimating a population parameter, whereby inferences are based strictly on information obtained from a random sample selected from the population, and the **Bayesian method**, which utilizes prior subjective knowledge about the probability distribution of the unknown parameters in conjunction with the information provided by the sample data. Throughout most of this chapter, we shall use classical methods to estimate unknown population parameters such as the mean, the proportion, and the variance by computing statistics from random samples and applying the theory of sampling distributions, much of which was covered in Chapter 8. Bayesian estimation will be discussed in Chapter 18.

Statistical inference may be divided into two major areas: **estimation** and **tests of hypotheses**. We treat these two areas separately, dealing with theory and applications of estimation in this chapter and hypothesis testing in Chapter 10. To distinguish clearly between the two areas, consider the following examples. A candidate for public office may wish to estimate the true proportion of voters favoring him by obtaining opinions from a random sample of 100 eligible voters. The fraction of voters in the sample favoring the candidate could be used as an estimate of the true proportion in the population of voters. A knowledge of the sampling distribution of a proportion enables one to establish the degree of accuracy of such an estimate. This problem falls in the area of estimation.

Now consider the case in which one is interested in finding out whether brand A floor wax is more scuff-resistant than brand B floor wax. He or she might hypothesize that brand A is better than brand B and, after proper testing, accept or reject this hypothesis. In this example, we do not attempt to estimate a parameter, but instead we try to arrive at a correct decision about a prestated hypothesis. Once again we are dependent on sampling theory and the use of data to provide us with some measure of accuracy for our decision.

9.3 Classical Methods of Estimation

A **point estimate** of some population parameter θ is a single value $\hat{\theta}$ of a statistic $\hat{\Theta}$. For example, the value $\bar{x}$ of the statistic $\bar{X}$, computed from a sample of size n, is a point estimate of the population parameter μ. Similarly, $\hat{p} = x/n$ is a point estimate of the true proportion p for a binomial experiment.

An estimator is not expected to estimate the population parameter without error. We do not expect $\bar{X}$ to estimate μ exactly, but we certainly hope that it is not far off. For a particular sample, it is possible to obtain a closer estimate of μ by using the sample median $\tilde{X}$ as an estimator. Consider, for instance, a sample consisting of the values 2, 5, and 11 from a population whose mean is 4 but is supposedly unknown. We would estimate μ to be $\bar{x} = 6$, using the sample mean as our estimate, or $\tilde{x} = 5$, using the sample median as our estimate. In this case, the estimator $\tilde{X}$ produces an estimate closer to the true parameter than does the estimator $\bar{X}$. On the other hand, if our random sample contains the values 2, 6, and 7, then $\bar{x} = 5$ and $\tilde{x} = 6$, so $\bar{X}$ is the better estimator. Not knowing the true value of μ, we must decide in advance whether to use $\bar{X}$ or $\tilde{X}$ as our estimator.

Unbiased Estimator

What are the desirable properties of a "good" decision function that would influence us to choose one estimator rather than another? Let $\hat{\Theta}$ be an estimator whose value $\hat{\theta}$ is a point estimate of some unknown population parameter θ. Certainly, we would like the sampling distribution of $\hat{\Theta}$ to have a mean equal to the parameter estimated. An estimator possessing this property is said to be **unbiased**.

Definition 9.1: A statistic $\hat{\Theta}$ is said to be an **unbiased estimator** of the parameter θ if

$$\mu_{\hat{\Theta}} = E(\hat{\Theta}) = \theta.$$

Example 9.1: Show that S^2 is an unbiased estimator of the parameter σ^2.

Solution: In Section 8.5 on page 244, we showed that

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2.$$

Now

$$E(S^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$$
$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}E(X_i - \mu)^2 - nE(\bar{X} - \mu)^2\right] = \frac{1}{n-1}\left(\sum_{i=1}^{n}\sigma_{X_i}^2 - n\sigma_{\bar{X}}^2\right).$$

However,

$$\sigma_{X_i}^2 = \sigma^2, \text{ for } i = 1, 2, \ldots, n, \text{ and } \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}.$$

Therefore,

$$E(S^2) = \frac{1}{n-1}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) = \sigma^2.$$

Although S^2 is an unbiased estimator of σ^2, S, on the other hand, is usually a biased estimator of σ, with the bias becoming insignificant for large samples. This example illustrates **why we divide by $n-1$** rather than n when the variance is estimated.

Variance of a Point Estimator

If $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are two unbiased estimators of the same population parameter θ, we want to choose the estimator whose sampling distribution has the smaller variance. Hence, if $\sigma_{\hat{\theta}_1}^2 < \sigma_{\hat{\theta}_2}^2$, we say that $\hat{\Theta}_1$ is a **more efficient estimator** of θ than $\hat{\Theta}_2$.

Definition 9.2: If we consider all possible unbiased estimators of some parameter θ, the one with the smallest variance is called the **most efficient estimator** of θ.

Figure 9.1 illustrates the sampling distributions of three different estimators, $\hat{\Theta}_1$, $\hat{\Theta}_2$, and $\hat{\Theta}_3$, all estimating θ. It is clear that only $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased, since their distributions are centered at θ. The estimator $\hat{\Theta}_1$ has a smaller variance than $\hat{\Theta}_2$ and is therefore more efficient. Hence, our choice for an estimator of θ, among the three considered, would be $\hat{\Theta}_1$.

For normal populations, one can show that both $\bar{X}$ and $\tilde{X}$ are unbiased estimators of the population mean μ, but the variance of $\bar{X}$ is smaller than the variance

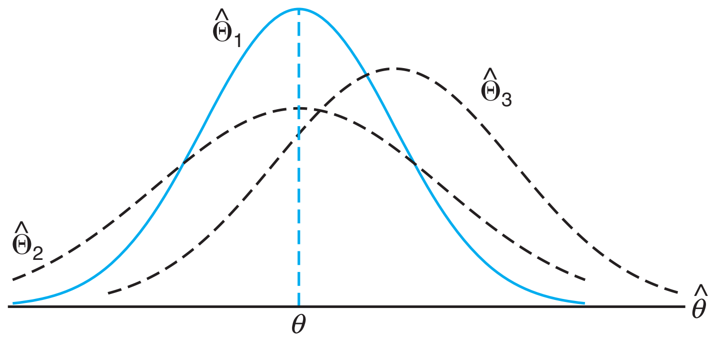

Figure 9.1: Sampling distributions of different estimators of θ.

of $\tilde{X}$. Thus, both estimates $\bar{x}$ and $\tilde{x}$ will, on average, equal the population mean μ, but $\bar{x}$ is likely to be closer to μ for a given sample, and thus $\bar{X}$ is more efficient than $\tilde{X}$.

Interval Estimation

Even the most efficient unbiased estimator is unlikely to estimate the population parameter exactly. It is true that estimation accuracy increases with large samples, but there is still no reason we should expect a **point estimate** from a given sample to be exactly equal to the population parameter it is supposed to estimate. There are many situations in which it is preferable to determine an interval within which we would expect to find the value of the parameter. Such an interval is called an **interval estimate**.

An interval estimate of a population parameter θ is an interval of the form $\hat{\theta}_L < \theta < \hat{\theta}_U$, where $\hat{\theta}_L$ and $\hat{\theta}_U$ depend on the value of the statistic $\hat{\Theta}$ for a particular sample and also on the sampling distribution of $\hat{\Theta}$. For example, a random sample of SAT verbal scores for students in the entering freshman class might produce an interval from 530 to 550, within which we expect to find the true average of all SAT verbal scores for the freshman class. The values of the endpoints, 530 and 550, will depend on the computed sample mean $\bar{x}$ and the sampling distribution of $\bar{X}$. As the sample size increases, we know that $\sigma^2_{\bar{X}} = \sigma^2/n$ decreases, and consequently our estimate is likely to be closer to the parameter μ, resulting in a shorter interval. Thus, the interval estimate indicates, by its length, the accuracy of the point estimate. An engineer will gain some insight into the population proportion defective by taking a sample and computing the *sample proportion defective*. But an interval estimate might be more informative.

Interpretation of Interval Estimates

Since different samples will generally yield different values of $\hat{\Theta}$ and, therefore, different values for $\hat{\theta}_L$ and $\hat{\theta}_U$, these endpoints of the interval are values of corresponding random variables $\hat{\Theta}_L$ and $\hat{\Theta}_U$. From the sampling distribution of $\hat{\Theta}$ we shall be able to determine $\hat{\Theta}_L$ and $\hat{\Theta}_U$ such that $P(\hat{\Theta}_L < \theta < \hat{\Theta}_U)$ is equal to any positive fractional value we care to specify. If, for instance, we find $\hat{\Theta}_L$ and $\hat{\Theta}_U$ such that

$$P(\hat{\Theta}_L < \theta < \hat{\Theta}_U) = 1 - \alpha,$$

for $0 < \alpha < 1$, then we have a probability of $1-\alpha$ of selecting a random sample that will produce an interval containing θ. The interval $\hat{\theta}_L < \theta < \hat{\theta}_U$, computed from the selected sample, is called a $100(1-\alpha)\%$ **confidence interval**, the fraction $1-\alpha$ is called the **confidence coefficient** or the **degree of confidence**, and the endpoints, $\hat{\theta}_L$ and $\hat{\theta}_U$, are called the lower and upper **confidence limits**. Thus, when $\alpha = 0.05$, we have a 95% confidence interval, and when $\alpha = 0.01$, we obtain a wider 99% confidence interval. The wider the confidence interval is, the more confident we can be that the interval contains the unknown parameter. Of course, it is better to be 95% confident that the average life of a certain television transistor is between 6 and 7 years than to be 99% confident that it is between 3 and 10 years. Ideally, we prefer a short interval with a high degree of confidence. Sometimes, restrictions on the size of our sample prevent us from achieving short intervals without sacrificing some degree of confidence.

In the sections that follow, we pursue the notions of point and interval estimation, with each section presenting a different special case. The reader should notice that while point and interval estimation represent different approaches to gaining information regarding a parameter, they are related in the sense that confidence interval estimators are based on point estimators. In the following section, for example, we will see that $\bar{X}$ is a very reasonable point estimator of μ. As a result, the important confidence interval estimator of μ depends on knowledge of the sampling distribution of $\bar{X}$.

We begin the following section with the simplest case of a confidence interval. The scenario is simple and yet unrealistic. We are interested in estimating a population mean μ and yet σ is known. Clearly, if μ is unknown, it is quite unlikely that σ is known. Any historical results that produced enough information to allow the assumption that σ is known would likely have produced similar information about μ. Despite this argument, we begin with this case because the concepts and indeed the resulting mechanics associated with confidence interval estimation remain the same for the more realistic situations presented later in Section 9.4 and beyond.

9.4 Single Sample: Estimating the Mean

The sampling distribution of $\bar{X}$ is centered at μ, and in most applications the variance is smaller than that of any other estimators of μ. Thus, the sample mean $\bar{x}$ will be used as a point estimate for the population mean μ. Recall that $\sigma^2_{\bar{X}} = \sigma^2/n$, so a large sample will yield a value of $\bar{X}$ that comes from a sampling distribution with a small variance. Hence, $\bar{x}$ is likely to be a very accurate estimate of μ when n is large.

Let us now consider the interval estimate of μ. If our sample is selected from a normal population or, failing this, if n is sufficiently large, we can establish a confidence interval for μ by considering the sampling distribution of $\bar{X}$.

According to the Central Limit Theorem, we can expect the sampling distribution of $\bar{X}$ to be approximately normally distributed with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$. Writing $z_{\alpha/2}$ for the z-value above which we find an area of $\alpha/2$ under the normal curve, we can see from Figure 9.2 that

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha,$$

where

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

Hence,

$$P\left(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1 - \alpha.$$

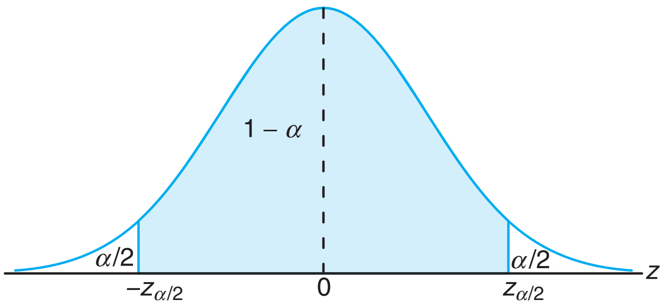

Figure 9.2: $P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$.

Multiplying each term in the inequality by $\sigma/\sqrt{n}$ and then subtracting $\bar{X}$ from each term and multiplying by -1 (reversing the sense of the inequalities), we obtain

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha.$$

A random sample of size n is selected from a population whose variance σ^2 is known, and the mean $\bar{x}$ is computed to give the $100(1-\alpha)\%$ confidence interval below. It is important to emphasize that we have invoked the Central Limit Theorem above. As a result, it is important to note the conditions for applications that follow.

Confidence Interval on μ, σ^2 Known

If $\bar{x}$ is the mean of a random sample of size n from a population with known variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is given by

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

For small samples selected from nonnormal populations, we cannot expect our degree of confidence to be accurate. However, for samples of size $n \geq 30$, with the shape of the distributions not too skewed, sampling theory guarantees good results.

Clearly, the values of the random variables $\hat{\Theta}_L$ and $\hat{\Theta}_U$, defined in Section 9.3, are the confidence limits

$$\hat{\theta}_L = \bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \quad \text{and} \quad \hat{\theta}_U = \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Different samples will yield different values of $\bar{x}$ and therefore produce different interval estimates of the parameter μ, as shown in Figure 9.3. The dot at the center of each interval indicates the position of the point estimate $\bar{x}$ for that random sample. Note that all of these intervals are of the same width, since their widths depend only on the choice of $z_{\alpha/2}$ once $\bar{x}$ is determined. The larger the value we choose for $z_{\alpha/2}$, the wider we make all the intervals and the more confident we can be that the particular sample selected will produce an interval that contains the unknown parameter μ. In general, for a selection of $z_{\alpha/2}$, $100(1-\alpha)\%$ of the intervals will cover μ.

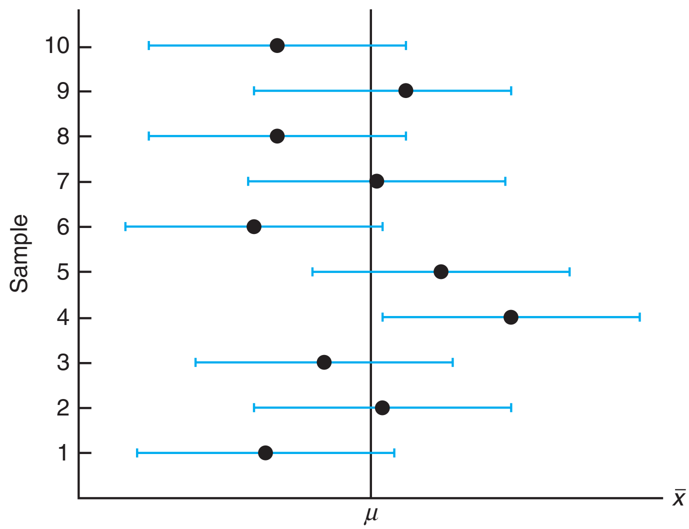

Figure 9.3: Interval estimates of μ for different samples.

Example 9.2: The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter.

Solution: The point estimate of μ is $\bar{x} = 2.6$. The z-value leaving an area of 0.025 to the right, and therefore an area of 0.975 to the left, is $z_{0.025} = 1.96$ (Table A.3). Hence, the 95% confidence interval is

$$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right),$$

which reduces to $2.50 < \mu < 2.70$. To find a 99% confidence interval, we find the z-value leaving an area of 0.005 to the right and 0.995 to the left. From Table A.3 again, $z_{0.005} = 2.575$, and the 99% confidence interval is

$$2.6 - (2.575)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (2.575)\left(\frac{0.3}{\sqrt{36}}\right),$$

or simply

$$2.47 < \mu < 2.73.$$

We now see that a longer interval is required to estimate μ with a higher degree of confidence.

The $100(1-\alpha)\%$ confidence interval provides an estimate of the accuracy of our point estimate. If μ is actually the center value of the interval, then $\bar{x}$ estimates μ without error. Most of the time, however, $\bar{x}$ will not be exactly equal to μ and the point estimate will be in error. The size of this error will be the absolute value of the difference between μ and $\bar{x}$, and we can be $100(1-\alpha)\%$ confident that this difference will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. We can readily see this if we draw a diagram of a hypothetical confidence interval, as in Figure 9.4.

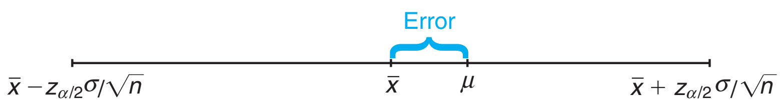

Figure 9.4: Error in estimating μ by $\bar{x}$.

Theorem 9.1: If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

In Example 9.2, we are 95% confident that the sample mean $\bar{x} = 2.6$ differs from the true mean μ by an amount less than $(1.96)(0.3)/\sqrt{36} = 0.1$ and 99% confident that the difference is less than $(2.575)(0.3)/\sqrt{36} = 0.13$.

Frequently, we wish to know how large a sample is necessary to ensure that the error in estimating μ will be less than a specified amount e. By Theorem 9.1, we must choose n such that $z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = e$. Solving this equation gives the following formula for n.

