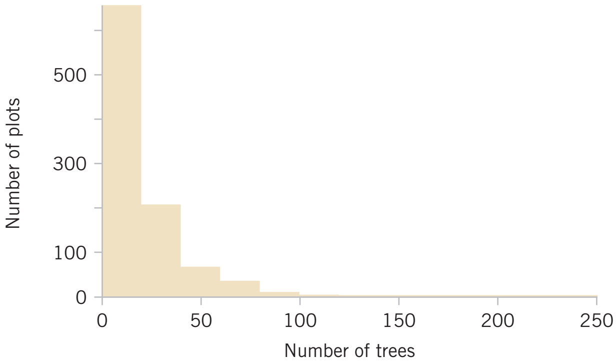

(b) Using the central limit theorem, what is the probability that the mean of the eight would be within 1 standard error of the mean?

(c) Why might you think that the probability that you calculated in (b) might not be very accurate?

7-19. Like hurricanes and earthquakes, geomagnetic storms are natural hazards with possible severe impact on the Earth. Severe storms can cause communication and utility breakdowns, leading to possible blackouts. The National Oceanic and Atmospheric Administration beams electron and proton flux data in various energy ranges to various stations on the Earth to help forecast possible disturbances. The following are 25 readings of proton flux in the 47-68 kEV range (units are in p/(cm2-sec-ster-MeV)) on the evening of December 28, 2011:

2310 2320 2010 10800 2190 3360 5640 2540 3360
11800 2010 3430 10600 7370 2160 3200 2020 2850
3500 10200 8550 9500 2260 7730 2250

(a) Find a point estimate of the mean proton flux in this time period.

(b) Find a point estimate of the standard deviation of the proton flux in this time period.

(c) Find an estimate of the standard error of the estimate in part (a).

(d) Find a point estimate for the median proton flux in this time period.

(e) Find a point estimate for the proportion of readings that are less than 5000 p/(cm2-sec-ster-MeV).

7-20. Wayne Collier designed an experiment to measure the fuel efficiency of his family car under different tire pressures. For each run, he set the tire pressure and then measured the miles he drove on a highway (I-95 between Mills River and Pisgah Forest, NC) until he ran out of fuel using 2 liters of fuel each time. To do this, he made some alterations to the normal flow of gasoline to the engine. In Wayne's words, "I inserted a T-junction into the fuel line just before the fuel filter, and a line into the passenger compartment of my car, where it joined with a graduated 2 liter Rubbermaid© bottle that I mounted in a box where the passenger seat is normally fastened. Then I sealed off the fuel-return line, which under normal operation sends excess fuel from the fuel pump back to the fuel tank."

Suppose that you call the mean miles that he can drive with normal pressure in the tires μ. An unbiased estimate for μ is the mean of the sample runs, $\bar{x}$. But Wayne has a different idea. He decides to use the following estimator: He flips a fair coin. If the coin comes up heads, he will add five miles to each observation. If tails come up, he will subtract five miles from each observation.

(a) Show that Wayne's estimate is, in fact, unbiased.

(b) Compare the standard deviation of Wayne's estimate with the standard deviation of the sample mean.

(c) Given your answer to (b), why does Wayne's estimate not make good sense scientifically?

7-21. Consider a Weibull distribution with shape parameter 1.5 and scale parameter 2.0. Generate a graph of the probability distribution. Does it look very much like a normal distribution? Construct a table similar to Table 7-1 by drawing 20 random samples of size $n = 10$ from this distribution. Compute the sample average from each sample and construct a normal probability plot of the sample averages. Do the sample averages seem to be normally distributed?

7-3 General Concepts of Point Estimation

7-3.1 UNBIASED ESTIMATORS

An estimator should be "close" in some sense to the true value of the unknown parameter. Formally, we say that $\hat{\Theta}$ is an unbiased estimator of θ if the expected value of $\hat{\Theta}$ is equal to θ. This is equivalent to saying that the mean of the probability distribution of $\hat{\Theta}$ (or the mean of the sampling distribution of $\hat{\Theta}$) is equal to θ.

