5.1 The Sampling Distribution of a Sample Mean

When you complete this section, you will be able to

- **Explain the difference between the sampling distribution of $\bar{x}$ and the population distribution.**
- **Determine the mean and standard deviation of $\bar{x}$ for an SRS of size n from a population with mean μ and standard deviation σ.**
- **Describe how much larger n has to be to reduce the standard deviation of $\bar{x}$ by a certain factor.**
- **Utilize the central limit theorem to approximate the sampling distribution of $\bar{x}$ and perform various probability calculations.**

THE DISTRIBUTION OF A STATISTIC

A statistic from a random sample or randomized experiment is a random variable. The probability distribution of the statistic is its **sampling distribution.**

POPULATION DISTRIBUTION

The **population distribution** of a variable is the distribution of its values for all members of the population. The population distribution is also the probability distribution of the variable when we choose one individual at random from the population.

FACTS ABOUT SAMPLE MEANS

1. Sample means are less variable than individual observations.
2. Sample means are more Normal than individual observations.

MEAN AND STANDARD DEVIATION OF A SAMPLE MEAN

Let $\bar{x}$ be the mean of an SRS of size n from a population having mean μ and standard deviation σ. The mean and standard deviation of $\bar{x}$ are

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

SAMPLING DISTRIBUTION OF A SAMPLE MEAN

If a population has the $N(\mu, \sigma)$ distribution, then the sample mean $\bar{x}$ of n independent observations has the $N(\mu, \sigma/\sqrt{n})$ distribution.

CENTRAL LIMIT THEOREM

Draw an SRS of size n from any population with mean μ and finite standard deviation σ. When n is large, the sampling distribution of the sample mean $\bar{x}$ is approximately Normal:

$$\bar{x} \text{ is approximately } N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

- **Take many random samples of size n from a population with mean μ and standard deviation σ.**
- **Find the sample mean $\bar{x}$ for each sample.**
- **Collect all the $\bar{x}$'s and display their distribution.**

The sampling distribution of $\bar{x}$ is shown on the right. Keep this figure in mind as you go forward.

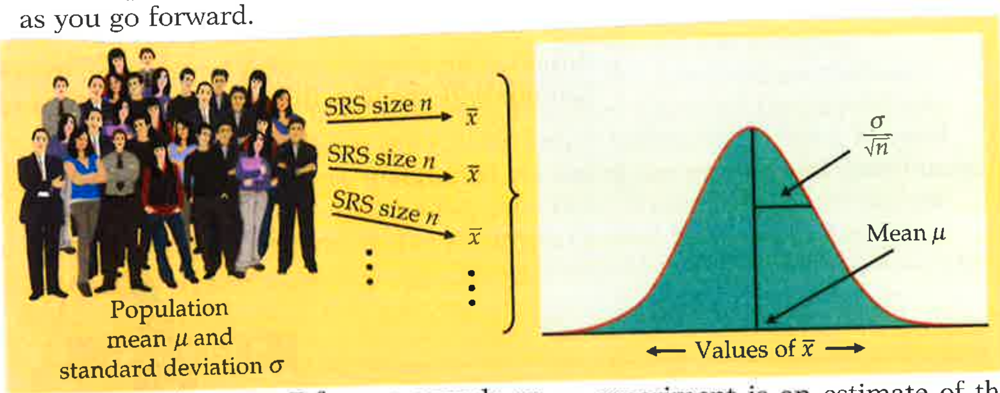

FIGURE 5.6 The sampling distribution of a sample mean $\bar{x}$ has mean μ and standard deviation $\sigma/\sqrt{n}$. The sampling distribution is Normal if the population distribution is Normal; it is approximately Normal for large samples in any case.

The sample mean $\bar{x}$ from a sample or an experiment is an estimate of the mean μ of the underlying population. The sampling distribution of $\bar{x}$ is determined by the design used to produce the data, the sample size n, and the population distribution.

Select an SRS of size n from a population, and measure a variable X on each individual in the sample. The n measurements are values of n random variables $X_1, X_2, \ldots, X_n$. A single X_i is a measurement on one individual selected at random from the population and therefore has the distribution of the population. If the population is large relative to the sample, we can consider $X_1, X_2, \ldots, X_n$ to be independent random variables each having the same distribution. This is our probability model for measurements on each individual in an SRS.

The sample mean of an SRS of size n is

$$\bar{x} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

LOOK BACK rules for means, p. 272

If the population has mean μ, then μ is the mean of the distribution of each observation X_i. To get the mean of $\bar{x}$, we use the rules for means of random variables. Specifically,

$$\mu_{\bar{x}} = \frac{1}{n}(\mu_{X_1} + \mu_{X_2} + \cdots + \mu_{X_n})$$

$$= \frac{1}{n}(\mu + \mu + \cdots + \mu) = \mu$$

LOOK BACK unbiased estimator, p. 210

That is, *the mean of $\bar{x}$ is the same as the mean of the population*. The sample mean $\bar{x}$ is therefore an unbiased estimator of the unknown population mean μ.

LOOK BACK rules for variances, p. 275

The observations are independent, so the addition rule for variances also applies:

$$\sigma^2_{\bar{x}} = \left(\frac{1}{n}\right)^2 (\sigma^2_{X_1} + \sigma^2_{X_2} + \cdots + \sigma^2_{X_n})$$

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

$$= \frac{\sigma^2}{n}$$

With n in the denominator, the variability of $\bar{x}$ about its mean decreases as the sample size grows. Thus, a sample mean from a large sample will usually be very close to the true population mean μ. Here is a summary of these facts.