Theorem 9.2: If $\bar{x}$ is used as an estimate of μ, we can be $100(1-\alpha)\%$ confident that the error will not exceed a specified amount e when the sample size is

$$n = \left(\frac{z_{\alpha/2}\sigma}{e}\right)^2.$$

When solving for the sample size, n, we round all fractional values up to the next whole number. By adhering to this principle, we can be sure that our degree of confidence never falls below $100(1-\alpha)\%$.

Strictly speaking, the formula in Theorem 9.2 is applicable only if we know the variance of the population from which we select our sample. Lacking this information, we could take a preliminary sample of size $n \geq 30$ to provide an estimate of σ. Then, using s as an approximation for σ in Theorem 9.2, we could determine approximately how many observations are needed to provide the desired degree of accuracy.

Example 9.3: How large a sample is required if we want to be 95% confident that our estimate of μ in Example 9.2 is off by less than 0.05?

Solution: The population standard deviation is $\sigma = 0.3$. Then, by Theorem 9.2,

$$n = \left[\frac{(1.96)(0.3)}{0.05}\right]^2 = 138.3.$$

Therefore, we can be 95% confident that a random sample of size 139 will provide an estimate $\bar{x}$ differing from μ by an amount less than 0.05.

One-Sided Confidence Bounds

The confidence intervals and resulting confidence bounds discussed thus far are *two-sided* (i.e., both upper and lower bounds are given). However, there are many applications in which only one bound is sought. For example, if the measurement of interest is tensile strength, the engineer receives better information from a lower bound only. This bound communicates the worst-case scenario. On the other hand, if the measurement is something for which a relatively large value of μ is not profitable or desirable, then an upper confidence bound is of interest. An example would be a case in which inferences need to be made concerning the mean mercury composition in a river. An upper bound is very informative in this case.

One-sided confidence bounds are developed in the same fashion as two-sided intervals. However, the source is a one-sided probability statement that makes use of the Central Limit Theorem:

$$P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_\alpha\right) = 1 - \alpha.$$

One can then manipulate the probability statement much as before and obtain

$$P(\mu > \bar{X} - z_\alpha \sigma/\sqrt{n}) = 1 - \alpha.$$

Similar manipulation of $P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} > -z_\alpha\right) = 1 - \alpha$ gives

$$P(\mu < \bar{X} + z_\alpha \sigma/\sqrt{n}) = 1 - \alpha.$$

As a result, the upper and lower one-sided bounds follow.

One-Sided Confidence Bounds on μ, σ^2 Known

If $\bar{X}$ is the mean of a random sample of size n from a population with variance σ^2, the one-sided $100(1-\alpha)\%$ confidence bounds for μ are given by

upper one-sided bound: $\bar{x} + z_\alpha \sigma/\sqrt{n}$;

lower one-sided bound: $\bar{x} - z_\alpha \sigma/\sqrt{n}$.

Example 9.4: In a psychological testing experiment, 25 subjects are selected randomly and their reaction time, in seconds, to a particular stimulus is measured. Past experience suggests that the variance in reaction times to these types of stimuli is 4 sec^2 and that the distribution of reaction times is approximately normal. The average time for the subjects is 6.2 seconds. Give an upper 95% bound for the mean reaction time.

***Solution*:** The upper 95% bound is given by

$$\begin{aligned}\bar{x} + z_\alpha \sigma/\sqrt{n} &= 6.2 + (1.645)\sqrt{4/25} = 6.2 + 0.658 \\ &= 6.858 \text{ seconds.}\end{aligned}$$

Hence, we are 95% confident that the mean reaction time is less than 6.858 seconds.

The Case of σ Unknown

Frequently, we must attempt to estimate the mean of a population when the variance is unknown. The reader should recall learning in Chapter 8 that if we have a random sample from a *normal distribution*, then the random variable

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has a Student t-distribution with $n-1$ degrees of freedom. Here S is the sample standard deviation. In this situation, with σ unknown, T can be used to construct a confidence interval on μ. The procedure is the same as that with σ known except that σ is replaced by S and the standard normal distribution is replaced by the t-distribution. Referring to Figure 9.5, we can assert that

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha,$$

where $t_{\alpha/2}$ is the t-value with $n-1$ degrees of freedom, above which we find an area of $\alpha/2$. Because of symmetry, an equal area of $\alpha/2$ will fall to the left of $-t_{\alpha/2}$. Substituting for T, we write

$$P\left(-t_{\alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}\right) = 1 - \alpha.$$

Multiplying each term in the inequality by $S/\sqrt{n}$, and then subtracting $\bar{X}$ from each term and multiplying by -1, we obtain

$$P\left(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\right) = 1 - \alpha.$$

For a particular random sample of size n, the mean $\bar{x}$ and standard deviation s are computed and the following $100(1-\alpha)\%$ confidence interval for μ is obtained.

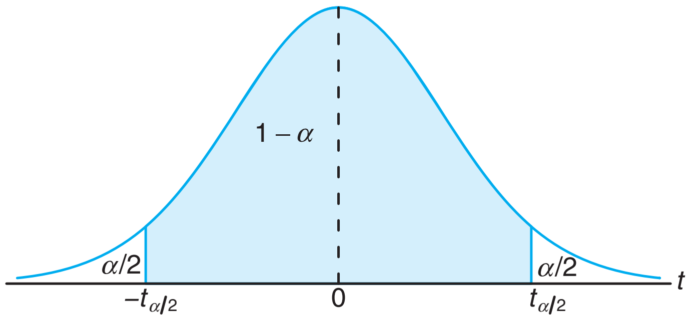

Figure 9.5: $P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha$.

Confidence Interval on μ, σ^2 Unknown

If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal population with unknown variance σ^2, a $100(1-\alpha)\%$ confidence interval for μ is

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$

where $t_{\alpha/2}$ is the t-value with $v = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

We have made a distinction between the cases of σ known and σ unknown in computing confidence interval estimates. We should emphasize that for σ known we exploited the Central Limit Theorem, whereas for σ unknown we made use of the sampling distribution of the random variable T. However, the use of the t-distribution is based on the premise that the sampling is from a normal distribution. As long as the distribution is approximately bell shaped, confidence intervals can be computed when σ^2 is unknown by using the t-distribution and we may expect very good results.

Computed one-sided confidence bounds for μ with σ unknown are as the reader would expect, namely

$$\bar{x} + t_{\alpha}\frac{s}{\sqrt{n}} \qquad \text{and} \qquad \bar{x} - t_{\alpha}\frac{s}{\sqrt{n}}.$$

They are the upper and lower $100(1-\alpha)\%$ bounds, respectively. Here t_α is the t-value having an area of α to the right.

Example 9.5: The contents of seven similar containers of sulfuric acid are 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, and 9.6 liters. Find a 95% confidence interval for the mean contents of all such containers, assuming an approximately normal distribution.

Solution: The sample mean and standard deviation for the given data are

$$\bar{x} = 10.0 \qquad \text{and} \qquad s = 0.283.$$

Using Table A.4, we find $t_{0.025} = 2.447$ for $v = 6$ degrees of freedom. Hence, the

95% confidence interval for μ is

$$10.0 - (2.447)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10.0 + (2.447)\left(\frac{0.283}{\sqrt{7}}\right),$$

which reduces to $9.74 < \mu < 10.26$.

Concept of a Large-Sample Confidence Interval

Often statisticians recommend that even when normality cannot be assumed, σ is unknown, and $n \geq 30$, s can replace σ and the confidence interval

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

may be used. This is often referred to as a *large-sample confidence interval*. The justification lies only in the presumption that with a sample as large as 30 and the population distribution not too skewed, s will be very close to the true σ and thus the Central Limit Theorem prevails. It should be emphasized that this is only an approximation and the quality of the result becomes better as the sample size grows larger.

Example 9.6: Scholastic Aptitude Test (SAT) mathematics scores of a random sample of 500 high school seniors in the state of Texas are collected, and the sample mean and standard deviation are found to be 501 and 112, respectively. Find a 99% confidence interval on the mean SAT mathematics score for seniors in the state of Texas.

***Solution*:** Since the sample size is large, it is reasonable to use the normal approximation. Using Table A.3, we find $z_{0.005} = 2.575$. Hence, a 99% confidence interval for μ is

$$501 \pm (2.575)\left(\frac{112}{\sqrt{500}}\right) = 501 \pm 12.9,$$

which yields $488.1 < \mu < 513.9$.

9.5 Standard Error of a Point Estimate

We have made a rather sharp distinction between the goal of a point estimate and that of a confidence interval estimate. The former supplies a single number extracted from a set of experimental data, and the latter provides an interval that is reasonable for the parameter, *given the experimental data*; that is, $100(1-\alpha)\%$ of such computed intervals "cover" the parameter.

These two approaches to estimation are related to each other. The common thread is the sampling distribution of the point estimator. Consider, for example, the estimator $\bar{X}$ of μ with σ known. We indicated earlier that a measure of the quality of an unbiased estimator is its variance. The variance of $\bar{X}$ is

$$\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}.$$

Thus, the standard deviation of $\bar{X}$, or *standard error* of $\bar{X}$, is $\sigma/\sqrt{n}$. Simply put, the standard error of an estimator is its standard deviation. For $\bar{X}$, the computed confidence limit

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \text{ is written as } \bar{x} \pm z_{\alpha/2} \text{ s.e.}(\bar{x}),$$

where "s.e." is the "standard error." The important point is that the width of the confidence interval on μ is dependent on the quality of the point estimator through its standard error. In the case where σ is unknown and sampling is from a normal distribution, s replaces σ and the *estimated standard error* $s/\sqrt{n}$ is involved. Thus, the confidence limits on μ are

Confidence Limits on μ, σ^2 Unknown

$$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = \bar{x} \pm t_{\alpha/2} \text{ s.e.}(\bar{x})$$

Again, the confidence interval is *no better* (in terms of width) *than the quality of the point estimate*, in this case through its estimated standard error. Computer packages often refer to estimated standard errors simply as "standard errors."

As we move to more complex confidence intervals, there is a prevailing notion that widths of confidence intervals become shorter as the quality of the corresponding point estimate becomes better, although it is not always quite as simple as we have illustrated here. It can be argued that a confidence interval is merely an augmentation of the point estimate to take into account the precision of the point estimate.

9.6 Prediction Intervals

The point and interval estimations of the mean in Sections 9.4 and 9.5 provide good information about the unknown parameter μ of a normal distribution or a nonnormal distribution from which a large sample is drawn. Sometimes, other than the population mean, the experimenter may also be interested in predicting the possible **value of a future observation**. For instance, in quality control, the experimenter may need to use the observed data to predict a new observation. A process that produces a metal part may be evaluated on the basis of whether the part meets specifications on tensile strength. On certain occasions, a customer may be interested in purchasing a **single part**. In this case, a confidence interval on the mean tensile strength does not capture the required information. The customer requires a statement regarding the uncertainty of a **single observation**. This type of requirement is nicely fulfilled by the construction of a **prediction interval**.

It is quite simple to obtain a prediction interval for the situations we have considered so far. Assume that the random sample comes from a normal population with unknown mean μ and known variance σ^2. A natural point estimator of a new observation is $\bar{X}$. It is known, from Section 8.4, that the variance of $\bar{X}$ is σ^2/n. However, to predict a new observation, not only do we need to account for the variation due to estimating the mean, but also we should account for the **variation of a future observation**. From the assumption, we know that the variance of the random error in a new observation is σ^2. The development of a prediction interval is best illustrated by beginning with a normal random variable $x_0 - \bar{x}$, where x_0 is the new observation and $\bar{x}$ comes from the sample. Since x_0 and $\bar{x}$ are independent, we know that

$$z = \frac{x_0 - \bar{x}}{\sqrt{\sigma^2 + \sigma^2/n}} = \frac{x_0 - \bar{x}}{\sigma\sqrt{1 + 1/n}}$$

is $n(z; 0, 1)$. As a result, if we use the probability statement

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

with the z-statistic above and place x_0 in the center of the probability statement, we have the following event occurring with probability $1 - \alpha$:

$$\bar{x} - z_{\alpha/2}\sigma\sqrt{1 + 1/n} < x_0 < \bar{x} + z_{\alpha/2}\sigma\sqrt{1 + 1/n}.$$

As a result, computation of the prediction interval is formalized as follows.

Prediction Interval of a Future Observation, σ^2 Known

For a normal distribution of measurements with unknown mean μ and known variance σ^2, a $100(1 - \alpha)\%$ **prediction interval** of a future observation x_0 is

$$\bar{x} - z_{\alpha/2}\sigma\sqrt{1 + 1/n} < x_0 < \bar{x} + z_{\alpha/2}\sigma\sqrt{1 + 1/n},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Example 9.7: Due to the decrease in interest rates, the First Citizens Bank received a lot of mortgage applications. A recent sample of 50 mortgage loans resulted in an average loan amount of \$257,300. Assume a population standard deviation of \$25,000. For the next customer who fills out a mortgage application, find a 95% prediction interval for the loan amount.

***Solution*:** The point prediction of the next customer's loan amount is $\bar{x} = \$257,300$. The z-value here is $z_{0.025} = 1.96$. Hence, a 95% prediction interval for the future loan amount is

$$257,300 - (1.96)(25,000)\sqrt{1 + 1/50} < x_0 < 257,300 + (1.96)(25,000)\sqrt{1 + 1/50},$$

which gives the interval (\$207,812.43, \$306,787.57).

The prediction interval provides a good estimate of the location of a future observation, which is quite different from the estimate of the sample mean value. It should be noted that the variation of this prediction is the sum of the variation due to an estimation of the mean and the variation of a single observation. However, as in the past, we first consider the case with known variance. It is also important to deal with the prediction interval of a future observation in the situation where the variance is unknown. Indeed a Student t-distribution may be used in this case, as described in the following result. The normal distribution is merely replaced by the t-distribution.

Prediction Interval of a Future Observation, σ^2 Unknown

For a normal distribution of measurements with unknown mean μ and unknown variance σ^2, a $100(1-\alpha)\%$ **prediction interval** of a future observation x_0 is

$$\bar{x} - t_{\alpha/2} s\sqrt{1 + 1/n} < x_0 < \bar{x} + t_{\alpha/2} s\sqrt{1 + 1/n},$$

where $t_{\alpha/2}$ is the t-value with $v = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

One-sided prediction intervals can also be constructed. Upper prediction bounds apply in cases where focus must be placed on future large observations. Concern over future small observations calls for the use of lower prediction bounds. The upper bound is given by

$$\bar{x} + t_{\alpha} s\sqrt{1 + 1/n}$$

and the lower bound by

$$\bar{x} - t_{\alpha} s\sqrt{1 + 1/n}.$$

Example 9.8: A meat inspector has randomly selected 30 packs of 95% lean beef. The sample resulted in a mean of 96.2% with a sample standard deviation of 0.8%. Find a 99% prediction interval for the leanness of a new pack. Assume normality.

***Solution*:** For $v = 29$ degrees of freedom, $t_{0.005} = 2.756$. Hence, a 99% prediction interval for a new observation x_0 is

$$96.2 - (2.756)(0.8)\sqrt{1 + \frac{1}{30}} < x_0 < 96.2 + (2.756)(0.8)\sqrt{1 + \frac{1}{30}},$$

which reduces to (93.96, 98.44).

Use of Prediction Limits for Outlier Detection

To this point in the text very little attention has been paid to the concept of **outliers**, or aberrant observations. The majority of scientific investigators are keenly sensitive to the existence of outlying observations or so-called faulty or "bad data." We deal with the concept of outlier detection extensively in Chapter 12. However, it is certainly of interest here since there is an important relationship between outlier detection and prediction intervals.

It is convenient for our purposes to view an outlying observation as one that comes from a population with a mean that is different from the mean that governs the rest of the sample of size n being studied. The prediction interval produces a bound that "covers" a future single observation with probability $1 - \alpha$ if it comes from the population from which the sample was drawn. As a result, a methodology for outlier detection involves the rule that **an observation is an outlier if it falls outside the prediction interval computed without including the questionable observation in the sample**. As a result, for the prediction interval of Example 9.8, if a new pack of beef is measured and its leanness is outside the interval $(93.96, 98.44)$, that observation can be viewed as an outlier.

9.7 Tolerance Limits

As discussed in Section 9.6, the scientist or engineer may be less interested in estimating parameters than in gaining a notion about where an individual *observation* or measurement might fall. Such situations call for the use of prediction intervals. However, there is yet a third type of interval that is of interest in many applications. Once again, suppose that interest centers around the manufacturing of a component part and specifications exist on a dimension of that part. In addition, there is little concern about the mean of the dimension. But unlike in the scenario in Section 9.6, one may be less interested in a single observation and more interested in where the majority of the population falls. If process specifications are important, the manager of the process is concerned about long-range performance, **not the next observation**. One must attempt to determine bounds that, in some probabilistic sense, "cover" values in the population (i.e., the measured values of the dimension).

One method of establishing the desired bounds is to determine a confidence interval on a *fixed proportion* of the measurements. This is best motivated by visualizing a situation in which we are doing random sampling from a normal distribution with known mean μ and variance σ^2. Clearly, a bound that covers the middle 95% of the population of observations is

$$\mu \pm 1.96\sigma.$$

This is called a **tolerance interval**, and indeed its coverage of 95% of measured observations is exact. However, in practice, μ and σ are seldom known; thus, the user must apply

$$\bar{x} \pm ks.$$

Now, of course, the interval is a random variable, and hence the *coverage* of a proportion of the population by the interval is not exact. As a result, a $100(1-\gamma)\%$ confidence interval must be used since $\bar{x} \pm ks$ cannot be expected to cover any specified proportion all the time. As a result, we have the following definition.