Bias of an Estimator

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$$E(\hat{\Theta}) = \theta \qquad (7\text{-}5)$$

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta}) - \theta \qquad (7\text{-}6)$$

is called the **bias** of the estimator $\hat{\Theta}$.

When an estimator is unbiased, the bias is zero; that is, $E(\hat{\Theta}) - \theta = 0$.

Example 7-4 **Sample Mean and Variance are Unbiased** Suppose that X is a random variable with mean μ and variance σ^2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the population represented by X. Show that the sample mean $\bar{X}$ and sample variance S^2 are unbiased estimators of μ and σ^2, respectively.

First consider the sample mean. In Section 5.5 in Chapter 5, we showed that $E(\bar{X}) = \mu$. Therefore, the sample mean $\bar{X}$ is an unbiased estimator of the population mean μ.

Now consider the sample variance. We have

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

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

The last equality follows the equation for the mean of a linear function in Chapter 5. However, because $E(X_i^2) = \mu^2 + \sigma^2$ and $E(\bar{X}^2) = \mu^2 + \sigma^2/n$, we have

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n)\right]$$

$$= \frac{1}{n-1}\left(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2\right) = \sigma^2$$

Therefore, the sample variance S^2 is an unbiased estimator of the population variance σ^2.

Although S^2 is unbiased for σ^2, S is a biased estimator of σ. For large samples, the bias is very small. However, there are good reasons for using S as an estimator of σ in samples from normal distributions as we will see in the next three chapters when we discuss confidence intervals and hypothesis testing.

Sometimes there are several unbiased estimators of the sample population parameter. For example, suppose that we take a random sample of size $n = 10$ from a normal population and obtain the data $x_1 = 12.8$, $x_2 = 9.4$, $x_3 = 8.7$, $x_4 = 11.6$, $x_5 = 13.1$, $x_6 = 9.8$, $x_7 = 14.1$, $x_8 = 8.5$, $x_9 = 12.1$, $x_{10} = 10.3$. Now the sample mean is

$$\bar{x} = \frac{12.8 + 9.4 + 8.7 + 11.6 + 13.1 + 9.8 + 14.1 + 8.5 + 12.1 + 10.3}{10} = 11.04$$

the sample median is

$$\tilde{x} = \frac{10.3 + 11.6}{2} = 10.95$$

and a 10% trimmed mean (obtained by discarding the smallest and largest 10% of the sample before averaging) is

$$\bar{x}_{tr(10)} = \frac{8.7 + 9.4 + 9.8 + 10.3 + 11.6 + 12.1 + 12.8 + 13.1}{8} = 10.98$$

We can show that all of these are unbiased estimates of μ. Because there is not a unique unbiased estimator, we cannot rely on the property of unbiasedness alone to select our estimator. We need a method to select among unbiased estimators. We suggest a method in the following section.

7-3.2 Variance of a Point Estimator

Suppose that $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased estimators of θ. This indicates that the distribution of each estimator is centered at the true value of zero. However, the variance of these distributions may be different. Figure 7-7 illustrates the situation. Because $\hat{\Theta}_1$ has a smaller variance than $\hat{\Theta}_2$, the estimator $\hat{\Theta}_1$ is more likely to produce an estimate close to the true value of θ. A logical principle of estimation when selecting among several unbiased estimators is to choose the estimator that has minimum variance.

Minimum Variance Unbiased Estimator

> If we consider all unbiased estimators of θ, the one with the smallest variance is called the **minimum variance unbiased estimator** (MVUE).

In a sense, the MVUE is most likely among all unbiased estimators to produce an estimate $\hat{\theta}$ that is close to the true value of θ. It has been possible to develop methodology to identify the MVUE in many practical situations. Although this methodology is beyond the scope of this book, we give one very important result concerning the normal distribution.

> If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, the sample mean $\overline{X}$ is the MVUE for μ.

