

The term *standard error* is used throughout statistical inference to refer to the standard deviation of a point estimator.

To compute $\sigma_{\bar{x}}$, we need to know σ , the standard deviation of the population. To further emphasize the difference between $\sigma_{\bar{x}}$ and σ , we refer to the standard deviation of \bar{x} , $\sigma_{\bar{x}}$, as the **standard error** of the mean. In general, the term *standard error* refers to the standard deviation of a point estimator. Later we will see that the value of the standard error of the mean is helpful in determining how far the sample mean may be from the population mean. Let us now return to the EAI example and compute the standard error of the mean associated with simple random samples of 30 EAI employees.

In Section 7.1 we saw that the standard deviation of annual salary for the population of 2500 EAI employees is $\sigma = 4000$. In this case, the population is finite, with $N = 2500$. However, with a sample size of 30, we have $n/N = 30/2500 = .012$. Because the sample size is less than 5% of the population size, we can ignore the finite population correction factor and use equation (7.3) to compute the standard error.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{30}} = 730.3$$

Form of the Sampling Distribution of \bar{x}

The preceding results concerning the expected value and standard deviation for the sampling distribution of \bar{x} are applicable for any population. The final step in identifying the characteristics of the sampling distribution of \bar{x} is to determine the form or shape of the sampling distribution. We will consider two cases: (1) The population has a normal distribution; and (2) the population does not have a normal distribution.

Population has a normal distribution In many situations it is reasonable to assume that the population from which we are selecting a random sample has a normal, or nearly normal, distribution. When the population has a normal distribution, the sampling distribution of \bar{x} is normally distributed for any sample size.

Population does not have a normal distribution When the population from which we are selecting a random sample does not have a normal distribution, the **central limit theorem** is helpful in identifying the shape of the sampling distribution of \bar{x} . A statement of the central limit theorem as it applies to the sampling distribution of \bar{x} follows.

CENTRAL LIMIT THEOREM

In selecting random samples of size n from a population, the sampling distribution of the sample mean \bar{x} can be approximated by a *normal distribution* as the sample size becomes large.

Figure 7.6 shows how the central limit theorem works for three different populations; each column refers to one of the populations. The top panel of the figure shows that none of the populations are normally distributed. Population I follows a uniform distribution. Population II is often called the rabbit-eared distribution. It is symmetric, but the more likely values fall in the tails of the distribution. Population III is shaped like the exponential distribution; it is skewed to the right.

The bottom three panels of Figure 7.6 show the shape of the sampling distribution for samples of size $n = 2$, $n = 5$, and $n = 30$. When the sample size is 2, we see that the shape of each sampling distribution is different from the shape of the corresponding population distribution. For samples of size 5, we see that the shapes of the sampling distributions

From a practitioner standpoint, we often want to know how large the sample size needs to be before the central limit theorem applies and we can assume that the shape of the sampling distribution is approximately normal. Statistical researchers have investigated this question by studying the sampling distribution of \bar{x} for a variety of populations and a variety of sample sizes. General statistical practice is to assume that, for most