Tolerance Limits For a normal distribution of measurements with unknown mean μ and unknown standard deviation σ, **tolerance limits** are given by $\bar{x} \pm ks$, where k is determined such that one can assert with $100(1-\gamma)\%$ confidence that the given limits contain at least the proportion $1-\alpha$ of the measurements.

Table A.7 gives values of k for $1-\alpha = 0.90, 0.95, 0.99$; $\gamma = 0.05, 0.01$; and selected values of n from 2 to 300.

Example 9.9: Consider Example 9.8. With the information given, find a tolerance interval that gives two-sided 95% bounds on 90% of the distribution of packages of 95% lean beef. Assume the data came from an approximately normal distribution.

Solution: Recall from Example 9.8 that $n = 30$, the sample mean is 96.2%, and the sample standard deviation is 0.8%. From Table A.7, $k = 2.14$. Using

$$\bar{x} \pm ks = 96.2 \pm (2.14)(0.8),$$

we find that the lower and upper bounds are 94.5 and 97.9.

We are 95% confident that the above range covers the central 90% of the distribution of 95% lean beef packages.

Distinction among Confidence Intervals, Prediction Intervals, and Tolerance Intervals

It is important to reemphasize the difference among the three types of intervals discussed and illustrated in the preceding sections. The computations are straightforward, but interpretation can be confusing. In real-life applications, these intervals are not interchangeable because their interpretations are quite distinct.

In the case of confidence intervals, one is attentive only to the **population mean**. For example, Exercise 9.13 on page 283 deals with an engineering process that produces shearing pins. A specification will be set on Rockwell hardness, below which a customer will not accept any pins. Here, a population parameter must take a backseat. It is important that the engineer know where the *majority of the values of Rockwell hardness are going to be*. Thus, tolerance limits should be used. Surely, when tolerance limits on any process output are tighter than process specifications, that is good news for the process manager.

It is true that the tolerance limit interpretation is somewhat related to the confidence interval. The $100(1-\alpha)\%$ tolerance interval on, say, the proportion 0.95 can be viewed as a confidence interval **on the middle 95%** of the corresponding normal distribution. One-sided tolerance limits are also relevant. In the case of the Rockwell hardness problem, it is desirable to have a lower bound of the form $\bar{x} - ks$ such that there is 99% confidence that at least 99% of Rockwell hardness values will exceed the computed value.

Prediction intervals are applicable when it is important to determine a bound on a **single value**. The mean is not the issue here, nor is the location of the majority of the population. Rather, the location of a single new observation is required.

Case Study 9.1: **Machine Quality**: A machine produces metal pieces that are cylindrical in shape. A sample of these pieces is taken and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. Use these data to calculate three interval types and draw interpretations that illustrate the distinction between them in the context of the system. For all computations, assume an approximately normal distribution. The sample mean and standard deviation for the given data are $\bar{x} = 1.0056$ and $s = 0.0246$.

(a) Find a 99% confidence interval on the mean diameter.

(b) Compute a 99% prediction interval on a measured diameter of a single metal piece taken from the machine.

(c) Find the 99% tolerance limits that will contain 95% of the metal pieces produced by this machine.

***Solution*:** (a) The 99% confidence interval for the mean diameter is given by

$$\bar{x} \pm t_{0.005} s/\sqrt{n} = 1.0056 \pm (3.355)(0.0246/3) = 1.0056 \pm 0.0275.$$

Thus, the 99% confidence bounds are 0.9781 and 1.0331.

(b) The 99% prediction interval for a future observation is given by

$$\bar{x} \pm t_{0.005} s\sqrt{1 + 1/n} = 1.0056 \pm (3.355)(0.0246)\sqrt{1 + 1/9},$$

with the bounds being 0.9186 and 1.0926.

(c) From Table A.7, for $n = 9$, $1 - \gamma = 0.99$, and $1 - \alpha = 0.95$, we find $k = 4.550$ for two-sided limits. Hence, the 99% tolerance limits are given by

$$\bar{x} + ks = 1.0056 \pm (4.550)(0.0246),$$

with the bounds being 0.8937 and 1.1175. We are 99% confident that the tolerance interval from 0.8937 to 1.1175 will contain the central 95% of the distribution of diameters produced.

This case study illustrates that the three types of limits can give appreciably different results even though they are all 99% bounds. In the case of the confidence interval on the mean, 99% of such intervals cover the population mean diameter. Thus, we say that we are 99% confident that the mean diameter produced by the process is between 0.9781 and 1.0331 centimeters. Emphasis is placed on the mean, with less concern about a single reading or the general nature of the distribution of diameters in the population. In the case of the prediction limits, the bounds 0.9186 and 1.0926 are based on the distribution of a single "new" metal piece taken from the process, and again 99% of such limits will cover the diameter of a new measured piece. On the other hand, the tolerance limits, as suggested in the previous section, give the engineer a sense of where the "majority," say the central 95%, of the diameters of measured pieces in the population reside. The 99% tolerance limits, 0.8937 and 1.1175, are numerically quite different from the other two bounds. If these bounds appear alarmingly wide to the engineer, it reflects negatively on process quality. On the other hand, if the bounds represent a desirable result, the engineer may conclude that a majority (95% in here) of the diameters are in a desirable range. Again, a confidence interval interpretation may be used: namely, 99% of such calculated bounds will cover the middle 95% of the population of diameters.

Exercises

9.1 A UCLA researcher claims that the life span of mice can be extended by as much as 25% when the calories in their diet are reduced by approximately 40% from the time they are weaned. The restricted diet is enriched to normal levels by vitamins and protein. Assuming that it is known from previous studies that $\sigma = 5.8$ months, how many mice should be included in our sample if we wish to be 99% confident that the mean life span of the sample will be within 2 months of the population mean for all mice subjected to this reduced diet?

9.2 An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed with a standard deviation of 40 hours. If a sample of 30 bulbs has an average life of 780 hours, find a 96% confidence interval for the population mean of all bulbs produced by this firm.

9.3 Many cardiac patients wear an implanted pacemaker to control their heartbeat. A plastic connector module mounts on the top of the pacemaker. Assuming a standard deviation of 0.0015 inch and an approximately normal distribution, find a 95% confidence interval for the mean of the depths of all connector modules made by a certain manufacturing company. A random sample of 75 modules has an average depth of 0.310 inch.

9.4 The heights of a random sample of 50 college students showed a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters.

(a) Construct a 98% confidence interval for the mean height of all college students.

(b) What can we assert with 98% confidence about the possible size of our error if we estimate the mean height of all college students to be 174.5 centimeters?

9.5 A random sample of 100 automobile owners in the state of Virginia shows that an automobile is driven on average 23,500 kilometers per year with a standard deviation of 3900 kilometers. Assume the distribution of measurements to be approximately normal.

(a) Construct a 99% confidence interval for the average number of kilometers an automobile is driven annually in Virginia.

(b) What can we assert with 99% confidence about the possible size of our error if we estimate the average number of kilometers driven by car owners in Virginia to be 23,500 kilometers per year?

9.6 How large a sample is needed in Exercise 9.2 if we wish to be 96% confident that our sample mean will be within 10 hours of the true mean?

9.7 How large a sample is needed in Exercise 9.3 if we wish to be 95% confident that our sample mean will be within 0.0005 inch of the true mean?

9.8 An efficiency expert wishes to determine the average time that it takes to drill three holes in a certain metal clamp. How large a sample will she need to be 95% confident that her sample mean will be within 15 seconds of the true mean? Assume that it is known from previous studies that $\sigma = 40$ seconds.

9.9 Regular consumption of presweetened cereals contributes to tooth decay, heart disease, and other degenerative diseases, according to studies conducted by Dr. W. H. Bowen of the National Institute of Health and Dr. J. Yudben, Professor of Nutrition and Dietetics at the University of London. In a random sample consisting of 20 similar single servings of Alpha-Bits, the average sugar content was 11.3 grams with a standard deviation of 2.45 grams. Assuming that the sugar contents are normally distributed, construct a 95% confidence interval for the mean sugar content for single servings of Alpha-Bits.

9.10 A random sample of 12 graduates of a certain secretarial school typed an average of 79.3 words per minute with a standard deviation of 7.8 words per minute. Assuming a normal distribution for the number of words typed per minute, find a 95% confidence interval for the average number of words typed by all graduates of this school.

9.11 A machine produces metal pieces that are cylindrical in shape. A sample of pieces is taken, and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. Find a 99% confidence interval for the mean diameter of pieces from this machine, assuming an approximately normal distribution.

9.12 A random sample of 10 chocolate energy bars of a certain brand has, on average, 230 calories per bar, with a standard deviation of 15 calories. Construct a 99% confidence interval for the true mean calorie content of this brand of energy bar. Assume that the distribution of the calorie content is approximately normal.

9.13 A random sample of 12 shearing pins is taken in a study of the Rockwell hardness of the pin head. Measurements on the Rockwell hardness are made for each of the 12, yielding an average value of 48.50 with a sample standard deviation of 1.5. Assuming the measurements to be normally distributed, construct a 90% confidence interval for the mean Rockwell hardness.

9.14 The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint:

3.4	2.5	4.8	2.9	3.6
2.8	3.3	5.6	3.7	2.8
4.4	4.0	5.2	3.0	4.8

Assuming that the measurements represent a random sample from a normal population, find a 95% prediction interval for the drying time for the next trial of the paint.

9.15 Referring to Exercise 9.5, construct a 99% prediction interval for the kilometers traveled annually by an automobile owner in Virginia.

9.16 Consider Exercise 9.10. Compute the 95% prediction interval for the next observed number of words per minute typed by a graduate of the secretarial school.

9.17 Consider Exercise 9.9. Compute a 95% prediction interval for the sugar content of the next single serving of Alpha-Bits.

9.18 Referring to Exercise 9.13, construct a 95% tolerance interval containing 90% of the measurements.

9.19 A random sample of 25 tablets of buffered aspirin contains, on average, 325.05 mg of aspirin per tablet, with a standard deviation of 0.5 mg. Find the 95% tolerance limits that will contain 90% of the tablet contents for this brand of buffered aspirin. Assume that the aspirin content is normally distributed.

9.20 Consider the situation of Exercise 9.11. Estimation of the mean diameter, while important, is not nearly as important as trying to pin down the location of the majority of the distribution of diameters. Find the 95% tolerance limits that contain 95% of the diameters.

9.21 In a study conducted by the Department of Zoology at Virginia Tech, fifteen samples of water were collected from a certain station in the James River in order to gain some insight regarding the amount of orthophosphorus in the river. The concentration of the chemical is measured in milligrams per liter. Let us suppose that the mean at the station is not as important as the upper extreme of the distribution of the concentration of the chemical at the station. Concern centers around whether the concentration at the extreme is too large. Readings for the fifteen water samples gave a sample mean of 3.84 milligrams per liter and a sample standard deviation of 3.07 milligrams per liter. Assume that the readings are a random sample from a normal distribution. Calculate a prediction interval (upper 95% prediction limit) and a tolerance limit (95% upper tolerance limit that exceeds 95% of the population of values). Interpret both; that is, tell what each communicates about the upper extreme of the distribution of orthophosphorus at the sampling station.

9.22 A type of thread is being studied for its tensile strength properties. Fifty pieces were tested under similar conditions, and the results showed an average tensile strength of 78.3 kilograms and a standard deviation of 5.6 kilograms. Assuming a normal distribution of tensile strengths, give a lower 95% prediction limit on a single observed tensile strength value. In addition, give a lower 95% tolerance limit that is exceeded by 99% of the tensile strength values.

9.23 Refer to Exercise 9.22. Why are the quantities requested in the exercise likely to be more important to the manufacturer of the thread than, say, a confidence interval on the mean tensile strength?

9.24 Refer to Exercise 9.22 again. Suppose that specifications by a buyer of the thread are that the tensile strength of the material must be at least 62 kilograms. The manufacturer is satisfied if at most 5% of the manufactured pieces have tensile strength less than 62 kilograms. Is there cause for concern? Use a one-sided 99% tolerance limit that is exceeded by 95% of the tensile strength values.

9.25 Consider the drying time measurements in Exercise 9.14. Suppose the 15 observations in the data set are supplemented by a 16th value of 6.9 hours. In the context of the original 15 observations, is the 16th value an outlier? Show work.

9.26 Consider the data in Exercise 9.13. Suppose the manufacturer of the shearing pins insists that the Rockwell hardness of the product be less than or equal to 44.0 only 5% of the time. What is your reaction? Use a tolerance limit calculation as the basis for your judgment.

9.27 Consider the situation of Case Study 9.1 on page 281 with a larger sample of metal pieces. The diameters are as follows: 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 1.01, 1.03, 0.99, 1.00, 1.00, 0.99, 0.98, 1.01, 1.02, 0.99 centimeters. Once again the normality assumption may be made. Do the following and compare your results to those of the case study. Discuss how they are different and why.

(a) Compute a 99% confidence interval on the mean diameter.

(b) Compute a 99% prediction interval on the next diameter to be measured.

(c) Compute a 99% tolerance interval for coverage of the central 95% of the distribution of diameters.

9.28 In Section 9.3, we emphasized the notion of "most efficient estimator" by comparing the variance of two unbiased estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$. However, this does not take into account bias in case one or both estimators are not unbiased. Consider the quantity

$$MSE = E(\hat{\Theta} - \theta),$$

where MSE denotes **mean squared error**. The MSE is often used to compare two estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$ of θ when either or both is unbiased because (i) it is intuitively reasonable and (ii) it accounts for bias. Show that MSE can be written

$$\begin{aligned} MSE &= E[\hat{\Theta} - E(\hat{\Theta})]^2 + [E(\hat{\Theta} - \theta)]^2 \\ &= \text{Var}(\hat{\Theta}) + [\text{Bias}(\hat{\Theta})]^2. \end{aligned}$$

9.29 Let us define $S'^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/n$. Show that

$$E(S'^2) = [(n-1)/n]\sigma^2,$$

and hence S'^2 is a biased estimator for σ^2.

9.30 Consider S'^2, the estimator of σ^2, from Exercise 9.29. Analysts often use S'^2 rather than dividing $\sum_{i=1}^{n}(X_i - \bar{X})^2$ by $n-1$, the degrees of freedom in the sample.

(a) What is the bias of S'^2?

(b) Show that the bias of S'^2 approaches zero as $n \to \infty$.

9.31 If X is a binomial random variable, show that

(a) $\widehat{P} = X/n$ is an unbiased estimator of p;

(b) $P' = \frac{X+\sqrt{n}/2}{n+\sqrt{n}}$ is a biased estimator of p.

9.32 Show that the estimator P' of Exercise 9.31(b) becomes unbiased as $n \to \infty$.

9.33 Compare S^2 and S'^2 (see Exercise 9.29), the two estimators of σ^2, to determine which is more efficient. Assume these estimators are found using $X_1, X_2, \ldots, X_n$, independent random variables from $n(x; \mu, \sigma)$. Which estimator is more efficient considering only the variance of the estimators? [*Hint*: Make use of Theorem 8.4 and the fact that the variance of χ^2_v is $2v$, from Section 6.7.]

9.34 Consider Exercise 9.33. Use the MSE discussed in Exercise 9.28 to determine which estimator is more efficient. Write out

$$\frac{MSE(S^2)}{MSE(S'^2)}.$$

9.8 Two Samples: Estimating the Difference between Two Means

If we have two populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively, a point estimator of the difference between μ_1 and μ_2 is given by the statistic $\bar{X}_1 - \bar{X}_2$. Therefore, to obtain a point estimate of $\mu_1 - \mu_2$, we shall select two independent random samples, one from each population, of sizes n_1 and n_2, and compute $\bar{x}_1 - \bar{x}_2$, the difference of the sample means. Clearly, we must consider the sampling distribution of $\bar{X}_1 - \bar{X}_2$.

According to Theorem 8.3, we can expect the sampling distribution of $\bar{X}_1 - \bar{X}_2$ to be approximately normally distributed with mean $\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$ and standard deviation $\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$. Therefore, we can assert with a probability of $1 - \alpha$ that the standard normal variable

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

will fall between $-z_{\alpha/2}$ and $z_{\alpha/2}$. Referring once again to Figure 9.2, we write

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha.$$

Substituting for Z, we state equivalently that

$$P\left(-z_{\alpha/2} < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} < z_{\alpha/2}\right) = 1 - \alpha,$$

which leads to the following $100(1 - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$.

Confidence Interval for $\mu_1 - \mu_2$, σ_1^2 and σ_2^2 Known

If $\bar{x}_1$ and $\bar{x}_2$ are means of independent random samples of sizes n_1 and n_2 from populations with known variances σ_1^2 and σ_2^2, respectively, a $100(l - \alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

The degree of confidence is exact when samples are selected from normal populations. For nonnormal populations, the Central Limit Theorem allows for a good approximation for reasonable size samples.