When we do not know whether an MVUE exists, we could still use a minimum variance principle to choose among competing estimators. Suppose, for example, we wish to estimate the mean of a population (not necessarily a *normal* population). We have a random sample of n observations $X_1, X_2, \ldots, X_n$, and we wish to compare two possible estimators for μ: the sample mean $\overline{X}$ and a single observation from the sample, say, X_i. Note that both $\overline{X}$ and X_i are unbiased estimators of μ; for the sample mean, we have $V(\overline{X}) = \sigma^2/n$ from Chapter 5 and the variance of any observation is $V(X_i) = \sigma^2$. Because $V(\overline{X}) < V(X_i)$ for sample sizes $n \geq 2$, we would conclude that the sample mean is a better estimator of μ than a single observation X_i.

7-3.3 Standard Error: Reporting a Point Estimate

When the numerical value or point estimate of a parameter is reported, it is usually desirable to give some idea of the precision of estimation. The measure of precision usually employed is the standard error of the estimator that has been used.

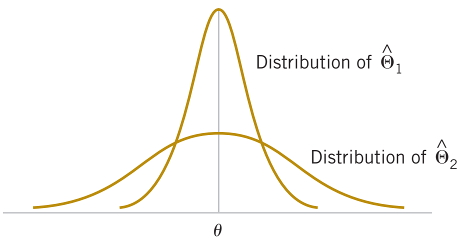

FIGURE 7-7 The sampling distributions of two unbiased estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$.

Standard Error of an Estimator

> The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation given by $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$. If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error**, denoted by $\hat{\sigma}_{\hat{\Theta}}$.

Sometimes the estimated standard error is denoted by $s_{\hat{\Theta}}$ or $se(\hat{\Theta})$.

Suppose that we are sampling from a normal distribution with mean μ and variance σ^2. Now the distribution of $\overline{X}$ is normal with mean μ and variance σ^2/n, so the **standard error** of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

If we did not know σ but substituted the sample standard deviation S into the preceding equation, the **estimated standard error** of $\overline{X}$ would be

$$\text{SE}(\overline{X}) = \hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$$

When the estimator follows a normal distribution as in the preceding situation, we can be reasonably confident that the true value of the parameter lies within two standard errors of the estimate. Because many point estimators are normally distributed (or approximately so) for large n, this is a very useful result. Even when the point estimator is not normally distributed, we can state that so long as the estimator is unbiased, the estimate of the parameter will deviate from the true value by as much as four standard errors at most 6 percent of the time. Thus, a very conservative statement is that the true value of the parameter differs from the point estimate by at most four standard errors. See Chebyshev's inequality in the supplemental material on the Web site.

Example 7-5 **Thermal Conductivity** An article in the *Journal of Heat Transfer* (Trans. ASME, Sec. C, 96, 1974, p. 59) described a new method of measuring the thermal conductivity of Armco iron. Using a temperature of 100°F and a power input of 550 watts, the following 10 measurements of thermal conductivity (in Btu/hr-ft-°F) were obtained:

$$41.60, 41.48, 42.34, 41.95, 41.86,$$
$$42.18, 41.72, 42.26, 41.81, 42.04$$

A point estimate of the mean thermal conductivity at 100°F and 550 watts is the sample mean or

$$\overline{x} = 41.924 \text{ Btu/hr-ft-°F}$$

The standard error of the sample mean is $\sigma_{\overline{X}} = \sigma/\sqrt{n}$, and because σ is unknown, we may replace it by the sample standard deviation $s = 0.284$ to obtain the estimated standard error of $\overline{X}$ as

$$SE(\overline{X}) = \hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{0.284}{\sqrt{10}} = 0.0898$$

Practical Interpretation: Notice that the standard error is about 0.2 percent of the sample mean, implying that we have obtained a relatively precise point estimate of thermal conductivity. If we can assume that thermal conductivity is normally distributed, 2 times the standard error is $2\hat{\sigma}_{\overline{X}} = 2(0.0898) = 0.1796$, and we are highly confident that the true mean thermal conductivity is within the interval 41.924 ± 0.1796 or between 41.744 and 42.104.

7.3.4 Bootstrap Standard Error

In some situations, the form of a point estimator is complicated, and standard statistical methods to find its standard error are difficult or impossible to apply. One example of these is S, the point estimator of the population standard deviation σ. Others occur with