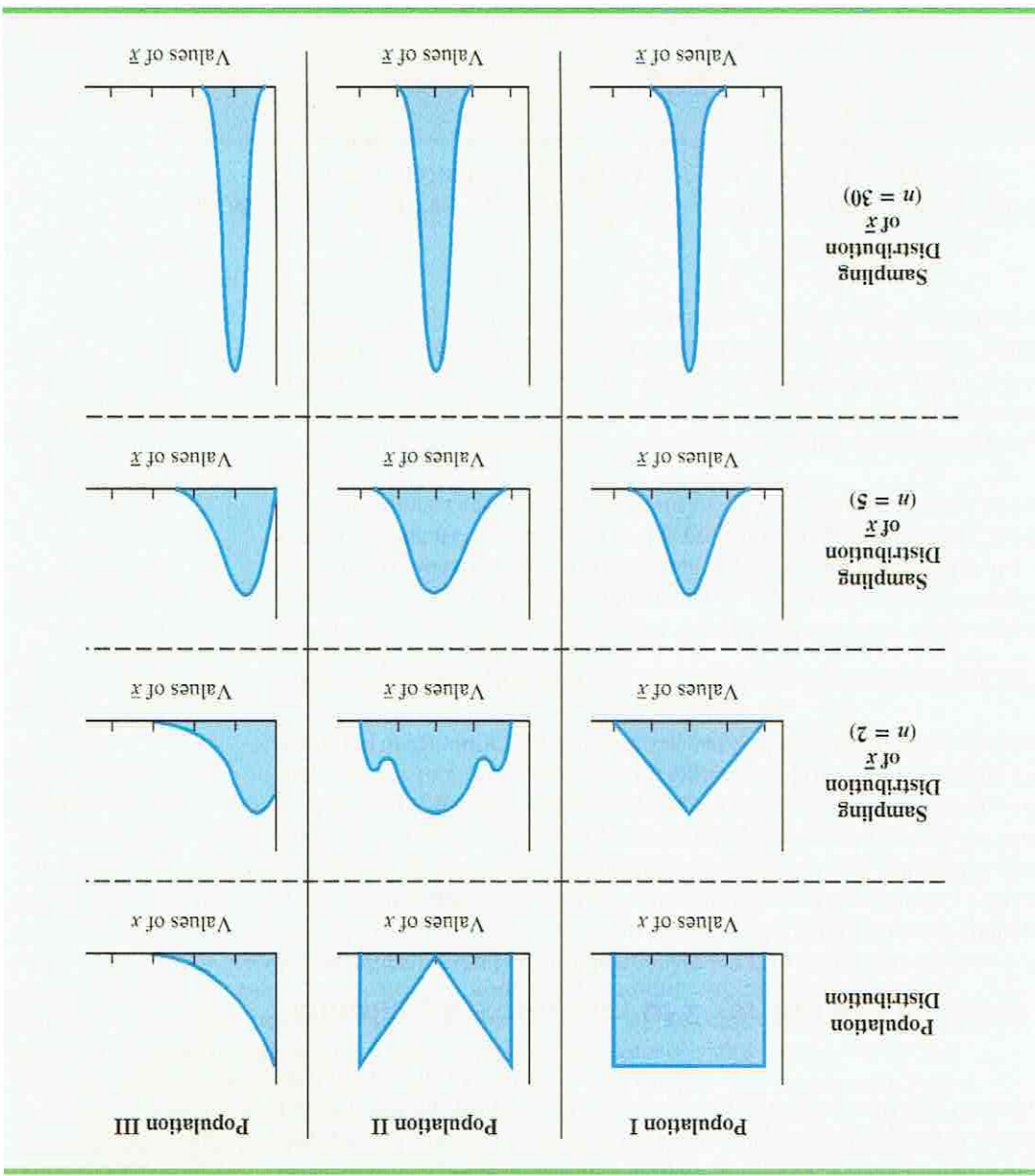


FIGURE 7.6 ILLUSTRATION OF THE CENTRAL LIMIT THEOREM FOR THREE POPULATIONS

applications, the sampling distribution of \bar{x} can be approximated by a normal distribution whenever the sample size is 30 or more. In cases where the population is highly skewed or outliers are present, samples of size 50 may be needed. Finally, if the population is discrete, the sample size needed for a normal approximation often depends on the population proportion. We say more about this issue when we discuss the sampling distribution of \bar{p} in Section 7.6.

Sampling Distribution of \bar{x} for the EAI Problem

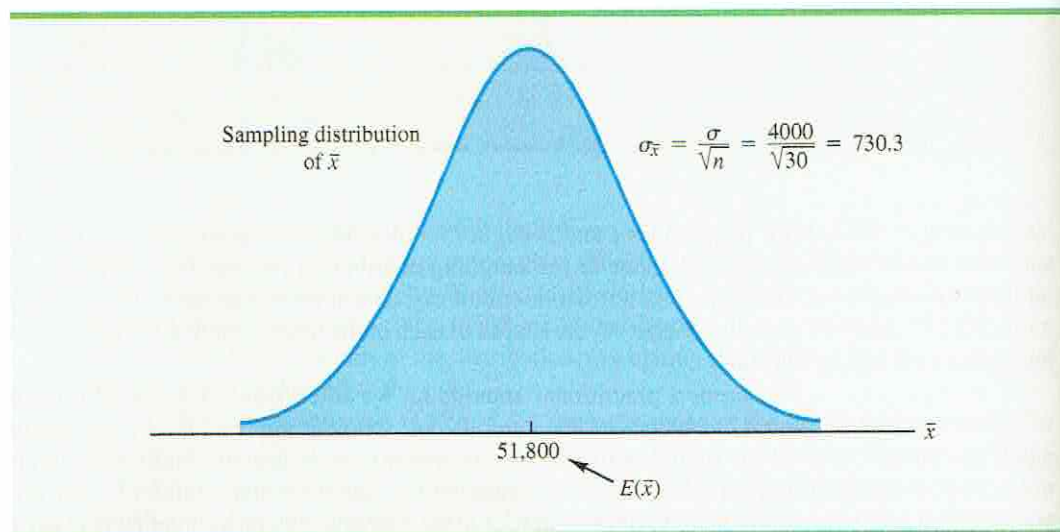
Let us return to the EAI problem where we previously showed that $E(\bar{x}) = \$51,800$ and $\sigma_{\bar{x}} = 730.3$. At this point, we do not have any information about the population distribution; it may or may not be normally distributed. If the population has a normal distribution, the sampling distribution of \bar{x} is normally distributed. If the population does not have a normal distribution, the simple random sample of 30 employees and the central limit theorem enable us to conclude that the sampling distribution of \bar{x} can be approximated by a normal distribution. In either case, we are comfortable proceeding with the conclusion that the sampling distribution of \bar{x} can be described by the normal distribution shown in Figure 7.7.