The Experimental Conditions and the Experimental Unit

For the case of confidence interval estimation on the difference between two means, we need to consider the experimental conditions in the data-taking process. It is assumed that we have two independent random samples from distributions with means μ_1 and μ_2, respectively. It is important that experimental conditions emulate this ideal described by these assumptions as closely as possible. Quite often, the experimenter should plan the strategy of the experiment accordingly. For almost any study of this type, there is a so-called *experimental unit*, which is that part of the experiment that produces experimental error and is responsible for the population variance we refer to as σ^2. In a drug study, the experimental unit is the patient or subject. In an agricultural experiment, it may be a plot of ground. In a chemical experiment, it may be a quantity of raw materials. It is important that differences between the experimental units have minimal impact on the results. The experimenter will have a degree of insurance that experimental units will not bias results if the conditions that define the two populations are *randomly assigned* to the experimental units. We shall again focus on randomization in future chapters that deal with hypothesis testing.

Example 9.10: A study was conducted in which two types of engines, A and B, were compared. Gas mileage, in miles per gallon, was measured. Fifty experiments were conducted using engine type A and 75 experiments were done with engine type B. The gasoline used and other conditions were held constant. The average gas mileage was 36 miles per gallon for engine A and 42 miles per gallon for engine B. Find a 96% confidence interval on $\mu_B - \mu_A$, where μ_A and μ_B are population mean gas mileages for engines A and B, respectively. Assume that the population standard deviations are 6 and 8 for engines A and B, respectively.

***Solution*:** The point estimate of $\mu_B - \mu_A$ is $\bar{x}_B - \bar{x}_A = 42 - 36 = 6$. Using $\alpha = 0.04$, we find $z_{0.02} = 2.05$ from Table A.3. Hence, with substitution in the formula above, the 96% confidence interval is

$$6 - 2.05\sqrt{\frac{64}{75} + \frac{36}{50}} < \mu_B - \mu_A < 6 + 2.05\sqrt{\frac{64}{75} + \frac{36}{50}},$$

or simply $3.43 < \mu_B - \mu_A < 8.57$.

This procedure for estimating the difference between two means is applicable if σ_1^2 and σ_2^2 are known. If the variances are not known and the two distributions involved are approximately normal, the t-distribution becomes involved, as in the case of a single sample. If one is not willing to assume normality, large samples (say greater than 30) will allow the use of s_1 and s_2 in place of σ_1 and σ_2, respectively, with the rationale that $s_1 \approx \sigma_1$ and $s_2 \approx \sigma_2$. Again, of course, the confidence interval is an approximate one.

Variances Unknown but Equal

Consider the case where σ_1^2 and σ_2^2 are unknown. If $\sigma_1^2 = \sigma_2^2 = \sigma^2$, we obtain a standard normal variable of the form

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2[(1/n_1) + (1/n_2)]}}.$$

According to Theorem 8.4, the two random variables

$$\frac{(n_1 - 1)S_1^2}{\sigma^2} \quad \text{and} \quad \frac{(n_2 - 1)S_2^2}{\sigma^2}$$

have chi-squared distributions with $n_1 - 1$ and $n_2 - 1$ degrees of freedom, respectively. Furthermore, they are independent chi-squared variables, since the random samples were selected independently. Consequently, their sum

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

has a chi-squared distribution with $v = n_1 + n_2 - 2$ degrees of freedom.

Since the preceding expressions for Z and V can be shown to be independent, it follows from Theorem 8.5 that the statistic

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2[(1/n_1) + (1/n_2)]}} \Bigg/ \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{\sigma^2(n_1 + n_2 - 2)}}$$

has the t-distribution with $v = n_1 + n_2 - 2$ degrees of freedom.

A point estimate of the unknown common variance σ^2 can be obtained by pooling the sample variances. Denoting the pooled estimator by S_p^2, we have the following.

Pooled Estimate of Variance

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

Substituting S_p^2 in the T statistic, we obtain the less cumbersome form

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{(1/n_1) + (1/n_2)}}.$$

Using the T statistic, we have

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha,$$

where $t_{\alpha/2}$ is the t-value with $n_1 + n_2 - 2$ degrees of freedom, above which we find an area of $\alpha/2$. Substituting for T in the inequality, we write

$$P\left[-t_{\alpha/2} < \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{(1/n_1) + (1/n_2)}} < t_{\alpha/2}\right] = 1 - \alpha.$$

After the usual mathematical manipulations, the difference of the sample means $\bar{x}_1 - \bar{x}_2$ and the pooled variance are computed and then the following $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is obtained.

The value of s_p^2 is easily seen to be a weighted average of the two sample variances s_1^2 and s_2^2, where the weights are the degrees of freedom.

Confidence Interval for $\mu_1 - \mu_2$, $\sigma_1^2 = \sigma_2^2$ but Both Unknown

If $\bar{x}_1$ and $\bar{x}_2$ are the means of independent random samples of sizes n_1 and n_2, respectively, from approximately normal populations with unknown but equal variances, a $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

where s_p is the pooled estimate of the population standard deviation and $t_{\alpha/2}$ is the t-value with $v = n_1 + n_2 - 2$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Example 9.11: The article "Macroinvertebrate Community Structure as an Indicator of Acid Mine Pollution," published in the *Journal of Environmental Pollution*, reports on an investigation undertaken in Cane Creek, Alabama, to determine the relationship between selected physiochemical parameters and different measures of macroinvertebrate community structure. One facet of the investigation was an evaluation of the effectiveness of a numerical species diversity index to indicate aquatic degradation due to acid mine drainage. Conceptually, a high index of macroinvertebrate species diversity should indicate an unstressed aquatic system, while a low diversity index should indicate a stressed aquatic system.

Two independent sampling stations were chosen for this study, one located downstream from the acid mine discharge point and the other located upstream. For 12 monthly samples collected at the downstream station, the species diversity index had a mean value $\bar{x}_1 = 3.11$ and a standard deviation $s_1 = 0.771$, while 10 monthly samples collected at the upstream station had a mean index value $\bar{x}_2 = 2.04$ and a standard deviation $s_2 = 0.448$. Find a 90% confidence interval for the difference between the population means for the two locations, assuming that the populations are approximately normally distributed with equal variances.

Solution: Let μ_1 and μ_2 represent the population means, respectively, for the species diversity indices at the downstream and upstream stations. We wish to find a 90% confidence interval for $\mu_1 - \mu_2$. Our point estimate of $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 = 3.11 - 2.04 = 1.07.$$

The pooled estimate, s_p^2, of the common variance, σ^2, is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(11)(0.771^2) + (9)(0.448^2)}{12 + 10 - 2} = 0.417.$$

Taking the square root, we obtain $s_p = 0.646$. Using $\alpha = 0.1$, we find in Table A.4 that $t_{0.05} = 1.725$ for $v = n_1 + n_2 - 2 = 20$ degrees of freedom. Therefore, the 90% confidence interval for $\mu_1 - \mu_2$ is

$$1.07 - (1.725)(0.646)\sqrt{\frac{1}{12} + \frac{1}{10}} < \mu_1 - \mu_2 < 1.07 + (1.725)(0.646)\sqrt{\frac{1}{12} + \frac{1}{10}},$$

which simplifies to $0.593 < \mu_1 - \mu_2 < 1.547$.

Interpretation of the Confidence Interval

For the case of a single parameter, the confidence interval simply provides error bounds on the parameter. Values contained in the interval should be viewed as reasonable values given the experimental data. In the case of a difference between two means, the interpretation can be extended to one of comparing the two means. For example, if we have high confidence that a difference $\mu_1 - \mu_2$ is positive, we would certainly infer that $\mu_1 > \mu_2$ with little risk of being in error. For example, in Example 9.11, we are 90% confident that the interval from 0.593 to 1.547 contains the difference of the population means for values of the species diversity index at the two stations. The fact that both confidence limits are positive indicates that, on the average, the index for the station located downstream from the discharge point is greater than the index for the station located upstream.

Equal Sample Sizes

The procedure for constructing confidence intervals for $\mu_1 - \mu_2$ with $\sigma_1 = \sigma_2 = \sigma$ unknown requires the assumption that the populations are normal. Slight departures from either the equal variance or the normality assumption do not seriously alter the degree of confidence for our interval. (A procedure is presented in Chapter 10 for testing the equality of two unknown population variances based on the information provided by the sample variances.) If the population variances are considerably different, we still obtain reasonable results when the populations are normal, provided that $n_1 = n_2$. Therefore, in planning an experiment, one should make every effort to equalize the size of the samples.

Unknown and Unequal Variances

Let us now consider the problem of finding an interval estimate of $\mu_1 - \mu_2$ when the unknown population variances are not likely to be equal. The statistic most often used in this case is

$$T' = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(S_1^2/n_1) + (S_2^2/n_2)}},$$

which has approximately a t-distribution with v degrees of freedom, where

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1 - 1)] + [(s_2^2/n_2)^2/(n_2 - 1)]}.$$

Since v is seldom an integer, we *round it down* to the nearest whole number. The above estimate of the degrees of freedom is called the Satterthwaite approximation (Satterthwaite, 1946, in the Bibliography).

Using the statistic T', we write

$$P(-t_{\alpha/2} < T' < t_{\alpha/2}) \approx 1 - \alpha,$$

where $t_{\alpha/2}$ is the value of the t-distribution with v degrees of freedom, above which we find an area of $\alpha/2$. Substituting for T' in the inequality and following the same steps as before, we state the final result.

Confidence Interval for $\mu_1 - \mu_2$, $\sigma_1^2 \neq \sigma_2^2$ and Both Unknown

If $\bar{x}_1$ and s_1^2 and $\bar{x}_2$ and s_2^2 are the means and variances of independent random samples of sizes n_1 and n_2, respectively, from approximately normal populations with unknown and unequal variances, an approximate $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is given by

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

where $t_{\alpha/2}$ is the t-value with

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1-1)] + [(s_2^2/n_2)^2/(n_2-1)]}$$

degrees of freedom, leaving an area of $\alpha/2$ to the right.

Note that the expression for v above involves random variables, and thus v is an *estimate* of the degrees of freedom. In applications, this estimate will not result in a whole number, and thus the analyst must round down to the nearest integer to achieve the desired confidence.

Before we illustrate the above confidence interval with an example, we should point out that all the confidence intervals on $\mu_1 - \mu_2$ are of the same general form as those on a single mean; namely, they can be written as

$$\text{point estimate} \ \pm \ t_{\alpha/2} \ \widehat{\text{s.e.}}(\text{point estimate})$$

or

$$\text{point estimate} \ \pm \ z_{\alpha/2} \ \text{s.e.}(\text{point estimate}).$$

For example, in the case where $\sigma_1 = \sigma_2 = \sigma$, the estimated standard error of $\bar{x}_1 - \bar{x}_2$ is $s_p\sqrt{1/n_1 + 1/n_2}$. For the case where $\sigma_1^2 \neq \sigma_2^2$,

$$\widehat{\text{s.e.}}(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}.$$

Example 9.12: A study was conducted by the Department of Zoology at the Virginia Tech to estimate the difference in the amounts of the chemical orthophosphorus measured at two different stations on the James River. Orthophosphorus was measured in milligrams per liter. Fifteen samples were collected from station 1, and 12 samples were obtained from station 2. The 15 samples from station 1 had an average orthophosphorus content of 3.84 milligrams per liter and a standard deviation of 3.07 milligrams per liter, while the 12 samples from station 2 had an average content of 1.49 milligrams per liter and a standard deviation of 0.80 milligram per liter. Find a 95% confidence interval for the difference in the true average orthophosphorus contents at these two stations, assuming that the observations came from normal populations with different variances.

***Solution*:** For station 1, we have $\bar{x}_1 = 3.84$, $s_1 = 3.07$, and $n_1 = 15$. For station 2, $\bar{x}_2 = 1.49$, $s_2 = 0.80$, and $n_2 = 12$. We wish to find a 95% confidence interval for $\mu_1 - \mu_2$.

Since the population variances are assumed to be unequal, we can only find an approximate 95% confidence interval based on the t-distribution with v degrees of freedom, where

$$v = \frac{(3.07^2/15 + 0.80^2/12)^2}{[(3.07^2/15)^2/14] + [(0.80^2/12)^2/11]} = 16.3 \approx 16.$$

Our point estimate of $\mu_1 - \mu_2$ is

$$\bar{x}_1 - \bar{x}_2 = 3.84 - 1.49 = 2.35.$$

Using $\alpha = 0.05$, we find in Table A.4 that $t_{0.025} = 2.120$ for $v = 16$ degrees of freedom. Therefore, the 95% confidence interval for $\mu_1 - \mu_2$ is

$$2.35 - 2.120\sqrt{\frac{3.07^2}{15} + \frac{0.80^2}{12}} < \mu_1 - \mu_2 < 2.35 + 2.120\sqrt{\frac{3.07^2}{15} + \frac{0.80^2}{12}},$$

which simplifies to $0.60 < \mu_1 - \mu_2 < 4.10$. Hence, we are 95% confident that the interval from 0.60 to 4.10 milligrams per liter contains the difference of the true average orthophosphorus contents for these two locations. ■

When two population variances are unknown, the assumption of equal variances or unequal variances may be precarious. In Section 10.10, a procedure will be introduced that will aid in discriminating between the equal variance and the unequal variance situation.

9.9 Paired Observations

At this point, we shall consider estimation procedures for the difference of two means when the samples are not independent and the variances of the two populations are not necessarily equal. The situation considered here deals with a very special experimental condition, namely that of *paired observations.* Unlike in the situation described earlier, the conditions of the two populations are not assigned randomly to experimental units. Rather, each homogeneous experimental unit receives both population conditions; as a result, each experimental unit has a pair of observations, one for each population. For example, if we run a test on a new diet using 15 individuals, the weights before and after going on the diet form the information for our two samples. The two populations are "before" and "after," and the experimental unit is the individual. Obviously, the observations in a pair have something in common. To determine if the diet is effective, we consider the differences $d_1, d_2, \ldots, d_n$ in the paired observations. These differences are the values of a random sample $D_1, D_2, \ldots, D_n$ from a population of differences that we shall assume to be normally distributed with mean $\mu_D = \mu_1 - \mu_2$ and variance σ_D^2. We estimate σ_D^2 by s_d^2, the variance of the differences that constitute our sample. The point estimator of μ_D is given by $\bar{D}$.

When Should Pairing Be Done?

Pairing observations in an experiment is a strategy that can be employed in many fields of application. The reader will be exposed to this concept in material related to hypothesis testing in Chapter 10 and experimental design issues in Chapters 13 and 15. Selecting experimental units that are relatively homogeneous (within the units) and allowing each unit to experience both population conditions reduces the effective experimental error variance (in this case, σ_D^2). The reader may visualize the ith pair difference as

$$D_i = X_{1i} - X_{2i}.$$

Since the two observations are taken on the sample experimental unit, they are not independent and, in fact,

$$\text{Var}(D_i) = \text{Var}(X_{1i} - X_{2i}) = \sigma_1^2 + \sigma_2^2 - 2\ \text{Cov}(X_{1i}, X_{2i}).$$

Now, intuitively, we expect that σ_D^2 should be reduced because of the similarity in nature of the "errors" of the two observations within a given experimental unit, and this comes through in the expression above. One certainly expects that if the unit is homogeneous, the covariance is positive. As a result, the gain in quality of the confidence interval over that obtained without pairing will be greatest when there is homogeneity within units and large differences as one goes from unit to unit. One should keep in mind that the performance of the confidence interval will depend on the standard error of $\bar{D}$, which is, of course, $\sigma_D/\sqrt{n}$, where n is the number of pairs. As we indicated earlier, the intent of pairing is to reduce σ_D.

Tradeoff between Reducing Variance and Losing Degrees of Freedom

Comparing the confidence intervals obtained with and without pairing makes apparent that there is a tradeoff involved. Although pairing should indeed reduce variance and hence reduce the standard error of the point estimate, the degrees of freedom are reduced by reducing the problem to a one-sample problem. As a result, the $t_{\alpha/2}$ point attached to the standard error is adjusted accordingly. Thus, pairing may be counterproductive. This would certainly be the case if one experienced only a modest reduction in variance (through σ_D^2) by pairing.

Another illustration of pairing involves choosing n pairs of subjects, with each pair having a similar characteristic such as IQ, age, or breed, and then selecting one member of each pair at random to yield a value of X_1, leaving the other member to provide the value of X_2. In this case, X_1 and X_2 might represent the grades obtained by two individuals of equal IQ when one of the individuals is assigned at random to a class using the conventional lecture approach while the other individual is assigned to a class using programmed materials.

A $100(1-\alpha)\%$ confidence interval for μ_D can be established by writing

$$P(-t_{\alpha/2} < T < t_{\alpha/2}) = 1 - \alpha,$$

where $T = \frac{\bar{D}-\mu_D}{S_d/\sqrt{n}}$ and $t_{\alpha/2}$, as before, is a value of the t-distribution with $n-1$ degrees of freedom.

It is now a routine procedure to replace T by its definition in the inequality above and carry out the mathematical steps that lead to the following $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2 = \mu_D$.

Confidence Interval for $\mu_D = \mu_1 - \mu_2$ for Paired Observations

If $\bar{d}$ and s_d are the mean and standard deviation, respectively, of the normally distributed differences of n random pairs of measurements, a $100(1-\alpha)\%$ confidence interval for $\mu_D = \mu_1 - \mu_2$ is

$$\bar{d} - t_{\alpha/2}\frac{s_d}{\sqrt{n}} < \mu_D < \bar{d} + t_{\alpha/2}\frac{s_d}{\sqrt{n}},$$

where $t_{\alpha/2}$ is the t-value with $v = n - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right.

Example 9.13: A study published in *Chemosphere* reported the levels of the dioxin TCDD of 20 Massachusetts Vietnam veterans who were possibly exposed to Agent Orange. The TCDD levels in plasma and in fat tissue are listed in Table 9.1.

Find a 95% confidence interval for $\mu_1 - \mu_2$, where μ_1 and μ_2 represent the true mean TCDD levels in plasma and in fat tissue, respectively. Assume the distribution of the differences to be approximately normal.

Table 9.1: Data for Example 9.13

Veteran	TCDD Levels in Plasma	TCDD Levels in Fat Tissue	d_i	Veteran	TCDD Levels in Plasma	TCDD Levels in Fat Tissue	d_i
1	2.5	4.9	−2.4	11	6.9	7.0	−0.1
2	3.1	5.9	−2.8	12	3.3	2.9	0.4
3	2.1	4.4	−2.3	13	4.6	4.6	0.0
4	3.5	6.9	−3.4	14	1.6	1.4	0.2
5	3.1	7.0	−3.9	15	7.2	7.7	−0.5
6	1.8	4.2	−2.4	16	1.8	1.1	0.7
7	6.0	10.0	−4.0	17	20.0	11.0	9.0
8	3.0	5.5	−2.5	18	2.0	2.5	−0.5
9	36.0	41.0	−5.0	19	2.5	2.3	0.2
10	4.7	4.4	0.3	20	4.1	2.5	1.6

Source: Schecter, A. et al. "Partitioning of 2,3,7,8-chlorinated dibenzo-*p*-dioxins and dibenzofurans between adipose tissue and plasma lipid of 20 Massachusetts Vietnam veterans," *Chemosphere*, Vol. 20, Nos. 7–9, 1990, pp. 954–955 (Tables I and II).

Solution: We wish to find a 95% confidence interval for $\mu_1 - \mu_2$. Since the observations are paired, $\mu_1 - \mu_2 = \mu_D$. The point estimate of μ_D is $\bar{d} = -0.87$. The standard deviation, s_d, of the sample differences is

$$s_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2} = \sqrt{\frac{168.4220}{19}} = 2.9773.$$

Using $\alpha = 0.05$, we find in Table A.4 that $t_{0.025} = 2.093$ for $v = n - 1 = 19$ degrees of freedom. Therefore, the 95% confidence interval is

$$-0.8700 - (2.093)\left(\frac{2.9773}{\sqrt{20}}\right) < \mu_D < -0.8700 + (2.093)\left(\frac{2.9773}{\sqrt{20}}\right),$$

or simply $-2.2634 < \mu_D < 0.5234$, from which we can conclude that there is no significant difference between the mean TCDD level in plasma and the mean TCDD level in fat tissue.

Exercises

9.35 A random sample of size $n_1 = 25$, taken from a normal population with a standard deviation $\sigma_1 = 5$, has a mean $\bar{x}_1 = 80$. A second random sample of size $n_2 = 36$, taken from a different normal population with a standard deviation $\sigma_2 = 3$, has a mean $\bar{x}_2 = 75$. Find a 94% confidence interval for $\mu_1 - \mu_2$.

9.36 Two kinds of thread are being compared for strength. Fifty pieces of each type of thread are tested under similar conditions. Brand A has an average tensile strength of 78.3 kilograms with a standard deviation of 5.6 kilograms, while brand B has an average tensile strength of 87.2 kilograms with a standard deviation of 6.3 kilograms. Construct a 95% confidence interval for the difference of the population means.

9.37 A study was conducted to determine if a certain treatment has any effect on the amount of metal removed in a pickling operation. A random sample of 100 pieces was immersed in a bath for 24 hours without the treatment, yielding an average of 12.2 millimeters of metal removed and a sample standard deviation of 1.1 millimeters. A second sample of 200 pieces was exposed to the treatment, followed by the 24-hour immersion in the bath, resulting in an average removal of 9.1 millimeters of metal with a sample standard deviation of 0.9 millimeter. Compute a 98% confidence interval estimate for the difference between the population means. Does the treatment appear to reduce the mean amount of metal removed?

9.38 Two catalysts in a batch chemical process, are being compared for their effect on the output of the process reaction. A sample of 12 batches was prepared using catalyst 1, and a sample of 10 batches was prepared using catalyst 2. The 12 batches for which catalyst 1 was used in the reaction gave an average yield of 85 with a sample standard deviation of 4, and the 10 batches for which catalyst 2 was used gave an average yield of 81 and a sample standard deviation of 5. Find a 90% confidence interval for the difference between the population means, assuming that the populations are approximately normally distributed with equal variances.

9.39 Students may choose between a 3-semester-hour physics course without labs and a 4-semester-hour course with labs. The final written examination is the same for each section. If 12 students in the section with labs made an average grade of 84 with a standard deviation of 4, and 18 students in the section without labs made an average grade of 77 with a standard deviation of 6, find a 99% confidence interval for the difference between the average grades for the two courses. Assume the populations to be approximately normally distributed with equal variances.

9.40 In a study conducted at Virginia Tech on the development of ectomycorrhizal, a symbiotic relationship between the roots of trees and a fungus, in which minerals are transferred from the fungus to the trees and sugars from the trees to the fungus, 20 northern red oak seedlings exposed to the fungus *Pisolithus tinctorus* were grown in a greenhouse. All seedlings were planted in the same type of soil and received the same amount of sunshine and water. Half received no nitrogen at planting time, to serve as a control, and the other half received 368 ppm of nitrogen in the form $NaNO_3$. The stem weights, in grams, at the end of 140 days were recorded as follows:

No Nitrogen	Nitrogen
0.32	0.26
0.53	0.43
0.28	0.47
0.37	0.49
0.47	0.52
0.43	0.75
0.36	0.79
0.42	0.86
0.38	0.62
0.43	0.46

Construct a 95% confidence interval for the difference in the mean stem weight between seedlings that receive no nitrogen and those that receive 368 ppm of nitrogen. Assume the populations to be normally distributed with equal variances.

9.41 The following data represent the length of time, in days, to recovery for patients randomly treated with one of two medications to clear up severe bladder infections:

Medication 1	Medication 2
$n_1 = 14$	$n_2 = 16$
$\bar{x}_1 = 17$	$\bar{x}_2 = 19$
$s_1^2 = 1.5$	$s_2^2 = 1.8$

Find a 99% confidence interval for the difference $\mu_2 - \mu_1$

in the mean recovery times for the two medications, assuming normal populations with equal variances.

9.42 An experiment reported in *Popular Science* compared fuel economies for two types of similarly equipped diesel mini-trucks. Let us suppose that 12 Volkswagen and 10 Toyota trucks were tested in 90-kilometer-per-hour steady-paced trials. If the 12 Volkswagen trucks averaged 16 kilometers per liter with a standard deviation of 1.0 kilometer per liter and the 10 Toyota trucks averaged 11 kilometers per liter with a standard deviation of 0.8 kilometer per liter, construct a 90% confidence interval for the difference between the average kilometers per liter for these two mini-trucks. Assume that the distances per liter for the truck models are approximately normally distributed with equal variances.

9.43 A taxi company is trying to decide whether to purchase brand A or brand B tires for its fleet of taxis. To estimate the difference in the two brands, an experiment is conducted using 12 of each brand. The tires are run until they wear out. The results are

Brand A: $\bar{x}_1 = 36,300$ kilometers, $s_1 = 5000$ kilometers.
Brand B: $\bar{x}_2 = 38,100$ kilometers, $s_2 = 6100$ kilometers.

Compute a 95% confidence interval for $\mu_A - \mu_B$ assuming the populations to be approximately normally distributed. You may not assume that the variances are equal.

9.44 Referring to Exercise 9.43, find a 99% confidence interval for $\mu_1 - \mu_2$ if tires of the two brands are assigned at random to the left and right rear wheels of 8 taxis and the following distances, in kilometers, are recorded:

Taxi	Brand A	Brand B
1	34,400	36,700
2	45,500	46,800
3	36,700	37,700
4	32,000	31,100
5	48,400	47,800
6	32,800	36,400
7	38,100	38,900
8	30,100	31,500

Assume that the differences of the distances are approximately normally distributed.

9.45 The federal government awarded grants to the agricultural departments of 9 universities to test the yield capabilities of two new varieties of wheat. Each variety was planted on a plot of equal area at each university, and the yields, in kilograms per plot, were recorded as follows:

	University								
Variety	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	38	23	35	41	44	29	37	31	38
2	45	25	31	38	50	33	36	40	43

Find a 95% confidence interval for the mean difference between the yields of the two varieties, assuming the differences of yields to be approximately normally distributed. Explain why pairing is necessary in this problem.

9.46 The following data represent the running times of films produced by two motion-picture companies.

Company	Time (minutes)						
I	103	94	110	87	98		
II	97	82	123	92	175	88	118

Compute a 90% confidence interval for the difference between the average running times of films produced by the two companies. Assume that the running-time differences are approximately normally distributed with unequal variances.

9.47 *Fortune* magazine (March 1997) reported the total returns to investors for the 10 years prior to 1996 and also for 1996 for 431 companies. The total returns for 10 of the companies are listed below. Find a 95% confidence interval for the mean change in percent return to investors.

	Total Return to Investors	
Company	**1986–96**	**1996**
Coca-Cola	29.8%	43.3%
Mirage Resorts	27.9%	25.4%
Merck	22.1%	24.0%
Microsoft	44.5%	88.3%
Johnson & Johnson	22.2%	18.1%
Intel	43.8%	131.2%
Pfizer	21.7%	34.0%
Procter & Gamble	21.9%	32.1%
Berkshire Hathaway	28.3%	6.2%
S&P 500	11.8%	20.3%

9.48 An automotive company is considering two types of batteries for its automobile. Sample information on battery life is collected for 20 batteries of type A and 20 batteries of type B. The summary statistics are $\bar{x}_A = 32.91$, $\bar{x}_B = 30.47$, $s_A = 1.57$, and $s_B = 1.74$. Assume the data on each battery are normally distributed and assume $\sigma_A = \sigma_B$.

(a) Find a 95% confidence interval on $\mu_A - \mu_B$.

(b) Draw a conclusion from (a) that provides insight into whether A or B should be adopted.

9.49 Two different brands of latex paint are being considered for use. Fifteen specimens of each type of paint were selected, and the drying times, in hours, were as follows:

Paint A					Paint B				
3.5	2.7	3.9	4.2	3.6	4.7	3.9	4.5	5.5	4.0
2.7	3.3	5.2	4.2	2.9	5.3	4.3	6.0	5.2	3.7
4.4	5.2	4.0	4.1	3.4	5.5	6.2	5.1	5.4	4.8

Assume the drying time is normally distributed with $\sigma_A = \sigma_B$. Find a 95% confidence interval on $\mu_B - \mu_A$, where μ_A and μ_B are the mean drying times.

9.50 Two levels (low and high) of insulin doses are given to two groups of diabetic rats to check the insulin-binding capacity, yielding the following data:

Low dose:	$n_1 = 8$	$\bar{x}_1 = 1.98$	$s_1 = 0.51$
High dose:	$n_2 = 13$	$\bar{x}_2 = 1.30$	$s_2 = 0.35$

Assume that the variances are equal. Give a 95% confidence interval for the difference in the true average insulin-binding capacity between the two samples.

9.10 Single Sample: Estimating a Proportion

A point estimator of the proportion p in a binomial experiment is given by the statistic $\widehat{P} = X/n$, where X represents the number of successes in n trials. Therefore, the sample proportion $\hat{p} = x/n$ will be used as the point estimate of the parameter p.

If the unknown proportion p is not expected to be too close to 0 or 1, we can establish a confidence interval for p by considering the sampling distribution of $\widehat{P}$. Designating a failure in each binomial trial by the value 0 and a success by the value 1, the number of successes, x, can be interpreted as the sum of n values consisting only of 0 and 1s, and $\hat{p}$ is just the sample mean of these n values. Hence, by the Central Limit Theorem, for n sufficiently large, $\widehat{P}$ is approximately normally distributed with mean

$$\mu_{\widehat{P}} = E(\widehat{P}) = E\left(\frac{X}{n}\right) = \frac{np}{n} = p$$

and variance

$$\sigma^2_{\widehat{P}} = \sigma^2_{X/n} = \frac{\sigma^2_X}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}.$$

Therefore, we can assert that

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha, \text{ with } Z = \frac{\widehat{P} - p}{\sqrt{pq/n}},$$

and $z_{\alpha/2}$ is the value above which we find an area of $\alpha/2$ under the standard normal curve. Substituting for Z, we write

$$P\left(-z_{\alpha/2} < \frac{\widehat{P} - p}{\sqrt{pq/n}} < z_{\alpha/2}\right) = 1 - \alpha.$$

When n is large, very little error is introduced by substituting the point estimate $\hat{p} = x/n$ for the p under the radical sign. Then we can write

$$P\left(\widehat{P} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \widehat{P} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \approx 1 - \alpha.$$

On the other hand, by solving for p in the quadratic inequality above,

$$-z_{\alpha/2} < \frac{\widehat{P} - p}{\sqrt{pq/n}} < z_{\alpha/2},$$

we obtain another form of the confidence interval for p with limits

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} \pm \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}}.$$

For a random sample of size n, the sample proportion $\hat{p} = x/n$ is computed, and the following approximate $100(1-\alpha)\%$ confidence intervals for p can be obtained.

Large-Sample Confidence Intervals for p

If $\hat{p}$ is the proportion of successes in a random sample of size n and $\hat{q} = 1 - \hat{p}$, an approximate $100(1-\alpha)\%$ confidence interval, for the binomial parameter p is given by (method 1)

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

or by (method 2)

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} - \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}} < p < \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n}}{1 + \frac{z_{\alpha/2}^2}{n}} + \frac{z_{\alpha/2}}{1 + \frac{z_{\alpha/2}^2}{n}} \sqrt{\frac{\hat{p}\hat{q}}{n} + \frac{z_{\alpha/2}^2}{4n^2}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

When n is small and the unknown proportion p is believed to be close to 0 or to 1, the confidence-interval procedure established here is unreliable and, therefore, should not be used. To be on the safe side, one should require both $n\hat{p}$ and $n\hat{q}$ to be greater than or equal to 5. The methods for finding a confidence interval for the binomial parameter p are also applicable when the binomial distribution is being used to approximate the hypergeometric distribution, that is, when n is small relative to N, as illustrated by Example 9.14.

Note that although method 2 yields more accurate results, it is more complicated to calculate, and the gain in accuracy that it provides diminishes when the sample size is large enough. Hence, method 1 is commonly used in practice.

Example 9.14: In a random sample of $n = 500$ families owning television sets in the city of Hamilton, Canada, it is found that $x = 340$ subscribe to HBO. Find a 95% confidence interval for the actual proportion of families with television sets in this city that subscribe to HBO.

***Solution*:** The point estimate of p is $\hat{p} = 340/500 = 0.68$. Using Table A.3, we find that $z_{0.025} = 1.96$. Therefore, using method 1, the 95% confidence interval for p is

$$0.68 - 1.96\sqrt{\frac{(0.68)(0.32)}{500}} < p < 0.68 + 1.96\sqrt{\frac{(0.68)(0.32)}{500}},$$

which simplifies to $0.6391 < p < 0.7209$.

If we use method 2, we can obtain

$$\frac{0.68 + \frac{1.96^2}{(2)(500)}}{1 + \frac{1.96^2}{500}} \pm \frac{1.96}{1 + \frac{1.96^2}{500}} \sqrt{\frac{(0.68)(0.32)}{500} + \frac{1.96^2}{(4)(500^2)}} = 0.6786 \pm 0.0408,$$

which simplifies to $0.6378 < p < 0.7194$. Apparently, when n is large (500 here), both methods yield very similar results.

If p is the center value of a $100(1-\alpha)\%$ confidence interval, then $\hat{p}$ estimates p without error. Most of the time, however, $\hat{p}$ will not be exactly equal to p and the point estimate will be in error. The size of this error will be the positive difference that separates p and $\hat{p}$, and we can be $100(1-\alpha)\%$ confident that this difference will not exceed $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$. We can readily see this if we draw a diagram of a typical confidence interval, as in Figure 9.6. Here we use method 1 to estimate the error.

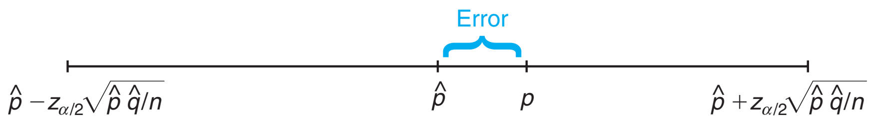

Figure 9.6: Error in estimating p by $\hat{p}$.

Theorem 9.3: If $\hat{p}$ is used as an estimate of p, we can be $100(1-\alpha)\%$ confident that the error will not exceed $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$.

In Example 9.14, we are 95% confident that the sample proportion $\hat{p} = 0.68$ differs from the true proportion p by an amount not exceeding 0.04.

Choice of Sample Size

Let us now determine how large a sample is necessary to ensure that the error in estimating p will be less than a specified amount e. By Theorem 9.3, we must choose n such that $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} = e$.

Theorem 9.4: If $\hat{p}$ is used as an estimate of p, we can be $100(1-\alpha)\%$ confident that the error will be less than a specified amount e when the sample size is approximately

$$n = \frac{z_{\alpha/2}^2 \hat{p}\hat{q}}{e^2}.$$

Theorem 9.4 is somewhat misleading in that we must use $\hat{p}$ to determine the sample size n, but $\hat{p}$ is computed from the sample. If a crude estimate of p can be made without taking a sample, this value can be used to determine n. Lacking such an estimate, we could take a preliminary sample of size $n \geq 30$ to provide an estimate of p. Using Theorem 9.4, we could determine approximately how many observations are needed to provide the desired degree of accuracy. Note that fractional values of n are rounded up to the next whole number.

Example 9.15: How large a sample is required if we want to be 95% confident that our estimate of p in Example 9.14 is within 0.02 of the true value?

***Solution*:** Let us treat the 500 families as a preliminary sample, providing an estimate $\hat{p} = 0.68$. Then, by Theorem 9.4,

$$n = \frac{(1.96)^2(0.68)(0.32)}{(0.02)^2} = 2089.8 \approx 2090.$$

Therefore, if we base our estimate of p on a random sample of size 2090, we can be 95% confident that our sample proportion will not differ from the true proportion by more than 0.02.

Occasionally, it will be impractical to obtain an estimate of p to be used for determining the sample size for a specified degree of confidence. If this happens, an upper bound for n is established by noting that $\hat{p}\hat{q} = \hat{p}(1 - \hat{p})$, which must be at most 1/4, since $\hat{p}$ must lie between 0 and 1. This fact may be verified by completing the square. Hence

$$\hat{p}(1 - \hat{p}) = -(\hat{p}^2 - \hat{p}) = \frac{1}{4} - \left(\hat{p}^2 - \hat{p} + \frac{1}{4}\right) = \frac{1}{4} - \left(\hat{p} - \frac{1}{2}\right)^2,$$

which is always less than 1/4 except when $\hat{p} = 1/2$, and then $\hat{p}\hat{q} = 1/4$. Therefore, if we substitute $\hat{p} = 1/2$ into the formula for n in Theorem 9.4 when, in fact, p actually differs from l/2, n will turn out to be larger than necessary for the specified degree of confidence; as a result, our degree of confidence will increase.

Theorem 9.5: If $\hat{p}$ is used as an estimate of p, we can be **at least** $100(1 - \alpha)\%$ confident that the error will not exceed a specified amount e when the sample size is

$$n = \frac{z_{\alpha/2}^2}{4e^2}.$$

Example 9.16: How large a sample is required if we want to be at least 95% confident that our estimate of p in Example 9.14 is within 0.02 of the true value?

***Solution*:** Unlike in Example 9.15, we shall now assume that no preliminary sample has been taken to provide an estimate of p. Consequently, we can be at least 95% confident that our sample proportion will not differ from the true proportion by more than 0.02 if we choose a sample of size

$$n = \frac{(1.96)^2}{(4)(0.02)^2} = 2401.$$

Comparing the results of Examples 9.15 and 9.16, we see that information concerning p, provided by a preliminary sample or from experience, enables us to choose a smaller sample while maintaining our required degree of accuracy.

9.11 Two Samples: Estimating the Difference between Two Proportions

Consider the problem where we wish to estimate the difference between two binomial parameters p_1 and p_2. For example, p_1 might be the proportion of smokers with lung cancer and p_2 the proportion of nonsmokers with lung cancer, and the problem is to estimate the difference between these two proportions. First, we select independent random samples of sizes n_1 and n_2 from the two binomial populations with means n_1p_1 and n_2p_2 and variances $n_1p_1q_1$ and $n_2p_2q_2$, respectively; then we determine the numbers x_1 and x_2 of people in each sample with lung cancer and form the proportions $\hat{p}_1 = x_1/n$ and $\hat{p}_2 = x_2/n$. A point estimator of the difference between the two proportions, $p_1 - p_2$, is given by the statistic $\widehat{P}_1 - \widehat{P}_2$. Therefore, the difference of the sample proportions, $\hat{p}_1 - \hat{p}_2$, will be used as the point estimate of $p_1 - p_2$.

A confidence interval for $p_1 - p_2$ can be established by considering the sampling distribution of $\widehat{P}_1 - \widehat{P}_2$. From Section 9.10 we know that $\widehat{P}_1$ and $\widehat{P}_2$ are each approximately normally distributed, with means p_1 and p_2 and variances p_1q_1/n_1 and p_2q_2/n_2, respectively. Choosing independent samples from the two populations ensures that the variables $\widehat{P}_1$ and $\widehat{P}_2$ will be independent, and then by the reproductive property of the normal distribution established in Theorem 7.11, we conclude that $\widehat{P}_1 - \widehat{P}_2$ is approximately normally distributed with mean

$$\mu_{\widehat{P}_1-\widehat{P}_2} = p_1 - p_2$$

and variance

$$\sigma^2_{\widehat{P}_1-\widehat{P}_2} = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}.$$

Therefore, we can assert that

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha,$$

where

$$Z = \frac{(\widehat{P}_1 - \widehat{P}_2) - (p_1 - p_2)}{\sqrt{p_1q_1/n_1 + p_2q_2/n_2}}$$

and $z_{\alpha/2}$ is the value above which we find an area of $\alpha/2$ under the standard normal curve. Substituting for Z, we write

$$P\left[-z_{\alpha/2} < \frac{(\widehat{P}_1 - \widehat{P}_2) - (p_1 - p_2)}{\sqrt{p_1q_1/n_1 + p_2q_2/n_2}} < z_{\alpha/2}\right] = 1 - \alpha.$$

After performing the usual mathematical manipulations, we replace p_1, p_2, q_1, and q_2 under the radical sign by their estimates $\hat{p}_1 = x_1/n_1$, $\hat{p}_2 = x_2/n_2$, $\hat{q}_1 = 1 - \hat{p}_1$, and $\hat{q}_2 = 1 - \hat{p}_2$, provided that $n_1\hat{p}_1$, $n_1\hat{q}_1$, $n_2\hat{p}_2$, and $n_2\hat{q}_2$ are all greater than or equal to 5, and the following approximate $100(1-\alpha)\%$ confidence interval for $p_1 - p_2$ is obtained.

Large-Sample Confidence Interval for $p_1 - p_2$

If $\hat{p}_1$ and $\hat{p}_2$ are the proportions of successes in random samples of sizes n_1 and n_2, respectively, $\hat{q}_1 = 1 - \hat{p}_1$, and $\hat{q}_2 = 1 - \hat{p}_2$, an approximate $100(1-\alpha)\%$ confidence interval for the difference of two binomial parameters, $p_1 - p_2$, is given by

$$(\hat{p}_1 - \hat{p}_2) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} < p_1 - p_2 < (\hat{p}_1 - \hat{p}_2) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}},$$

where $z_{\alpha/2}$ is the z-value leaving an area of $\alpha/2$ to the right.

Example 9.17: A certain change in a process for manufacturing component parts is being considered. Samples are taken under both the existing and the new process so as to determine if the new process results in an improvement. If 75 of 1500 items from the existing process are found to be defective and 80 of 2000 items from the new process are found to be defective, find a 90% confidence interval for the true difference in the proportion of defectives between the existing and the new process.

Solution: Let p_1 and p_2 be the true proportions of defectives for the existing and new processes, respectively. Hence, $\hat{p}_1 = 75/1500 = 0.05$ and $\hat{p}_2 = 80/2000 = 0.04$, and the point estimate of $p_1 - p_2$ is

$$\hat{p}_1 - \hat{p}_2 = 0.05 - 0.04 = 0.01.$$

Using Table A.3, we find $z_{0.05} = 1.645$. Therefore, substituting into the formula, with

$$1.645\sqrt{\frac{(0.05)(0.95)}{1500} + \frac{(0.04)(0.96)}{2000}} = 0.0117,$$

we find the 90% confidence interval to be $-0.0017 < p_1 - p_2 < 0.0217$. Since the interval contains the value 0, there is no reason to believe that the new process produces a significant decrease in the proportion of defectives over the existing method.

Up to this point, all confidence intervals presented were of the form

$$\text{point estimate} \ \pm \ K \ \text{s.e.(point estimate)},$$

where K is a constant (either t or normal percent point). This form is valid when the parameter is a mean, a difference between means, a proportion, or a difference between proportions, due to the symmetry of the t- and Z-distributions. However, it does not extend to variances and ratios of variances, which will be discussed in Sections 9.12 and 9.13.

Exercises

In this set of exercises, for estimation concerning one proportion, use only method 1 to obtain the confidence intervals, unless instructed otherwise.

9.51 In a random sample of 1000 homes in a certain city, it is found that 228 are heated by oil. Find 99% confidence intervals for the proportion of homes in this city that are heated by oil using both methods presented on page 297.

9.52 Compute 95% confidence intervals, using both methods on page 297, for the proportion of defective items in a process when it is found that a sample of size 100 yields 8 defectives.

9.53 (a) A random sample of 200 voters in a town is selected, and 114 are found to support an annexation suit. Find the 96% confidence interval for the fraction of the voting population favoring the suit.

(b) What can we assert with 96% confidence about the possible size of our error if we estimate the fraction of voters favoring the annexation suit to be 0.57?

9.54 A manufacturer of MP3 players conducts a set of comprehensive tests on the electrical functions of its product. All MP3 players must pass all tests prior to being sold. Of a random sample of 500 MP3 players, 15 failed one or more tests. Find a 90% confidence interval for the proportion of MP3 players from the population that pass all tests.

9.55 A new rocket-launching system is being considered for deployment of small, short-range rockets. The existing system has $p = 0.8$ as the probability of a successful launch. A sample of 40 experimental launches is made with the new system, and 34 are successful.

(a) Construct a 95% confidence interval for p.

(b) Would you conclude that the new system is better?

9.56 A geneticist is interested in the proportion of African males who have a certain minor blood disorder. In a random sample of 100 African males, 24 are found to be afflicted.

(a) Compute a 99% confidence interval for the proportion of African males who have this blood disorder.

(b) What can we assert with 99% confidence about the possible size of our error if we estimate the proportion of African males with this blood disorder to be 0.24?

9.57 (a) According to a report in the *Roanoke Times & World-News*, approximately 2/3 of 1600 adults polled by telephone said they think the space shuttle program is a good investment for the country. Find a 95% confidence interval for the proportion of American adults who think the space shuttle program is a good investment for the country.

(b) What can we assert with 95% confidence about the possible size of our error if we estimate the proportion of American adults who think the space shuttle program is a good investment to be 2/3?

9.58 In the newspaper article referred to in Exercise 9.57, 32% of the 1600 adults polled said the U.S. space program should emphasize scientific exploration. How large a sample of adults is needed for the poll if one wishes to be 95% confident that the estimated percentage will be within 2% of the true percentage?

9.59 How large a sample is needed if we wish to be 96% confident that our sample proportion in Exercise 9.53 will be within 0.02 of the true fraction of the voting population?

9.60 How large a sample is needed if we wish to be 99% confident that our sample proportion in Exercise 9.51 will be within 0.05 of the true proportion of homes in the city that are heated by oil?

9.61 How large a sample is needed in Exercise 9.52 if we wish to be 98% confident that our sample proportion will be within 0.05 of the true proportion defective?

9.62 A conjecture by a faculty member in the microbiology department at Washington University School of Dental Medicine in St. Louis, Missouri, states that a couple of cups of either green or oolong tea each day will provide sufficient fluoride to protect your teeth from decay. How large a sample is needed to estimate the percentage of citizens in a certain town who favor having their water fluoridated if one wishes to be at least 99% confident that the estimate is within 1% of the true percentage?

9.63 A study is to be made to estimate the percentage of citizens in a town who favor having their water fluoridated. How large a sample is needed if one wishes to be at least 95% confident that the estimate is within 1% of the true percentage?

9.64 A study is to be made to estimate the proportion of residents of a certain city and its suburbs who favor the construction of a nuclear power plant near the city. How large a sample is needed if one wishes to be at least 95% confident that the estimate is within 0.04 of the true proportion of residents who favor the construction of the nuclear power plant?

9.65 A certain geneticist is interested in the proportion of males and females in the population who have a minor blood disorder. In a random sample of 1000 males, 250 are found to be afflicted, whereas 275 of 1000 females tested appear to have the disorder. Compute a 95% confidence interval for the difference between the proportions of males and females who have the blood disorder.

9.66 Ten engineering schools in the United States were surveyed. The sample contained 250 electrical engineers, 80 being women; 175 chemical engineers, 40 being women. Compute a 90% confidence interval for the difference between the proportions of women in these two fields of engineering. Is there a significant difference between the two proportions?

9.67 A clinical trial was conducted to determine if a certain type of inoculation has an effect on the incidence of a certain disease. A sample of 1000 rats was kept in a controlled environment for a period of 1 year, and 500 of the rats were given the inoculation. In the group not inoculated, there were 120 incidences of the disease, while 98 of the rats in the inoculated group contracted it. If p_1 is the probability of incidence of the disease in uninoculated rats and p_2 the probability of incidence in inoculated rats, compute a 90% confidence interval for $p_1 - p_2$.

9.68 In the study *Germination and Emergence of Broccoli*, conducted by the Department of Horticulture at Virginia Tech, a researcher found that at 5°C, 10 broccoli seeds out of 20 germinated, while at 15°C, 15 out of 20 germinated. Compute a 95% confidence interval for the difference between the proportions of germination at the two different temperatures and decide if there is a significant difference.

9.69 A survey of 1000 students found that 274 chose professional baseball team A as their favorite team. In a similar survey involving 760 students, 240 of them chose team A as their favorite. Compute a 95% confidence interval for the difference between the proportions of students favoring team A in the two surveys. Is there a significant difference?

9.70 According to *USA Today* (March 17, 1997), women made up 33.7% of the editorial staff at local TV stations in the United States in 1990 and 36.2% in 1994. Assume 20 new employees were hired as editorial staff.

(a) Estimate the number that would have been women in 1990 and 1994, respectively.

(b) Compute a 95% confidence interval to see if there is evidence that the proportion of women hired as editorial staff was higher in 1994 than in 1990.

9.12 Single Sample: Estimating the Variance

If a sample of size n is drawn from a normal population with variance σ^2 and the sample variance s^2 is computed, we obtain a value of the statistic S^2. This computed sample variance is used as a point estimate of σ^2. Hence, the statistic S^2 is called an estimator of σ^2.

An interval estimate of σ^2 can be established by using the statistic

$$X^2 = \frac{(n-1)S^2}{\sigma^2}.$$

According to Theorem 8.4, the statistic X^2 has a chi-squared distribution with $n-1$ degrees of freedom when samples are chosen from a normal population. We may write (see Figure 9.7)

$$P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha,$$

where $\chi^2_{1-\alpha/2}$ and $\chi^2_{\alpha/2}$ are values of the chi-squared distribution with $n-1$ degrees of freedom, leaving areas of $1-\alpha/2$ and $\alpha/2$, respectively, to the right. Substituting for X^2, we write

$$P\left[\chi^2_{1-\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2}\right] = 1 - \alpha.$$

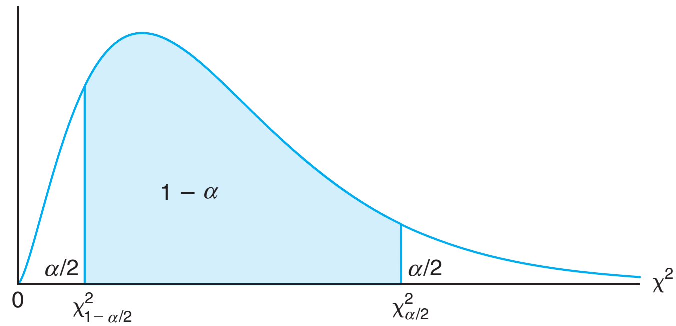

Figure 9.7: $P(\chi^2_{1-\alpha/2} < X^2 < \chi^2_{\alpha/2}) = 1 - \alpha$.

Dividing each term in the inequality by $(n-1)S^2$ and then inverting each term (thereby changing the sense of the inequalities), we obtain

$$P\left[\frac{(n-1)S^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\right] = 1 - \alpha.$$

For a random sample of size n from a normal population, the sample variance s^2 is computed, and the following $100(1-\alpha)\%$ confidence interval for σ^2 is obtained.

Confidence Interval for σ^2

If s^2 is the variance of a random sample of size n from a normal population, a $100(1-\alpha)\%$ confidence interval for σ^2 is

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

where $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are χ^2-values with $v = n-1$ degrees of freedom, leaving areas of $\alpha/2$ and $1-\alpha/2$, respectively, to the right.

An approximate $100(1-\alpha)\%$ confidence interval for σ is obtained by taking the square root of each endpoint of the interval for σ^2.

Example 9.18: The following are the weights, in decagrams, of 10 packages of grass seed distributed by a certain company: 46.4, 46.1, 45.8, 47.0, 46.1, 45.9, 45.8, 46.9, 45.2, and 46.0. Find a 95% confidence interval for the variance of the weights of all such packages of grass seed distributed by this company, assuming a normal population.

Solution: First we find

$$s^2 = \frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)} = \frac{(10)(21,273.12) - (461.2)^2}{(10)(9)} = 0.286.$$

To obtain a 95% confidence interval, we choose $\alpha = 0.05$. Then, using Table A.5 with $v = 9$ degrees of freedom, we find $\chi^2_{0.025} = 19.023$ and $\chi^2_{0.975} = 2.700$. Therefore, the 95% confidence interval for σ^2 is

$$\frac{(9)(0.286)}{19.023} < \sigma^2 < \frac{(9)(0.286)}{2.700},$$

or simply $0.135 < \sigma^2 < 0.953$.

9.13 Two Samples: Estimating the Ratio of Two Variances

A point estimate of the ratio of two population variances σ_1^2/σ_2^2 is given by the ratio s_1^2/s_2^2 of the sample variances. Hence, the statistic S_1^2/S_2^2 is called an estimator of σ_1^2/σ_2^2.

If σ_1^2 and σ_2^2 are the variances of normal populations, we can establish an interval estimate of σ_1^2/σ_2^2 by using the statistic

$$F = \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2}.$$

According to Theorem 8.8, the random variable F has an F-distribution with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom. Therefore, we may write (see Figure 9.8)

$$P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha,$$

where $f_{1-\alpha/2}(v_1, v_2)$ and $f_{\alpha/2}(v_1, v_2)$ are the values of the F-distribution with v_1 and v_2 degrees of freedom, leaving areas of $1 - \alpha/2$ and $\alpha/2$, respectively, to the right.

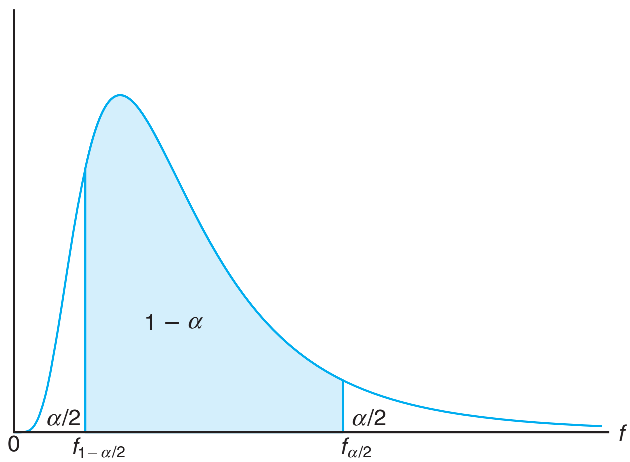

Figure 9.8: $P[f_{1-\alpha/2}(v_1, v_2) < F < f_{\alpha/2}(v_1, v_2)] = 1 - \alpha$.

Substituting for F, we write

$$P\left[f_{1-\alpha/2}(v_1, v_2) < \frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} < f_{\alpha/2}(v_1, v_2)\right] = 1 - \alpha.$$

Multiplying each term in the inequality by S_2^2/S_1^2 and then inverting each term, we obtain

$$P\left[\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{f_{1-\alpha/2}(v_1, v_2)}\right] = 1 - \alpha.$$

The results of Theorem 8.7 enable us to replace the quantity $f_{1-\alpha/2}(v_1, v_2)$ by $1/f_{\alpha/2}(v_2, v_1)$. Therefore,

$$P\left[\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} f_{\alpha/2}(v_2, v_1)\right] = 1 - \alpha.$$

For any two independent random samples of sizes n_1 and n_2 selected from two normal populations, the ratio of the sample variances s_1^2/s_2^2 is computed, and the following $100(1-\alpha)\%$ confidence interval for σ_1^2/σ_2^2 is obtained.

Confidence Interval for σ_1^2/σ_2^2

If s_1^2 and s_2^2 are the variances of independent samples of sizes n_1 and n_2, respectively, from normal populations, then a $100(1-\alpha)\%$ confidence interval for σ_1^2/σ_2^2 is

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(v_2, v_1),$$

where $f_{\alpha/2}(v_1, v_2)$ is an f-value with $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$ degrees of freedom, leaving an area of $\alpha/2$ to the right, and $f_{\alpha/2}(v_2, v_1)$ is a similar f-value with $v_2 = n_2 - 1$ and $v_1 = n_1 - 1$ degrees of freedom.

As in Section 9.12, an approximate $100(1-\alpha)\%$ confidence interval for σ_1/σ_2 is obtained by taking the square root of each endpoint of the interval for σ_1^2/σ_2^2.

Example 9.19: A confidence interval for the difference in the mean orthophosphorus contents, measured in milligrams per liter, at two stations on the James River was constructed in Example 9.12 on page 290 by assuming the normal population variance to be unequal. Justify this assumption by constructing 98% confidence intervals for σ_1^2/σ_2^2 and for σ_1/σ_2, where σ_1^2 and σ_2^2 are the variances of the populations of orthophosphorus contents at station 1 and station 2, respectively.

Solution: From Example 9.12, we have $n_1 = 15$, $n_2 = 12$, $s_1 = 3.07$, and $s_2 = 0.80$. For a 98% confidence interval, $\alpha = 0.02$. Interpolating in Table A.6, we find $f_{0.01}(14, 11) \approx 4.30$ and $f_{0.01}(11, 14) \approx 3.87$. Therefore, the 98% confidence interval for σ_1^2/σ_2^2 is

$$\left(\frac{3.07^2}{0.80^2}\right)\left(\frac{1}{4.30}\right) < \frac{\sigma_1^2}{\sigma_2^2} < \left(\frac{3.07^2}{0.80^2}\right)(3.87),$$

which simplifies to $3.425 < \frac{\sigma_1^2}{\sigma_2^2} < 56.991$. Taking square roots of the confidence limits, we find that a 98% confidence interval for σ_1/σ_2 is

$$1.851 < \frac{\sigma_1}{\sigma_2} < 7.549.$$

Since this interval does not allow for the possibility of σ_1/σ_2 being equal to 1, we were correct in assuming that $\sigma_1 \neq \sigma_2$ or $\sigma_1^2 \neq \sigma_2^2$ in Example 9.12.

Exercises

9.71 A manufacturer of car batteries claims that the batteries will last, on average, 3 years with a variance of 1 year. If 5 of these batteries have lifetimes of 1.9, 2.4, 3.0, 3.5, and 4.2 years, construct a 95% confidence interval for σ^2 and decide if the manufacturer's claim that $\sigma^2 = 1$ is valid. Assume the population of battery lives to be approximately normally distributed.

9.72 A random sample of 20 students yielded a mean of $\bar{x} = 72$ and a variance of $s^2 = 16$ for scores on a college placement test in mathematics. Assuming the scores to be normally distributed, construct a 98% confidence interval for σ^2.

9.73 Construct a 95% confidence interval for σ^2 in Exercise 9.9 on page 283.

9.74 Construct a 99% confidence interval for σ^2 in Exercise 9.11 on page 283.

9.75 Construct a 99% confidence interval for σ in Exercise 9.12 on page 283.

9.76 Construct a 90% confidence interval for σ in Exercise 9.13 on page 283.

9.77 Construct a 98% confidence interval for σ_1/σ_2 in Exercise 9.42 on page 295, where σ_1 and σ_2 are, respectively, the standard deviations for the distances traveled per liter of fuel by the Volkswagen and Toyota mini-trucks.

9.78 Construct a 90% confidence interval for σ_1^2/σ_2^2 in Exercise 9.43 on page 295. Were we justified in assuming that $\sigma_1^2 \neq \sigma_2^2$ when we constructed the confidence interval for $\mu_1 - \mu_2$?

9.79 Construct a 90% confidence interval for σ_1^2/σ_2^2 in Exercise 9.46 on page 295. Should we have assumed $\sigma_1^2 = \sigma_2^2$ in constructing our confidence interval for $\mu_I - \mu_{II}$?

9.80 Construct a 95% confidence interval for σ_A^2/σ_B^2 in Exercise 9.49 on page 295. Should the equal-variance assumption be used?

9.14 Maximum Likelihood Estimation (Optional)

Often the estimators of parameters have been those that appeal to intuition. The estimator $\bar{X}$ certainly seems reasonable as an estimator of a population mean μ. The virtue of S^2 as an estimator of σ^2 is underscored through the discussion of unbiasedness in Section 9.3. The estimator for a binomial parameter p is merely a sample proportion, which, of course, is an *average* and appeals to common sense. But there are many situations in which it is not at all obvious what the proper estimator should be. As a result, there is much to be learned by the student of statistics concerning different philosophies that produce different methods of estimation. In this section, we deal with the **method of maximum likelihood**.

Maximum likelihood estimation is one of the most important approaches to estimation in all of statistical inference. We will not give a thorough development of the method. Rather, we will attempt to communicate the philosophy of maximum likelihood and illustrate with examples that relate to other estimation problems discussed in this chapter.

The Likelihood Function

As the name implies, the method of maximum likelihood is that for which the *likelihood function* is maximized. The likelihood function is best illustrated through the use of an example with a discrete distribution and a single parameter. Denote by $X_1, X_2, \ldots, X_n$ the independent random variables taken from a discrete probability distribution represented by $f(\mathbf{x}, \theta)$, where θ is a single parameter of the distribution. Now

$$\begin{aligned} L(x_1, x_2, \ldots, x_n; \theta) &= f(x_1, x_2, \ldots, x_n; \theta) \\ &= f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta) \end{aligned}$$

is the *joint distribution of the random variables*, often referred to as the likelihood function. Note that the variable of the likelihood function is θ, not $\mathbf{x}$. Denote by $x_1, x_2, \ldots, x_n$ the observed values in a sample. In the case of a discrete random variable, the interpretation is very clear. The quantity $L(x_1, x_2, \ldots, x_n; \theta)$, *the likelihood of the sample*, is the following joint probability:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n \mid \theta),$$

which is the probability of obtaining the sample values $x_1, x_2, \ldots, x_n$. For the discrete case, the maximum likelihood estimator is one that results in a maximum value for this joint probability or maximizes the likelihood of the sample.

Consider a fictitious example where three items from an assembly line are inspected. The items are ruled either defective or nondefective, and thus the Bernoulli process applies. Testing the three items results in two nondefective items followed by a defective item. It is of interest to estimate p, the proportion nondefective in the process. The likelihood of the sample for this illustration is given by

$$p \cdot p \cdot q = p^2 q = p^2 - p^3,$$

where $q = 1 - p$. Maximum likelihood estimation would give an estimate of p for which the likelihood is maximized. It is clear that if we differentiate the likelihood with respect to p, set the derivative to zero, and solve, we obtain the value

$$\hat{p} = \frac{2}{3}.$$

Now, of course, in this situation $\hat{p} = 2/3$ is the sample proportion defective and is thus a reasonable estimator of the probability of a defective. The reader should attempt to understand that the philosophy of maximum likelihood estimation evolves from the notion that the reasonable estimator of a parameter based on sample information *is that parameter value that produces the largest probability of obtaining the sample*. This is, indeed, the interpretation for the discrete case, since the likelihood is the probability of jointly observing the values in the sample.

Now, while the interpretation of the likelihood function as a joint probability is confined to the discrete case, the notion of maximum likelihood extends to the estimation of parameters of a continuous distribution. We now present a formal definition of maximum likelihood estimation.

Definition 9.3: Given independent observations $x_1, x_2, \ldots, x_n$ from a probability density function (continuous case) or probability mass function (discrete case) $f(\mathbf{x}; \theta)$, the maximum likelihood estimator $\hat{\theta}$ is that which maximizes the likelihood function

$$L(x_1, x_2, \ldots, x_n; \theta) = f(\mathbf{x}; \theta) = f(x_1, \theta) f(x_2, \theta) \cdots f(x_n, \theta).$$

Quite often it is convenient to work with the natural log of the likelihood function in finding the maximum of that function. Consider the following example dealing with the parameter μ of a Poisson distribution.

Example 9.20: Consider a Poisson distribution with probability mass function

$$f(x|\mu) = \frac{e^{-\mu}\mu^x}{x!}, \quad x = 0, 1, 2, \ldots.$$

Suppose that a random sample $x_1, x_2, \ldots, x_n$ is taken from the distribution. What is the maximum likelihood estimate of μ?

***Solution*:** The likelihood function is

$$L(x_1, x_2, \ldots, x_n; \mu) = \prod_{i=1}^{n} f(x_i|\mu) = \frac{e^{-n\mu}\mu^{\sum\limits_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}.$$

Now consider

$$\ln L(x_1, x_2, \ldots, x_n; \mu) = -n\mu + \sum_{i=1}^{n} x_i \ln \mu - \ln \prod_{i=1}^{n} x_i!$$

$$\frac{\partial \ln L(x_1, x_2, \ldots, x_n; \mu)}{\partial \mu} = -n + \sum_{i=1}^{n} \frac{x_i}{\mu}.$$

Solving for $\hat{\mu}$, the maximum likelihood estimator, involves setting the derivative to zero and solving for the parameter. Thus,

$$\hat{\mu} = \sum_{i=1}^{n} \frac{x_i}{n} = \bar{x}.$$

The second derivative of the log-likelihood function is negative, which implies that the solution above indeed is a maximum. Since μ is the mean of the Poisson distribution (Chapter 5), the sample average would certainly seem like a reasonable estimator.

The following example shows the use of the method of maximum likelihood for finding estimates of two parameters. We simply find the values of the parameters that maximize (jointly) the likelihood function.

Example 9.21: Consider a random sample $x_1, x_2, \ldots, x_n$ from a normal distribution $N(\mu, \sigma)$. Find the maximum likelihood estimators for μ and σ^2.

Solution: The likelihood function for the normal distribution is

$$L(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = \frac{1}{(2\pi)^{n/2}(\sigma^2)^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2\right].$$

Taking logarithms gives us

$$\ln L(x_1, x_2, \ldots, x_n; \mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^2.$$

Hence,

$$\frac{\partial \ln L}{\partial \mu} = \sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma^2}\right)$$

and

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

Setting both derivatives equal to 0, we obtain

$$\sum_{i=1}^{n} x_i - n\mu = 0 \quad \text{and} \quad n\sigma^2 = \sum_{i=1}^{n}(x_i - \mu)^2.$$

Thus, the maximum likelihood estimator of μ is given by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$$

which is a pleasing result since $\bar{x}$ has played such an important role in this chapter as a point estimate of μ. On the other hand, the maximum likelihood estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Checking the second-order partial derivative matrix confirms that the solution results in a maximum of the likelihood function.

It is interesting to note the distinction between the maximum likelihood estimator of σ^2 and the unbiased estimator S^2 developed earlier in this chapter. The numerators are identical, of course, and the denominator is the degrees of freedom $n-1$ for the unbiased estimator and n for the maximum likelihood estimator. Maximum likelihood estimators do not necessarily enjoy the property of unbiasedness. However, they do have very important asymptotic properties.

Example 9.22: Suppose 10 rats are used in a biomedical study where they are injected with cancer cells and then given a cancer drug that is designed to increase their survival rate. The survival times, in months, are 14, 17, 27, 18, 12, 8, 22, 13, 19, and 12. Assume that the exponential distribution applies. Give a maximum likelihood estimate of the mean survival time.

***Solution*:** From Chapter 6, we know that the probability density function for the exponential random variable X is

$$f(x, \beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Thus, the log-likelihood function for the data, given $n = 10$, is

$$\ln L(x_1, x_2, \ldots, x_{10}; \beta) = -10 \ln \beta - \frac{1}{\beta} \sum_{i=1}^{10} x_i.$$

Setting

$$\frac{\partial \ln L}{\partial \beta} = -\frac{10}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{10} x_i = 0$$

implies that

$$\hat{\beta} = \frac{1}{10} \sum_{i=1}^{10} x_i = \bar{x} = 16.2.$$

Evaluating the second derivative of the log-likelihood function at the value $\hat{\beta}$ above yields a negative value. As a result, the estimator of the parameter β, the population mean, is the sample average $\bar{x}$.

The following example shows the maximum likelihood estimator for a distribution that does not appear in previous chapters.

Example 9.23: It is known that a sample consisting of the values 12, 11.2, 13.5, 12.3, 13.8, and 11.9 comes from a population with the density function

$$f(x; \theta) = \begin{cases} \frac{\theta}{x^{\theta+1}}, & x > 1, \\ 0, & \text{elsewhere,} \end{cases}$$

where $\theta > 0$. Find the maximum likelihood estimate of θ.

***Solution*:** The likelihood function of n observations from this population can be written as

$$L(x_1, x_2, \ldots, x_{10}; \theta) = \prod_{i=1}^{n} \frac{\theta}{\boldsymbol{x}_i^{\theta+1}} = \frac{\theta^n}{(\prod_{i=1}^{n} x_i)^{\theta+1}},$$

which implies that

$$\ln L(x_1, x_2, \ldots, x_{10}; \theta) = n \ln(\theta) - (\theta + 1) \sum_{i=1}^{n} \ln(x_i).$$

Setting $0 = \frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} \ln(x_i)$ results in

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln(x_i)}$$
$$= \frac{6}{\ln(12) + \ln(11.2) + \ln(13.5) + \ln(12.3) + \ln(13.8) + \ln(11.9)} = 0.3970.$$

Since the second derivative of L is $-n/\theta^2$, which is always negative, the likelihood function does achieve its maximum value at $\hat{\theta}$.

Additional Comments Concerning Maximum Likelihood Estimation

A thorough discussion of the properties of maximum likelihood estimation is beyond the scope of this book and is usually a major topic of a course in the theory of statistical inference. The method of maximum likelihood allows the analyst to make use of knowledge of the distribution in determining an appropriate estimator. *The method of maximum likelihood cannot be applied without knowledge of the underlying distribution.* We learned in Example 9.21 that the maximum likelihood estimator is not necessarily unbiased. The maximum likelihood estimator is unbiased *asymptotically* or *in the limit*; that is, the amount of bias approaches zero as the sample size becomes large. Earlier in this chapter the notion of efficiency was discussed, efficiency being linked to the variance property of an estimator. Maximum likelihood estimators possess desirable variance properties in the limit. The reader should consult Lehmann and D'Abrera (1998) for details.

Exercises

9.81 Suppose that there are n trials $x_1, x_2, \ldots, x_n$ from a Bernoulli process with parameter p, the probability of a success. That is, the probability of r successes is given by $\binom{n}{r} p^r (1-p)^{n-r}$. Work out the maximum likelihood estimator for the parameter p.

9.82 Consider the lognormal distribution with the density function given in Section 6.9. Suppose we have a random sample $x_1, x_2, \ldots, x_n$ from a lognormal distribution.

(a) Write out the likelihood function.

(b) Develop the maximum likelihood estimators of μ and σ^2.

9.83 Consider a random sample of $x_1, \ldots, x_n$ coming from the gamma distribution discussed in Section 6.6. Suppose the parameter α is known, say 5, and determine the maximum likelihood estimation for parameter β.

9.84 Consider a random sample of $x_1, x_2, \ldots, x_n$ observations from a Weibull distribution with parameters α and β and density function

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}, & x > 0, \\ 0, & \text{elsewhere}, \end{cases}$$

for $\alpha, \beta > 0$.

(a) Write out the likelihood function.

(b) Write out the equations that, when solved, give the maximum likelihood estimators of α and β.

9.85 Consider a random sample of $x_1, \ldots, x_n$ from a uniform distribution $U(0, \theta)$ with unknown parameter θ, where $\theta > 0$. Determine the maximum likelihood estimator of θ.

9.86 Consider the independent observations $x_1, x_2, \ldots, x_n$ from the gamma distribution discussed in Section 6.6.

(a) Write out the likelihood function.

(b) Write out a set of equations that, when solved, give the maximum likelihood estimators of α and β.

9.87 Consider a hypothetical experiment where a man with a fungus uses an antifungal drug and is cured. Consider this, then, a sample of one from a Bernoulli distribution with probability function

$$f(x) = p^x q^{1-x}, \quad x = 0, 1,$$

where p is the probability of a success (cure) and $q = 1 - p$. Now, of course, the sample information gives $x = 1$. Write out a development that shows that $\hat{p} = 1.0$ is the maximum likelihood estimator of the probability of a cure.

9.88 Consider the observation X from the negative binomial distribution given in Section 5.4. Find the maximum likelihood estimator for p, assuming k is known.

Review Exercises

9.89 Consider two estimators of σ^2 for a sample $x_1, x_2, \ldots, x_n$, which is drawn from a normal distribution with mean μ and variance σ^2. The estimators are the unbiased estimator $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ and the maximum likelihood estimator $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Discuss the variance properties of these two estimators.

9.90 According to the *Roanoke Times*, McDonald's sold 42.1% of the market share of hamburgers. A random sample of 75 burgers sold resulted in 28 of them being from McDonald's. Use material in Section 9.10 to determine if this information supports the claim in the *Roanoke Times*.

9.91 It is claimed that a new diet will reduce a person's weight by 4.5 kilograms on average in a period of 2 weeks. The weights of 7 women who followed this diet were recorded before and after the 2-week period.

Woman	Weight Before	Weight After
1	58.5	60.0
2	60.3	54.9
3	61.7	58.1
4	69.0	62.1
5	64.0	58.5
6	62.6	59.9
7	56.7	54.4

Test the claim about the diet by computing a 95% confidence interval for the mean difference in weights. Assume the differences of weights to be approximately normally distributed.

9.92 A study was undertaken at Virginia Tech to determine if fire can be used as a viable management tool to increase the amount of forage available to deer during the critical months in late winter and early spring. Calcium is a required element for plants and animals. The amount taken up and stored in plants is closely correlated to the amount present in the soil. It was hypothesized that a fire may change the calcium levels present in the soil and thus affect the amount available to deer. A large tract of land in the Fishburn Forest was selected for a prescribed burn. Soil samples were taken from 12 plots of equal area just prior to the burn and analyzed for calcium. Postburn calcium levels were analyzed from the same plots. These values, in kilograms per plot, are presented in the following table:

	Calcium Level (kg/plot)	
Plot	**Preburn**	**Postburn**
1	50	9
2	50	18
3	82	45
4	64	18
5	82	18
6	73	9
7	77	32
8	54	9
9	23	18
10	45	9
11	36	9
12	54	9

Construct a 95% confidence interval for the mean difference in calcium levels in the soil prior to and after the prescribed burn. Assume the distribution of differences in calcium levels to be approximately normal.

9.93 A health spa claims that a new exercise program will reduce a person's waist size by 2 centimeters on average over a 5-day period. The waist sizes, in centimeters, of 6 men who participated in this exercise program are recorded before and after the 5-day period in the following table:

Man	Waist Size Before	Waist Size After
1	90.4	91.7
2	95.5	93.9
3	98.7	97.4
4	115.9	112.8
5	104.0	101.3
6	85.6	84.0

By computing a 95% confidence interval for the mean reduction in waist size, determine whether the health spa's claim is valid. Assume the distribution of differences in waist sizes before and after the program to be approximately normal.

9.94 The Department of Civil Engineering at Virginia Tech compared a modified (M-5 hr) assay technique for recovering fecal coliforms in stormwater runoff from an urban area to a most probable number (MPN) technique. A total of 12 runoff samples were collected and analyzed by the two techniques. Fecal coliform counts per 100 milliliters are recorded in the following table.

Sample	MPN Count	M-5 hr Count
1	2300	2010
2	1200	930
3	450	400
4	210	436
5	270	4100
6	450	2090
7	154	219
8	179	169
9	192	194
10	230	174
11	340	274
12	194	183

Construct a 90% confidence interval for the difference in the mean fecal coliform counts between the M-5 hr and the MPN techniques. Assume that the count differences are approximately normally distributed.

9.95 An experiment was conducted to determine whether surface finish has an effect on the endurance limit of steel. There is a theory that polishing increases the average endurance limit (for reverse bending). From a practical point of view, polishing should not have any effect on the standard deviation of the endurance limit, which is known from numerous endurance limit experiments to be 4000 psi. An experiment was performed on 0.4% carbon steel using both unpolished and polished smooth-turned specimens. The data are as follows:

Endurance Limit (psi)	
Polished 0.4% Carbon	**Unpolished 0.4% Carbon**
85,500	82,600
91,900	82,400
89,400	81,700
84,000	79,500
89,900	79,400
78,700	69,800
87,500	79,900
83,100	83,400

Find a 95% confidence interval for the difference between the population means for the two methods, assuming that the populations are approximately normally distributed.

9.96 An anthropologist is interested in the proportion of individuals in two Indian tribes with double occipital hair whorls. Suppose that independent samples are taken from each of the two tribes, and it is found that 24 of 100 Indians from tribe A and 36 of 120 Indians from tribe B possess this characteristic. Construct a 95% confidence interval for the difference $p_B - p_A$ between the proportions of these two tribes with occipital hair whorls.

9.97 A manufacturer of electric irons produces these items in two plants. Both plants have the same suppliers of small parts. A saving can be made by purchasing thermostats for plant B from a local supplier. A single lot was purchased from the local supplier, and a test was conducted to see whether or not these new thermostats were as accurate as the old. The thermostats were tested on tile irons on the 550°F setting, and the actual temperatures were read to the nearest 0.1°F with a thermocouple. The data are as follows:

New Supplier (°F)					
530.3	559.3	549.4	544.0	551.7	566.3
549.9	556.9	536.7	558.8	538.8	543.3
559.1	555.0	538.6	551.1	565.4	554.9
550.0	554.9	554.7	536.1	569.1	
Old Supplier (°F)					
559.7	534.7	554.8	545.0	544.6	538.0
550.7	563.1	551.1	553.8	538.8	564.6
554.5	553.0	538.4	548.3	552.9	535.1
555.0	544.8	558.4	548.7	560.3	

Find 95% confidence intervals for σ_1^2/σ_2^2 and for σ_1/σ_2, where σ_1^2 and σ_2^2 are the population variances of the thermostat readings for the new and old suppliers, respectively.

9.98 It is argued that the resistance of wire A is greater than the resistance of wire B. An experiment on the wires shows the following results (in ohms):

Wire A	Wire B
0.140	0.135
0.138	0.140
0.143	0.136
0.142	0.142
0.144	0.138
0.137	0.140

Assuming equal variances, what conclusions do you draw? Justify your answer.

9.99 An alternative form of estimation is accomplished through the method of moments. This method involves equating the population mean and variance to the corresponding sample mean $\bar{x}$ and sample variance

s^2 and solving for the parameters, the results being the **moment estimators**. In the case of a single parameter, only the means are used. Give an argument that in the case of the Poisson distribution the maximum likelihood estimator and moment estimators are the same.

9.100 Specify the moment estimators for μ and σ^2 for the normal distribution.

9.101 Specify the moment estimators for μ and σ^2 for the lognormal distribution.

9.102 Specify the moment estimators for α and β for the gamma distribution.

9.103 A survey was done with the hope of comparing salaries of chemical plant managers employed in two areas of the country, the northern and west central regions. An independent random sample of 300 plant managers was selected from each of the two regions. These managers were asked their annual salaries. The results are as follows

North	**West Central**
$\bar{x}_1 = \$102,300$	$\bar{x}_2 = \$98,500$
$s_1 = \$5700$	$s_2 = \$3800$

(a) Construct a 99% confidence interval for $\mu_1 - \mu_2$, the difference in the mean salaries.

(b) What assumption did you make in (a) about the distribution of annual salaries for the two regions? Is the assumption of normality necessary? Why or why not?

(c) What assumption did you make about the two variances? Is the assumption of equality of variances reasonable? Explain!

9.104 Consider Review Exercise 9.103. Let us assume that the data have not been collected yet and that previous statistics suggest that $\sigma_1 = \sigma_2 = \$4000$. Are the sample sizes in Review Exercise 9.103 sufficient to produce a 95% confidence interval on $\mu_1 - \mu_2$ having a width of only \$1000? Show all work.

9.105 A labor union is becoming defensive about gross absenteeism by its members. The union leaders had always claimed that, in a typical month, 95% of its members were absent less than 10 hours. The union decided to check this by monitoring a random sample of 300 of its members. The number of hours absent was recorded for each of the 300 members. The results were $\bar{x} = 6.5$ hours and $s = 2.5$ hours. Use the data to respond to this claim, using a one-sided tolerance limit and choosing the confidence level to be 99%. Be sure to interpret what you learn from the tolerance limit calculation.

9.106 A random sample of 30 firms dealing in wireless products was selected to determine the proportion of such firms that have implemented new software to improve productivity. It turned out that 8 of the 30 had implemented such software. Find a 95% confidence interval on p, the true proportion of such firms that have implemented new software.

9.107 Refer to Review Exercise 9.106. Suppose there is concern about whether the point estimate $\hat{p} = 8/30$ is accurate enough because the confidence interval around p is not sufficiently narrow. Using $\hat{p}$ as the estimate of p, how many companies would need to be sampled in order to have a 95% confidence interval with a width of only 0.05?

9.108 A manufacturer turns out a product item that is labeled either "defective" or "not defective." In order to estimate the proportion defective, a random sample of 100 items is taken from production, and 10 are found to be defective. Following implementation of a quality improvement program, the experiment is conducted again. A new sample of 100 is taken, and this time only 6 are found to be defective.

(a) Give a 95% confidence interval on $p_1 - p_2$, where p_1 is the population proportion defective before improvement and p_2 is the proportion defective after improvement.

(b) Is there information in the confidence interval found in (a) that would suggest that $p_1 > p_2$? Explain.

9.109 A machine is used to fill boxes with product in an assembly line operation. Much concern centers around the variability in the number of ounces of product in a box. The standard deviation in weight of product is known to be 0.3 ounce. An improvement is implemented, after which a random sample of 20 boxes is selected and the sample variance is found to be 0.045 ounce^2. Find a 95% confidence interval on the variance in the weight of the product. Does it appear from the range of the confidence interval that the improvement of the process enhanced quality as far as variability is concerned? Assume normality on the distribution of weights of product.

9.110 A consumer group is interested in comparing operating costs for two different types of automobile engines. The group is able to find 15 owners whose cars have engine type A and 15 whose cars have engine type B. All 30 owners bought their cars at roughly the same time, and all have kept good records for a certain 12-month period. In addition, these owners drove roughly the same number of miles. The cost statistics are $\bar{y}_A = \$87.00/1000$ miles, $\bar{y}_B = \$75.00/1000$ miles, $s_A = \$5.99$, and $s_B = \$4.85$. Compute a 95% confidence interval to estimate $\mu_A - \mu_B$, the difference in the mean operating costs. Assume normality and equal variances.

9.111 Consider the statistic S_p^2, the pooled estimate of σ^2 discussed in Section 9.8. It is used when one is willing to assume that $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Show that the estimator is unbiased for σ^2 [i.e., show that $E(S_p^2) = \sigma^2$]. You may make use of results from any theorem or example in this chapter.

9.112 A group of human factor researchers are concerned about reaction to a stimulus by airplane pilots in a certain cockpit arrangement. An experiment was conducted in a simulation laboratory, and 15 pilots were used with average reaction time of 3.2 seconds with a sample standard deviation of 0.6 second. It is of interest to characterize the extreme (i.e., worst case scenario). To that end, do the following:

(a) Give a particular important one-sided 99% confidence bound on the mean reaction time. What assumption, if any, must you make on the distribution of reaction times?

(b) Give a 99% one-sided prediction interval and give an interpretation of what it means. Must you make an assumption about the distribution of reaction times to compute this bound?

(c) Compute a one-sided tolerance bound with 99% confidence that involves 95% of reaction times. Again, give an interpretation and assumptions about the distribution, if any. (Note: The one-sided tolerance limit values are also included in Table A.7.)

9.113 A certain supplier manufactures a type of rubber mat that is sold to automotive companies. The material used to produce the mats must have certain hardness characteristics. Defective mats are occasionally discovered and rejected. The supplier claims that the proportion defective is 0.05. A challenge was made by one of the clients who purchased the mats, so an experiment was conducted in which 400 mats are tested and 17 were found defective.

(a) Compute a 95% two-sided confidence interval on the proportion defective.

(b) Compute an appropriate 95% one-sided confidence interval on the proportion defective.

(c) Interpret both intervals from (a) and (b) and comment on the claim made by the supplier.

9.15 Potential Misconceptions and Hazards; Relationship to Material in Other Chapters

The concept of a *large-sample confidence interval* on a population is often confusing to the beginning student. It is based on the notion that even when σ is unknown and one is not convinced that the distribution being sampled is normal, a confidence interval on μ can be computed from

$$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}.$$

In practice, this formula is often used when the sample is too small. The genesis of this large sample interval is, of course, the Central Limit Theorem (CLT), under which normality is not necessary. Here the CLT requires a known σ, of which s is only an estimate. Thus, n must be at least as large as 30 and the underlying distribution must be close to symmetric, in which case the interval is still an approximation.

There are instances in which the appropriateness of the practical application of material in this chapter depends very much on the specific context. One very important illustration is the use of the t-distribution for the confidence interval on μ when σ is unknown. Strictly speaking, the use of the t-distribution requires that the distribution sampled from be normal. However, it is well known that any application of the t-distribution is reasonably insensitive (i.e., **robust**) to the normality assumption. This represents one of those fortunate situations which occur often in the field of statistics in which a basic assumption does not hold and yet "everything turns out all right!" However, one population from which the sample is drawn cannot deviate substantially from normal. Thus, the normal probability plots discussed in Chapter 8 and the goodness-of-fit tests introduced in Chapter 10 often need be called upon to ascertain some sense of "nearness to normality." This idea of "robustness to normality" will reappear in Chapter 10.

It is our experience that one of the most serious "misuses of statistics" in practice evolves from confusion about distinctions in the interpretation of the types of statistical intervals. Thus, the subsection in this chapter where differences among the three types of intervals are discussed is important. It is very likely that in practice the **confidence interval is heavily overused**. That is, it is used when there is really no interest in the mean; rather, the question is "Where is the next observation going to fall?" or often, more importantly, "Where is the large bulk of the distribution?" These are crucial questions that are not answered by computing an interval on the mean. The interpretation of a confidence interval is often misunderstood. It is tempting to conclude that the parameter falls inside the interval with probability 0.95. While this is a correct interpretation of a **Bayesian posterior interval** (readers are referred to Chapter 18 for more information on Bayesian inference), it is not the proper frequency interpretation.

A confidence interval merely suggests that if the experiment is conducted and data are observed again and again, about 95% of such intervals will contain the true parameter. Any beginning student of practical statistics should be very clear on the difference among these statistical intervals.

Another potential serious misuse of statistics centers around the use of the χ^2-distribution for a confidence interval on a single variance. Again, normality of the distribution from which the sample is drawn is assumed. Unlike the use of the t-distribution, the use of the χ^2 test for this application is **not robust to the normality assumption** (i.e., the sampling distribution of $\frac{(n-1)S^2}{\sigma^2}$ deviates far from χ^2 if the underlying distribution is not normal). Thus, strict use of goodness-of-fit (Chapter 10) tests and/or normal probability plotting can be extremely important in such contexts. More information about this general issue will be given in future chapters.