Practical Value of the Sampling Distribution of \bar{x}

Whenever a simple random sample is selected and the value of the sample mean is used to estimate the value of the population mean μ , we cannot expect the sample mean to exactly equal the population mean. The practical reason we are interested in the sampling distribution of \bar{x} is that it can be used to provide probability information about the difference between the sample mean and the population mean. To demonstrate this use, let us return to the EAI problem.

Suppose the personnel director believes the sample mean will be an acceptable estimate of the population mean if the sample mean is within \$500 of the population mean. However, it is not possible to guarantee that the sample mean will be within \$500 of the population mean. Indeed, Table 7.5 and Figure 7.4 show that some of the 500 sample means differed by more than \$2000 from the population mean. So we must think of the personnel director's

FIGURE 7.7 SAMPLING DISTRIBUTION OF \bar{x} FOR THE MEAN ANNUAL SALARY OF A SIMPLE RANDOM SAMPLE OF 30 EAI EMPLOYEES



request in probability terms. That is, the personnel director is concerned with the following question: What is the probability that the sample mean computed using a simple random sample of 30 EAI employees will be within \$50 of the population mean?

Because we have identified the properties of the sampling distribution of \bar{x} (see Figure 7.7), we will use this distribution to answer the probability question. Refer to the sampling distribution of \bar{x} shown again in Figure 7.8. With a population mean of \$51,800, the personnel director wants to know the probability that \bar{x} is between \$51,300 and \$52,300. This probability is given by the darkly shaded area of the sampling distribution shown in Figure 7.8. Because the sampling distribution is normally distributed, with mean \$51,800 and standard error of the mean 730.3, we can use the standard normal probability table to find the area or probability.

We first calculate the z value at the upper endpoint of the interval (\$52,300) and use the table to find the cumulative probability at that point (left tail area). Then we compute the z value at the lower endpoint of the interval (\$51,300) and use the table to find the area under the curve to the left of that point (another left tail area). Subtracting the second tail area from the first gives us the desired probability.

At $\bar{x} = 52,300$, we have

$$z = \frac{52,300 - 51,800}{730.30} = .68$$

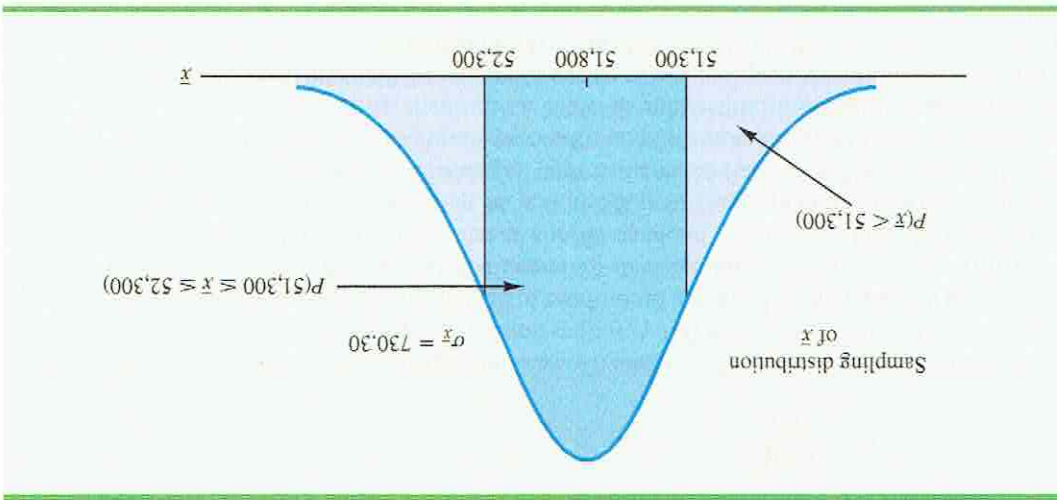
Referring to the standard normal probability table, we find a cumulative probability (area) to the left of $z = .68$ of .7517.

At $\bar{x} = 51,300$, we have

$$z = \frac{51,300 - 51,800}{730.30} = -.68$$

The area under the curve to the left of $z = -.68$ is .2483. Therefore, $P(51,300 \leq \bar{x} \leq 52,300) = P(z \leq .68) - P(z < -.68) = .7517 - .2483 = .5034$.

FIGURE 7.8 PROBABILITY OF A SAMPLE MEAN BEING WITHIN \$50 OF THE POPULATION MEAN FOR A SIMPLE RANDOM SAMPLE OF 30 EAI EMPLOYEES



Using Excel's NORM.DIST function is easier and provides more accurate results than using the tables with rounded values for z .

The desired probability can also be computed using Excel's NORM.DIST function. The advantage of using the NORM.DIST function is that we do not have to make a separate computation of the z value. Evaluating the NORM.DIST function at the upper endpoint of the interval provides the cumulative probability at 52,300. Entering the formula =NORM.DIST(52300,51800,730.30,TRUE) into a cell of an Excel worksheet provides .7532 for this cumulative probability. Evaluating the NORM.DIST function at the lower endpoint of the interval provides the area under the curve to the left of 51,300. Entering the formula =NORM.DIST(51300,51800,730.30,TRUE) into a cell of an Excel worksheet provides .2468 for this cumulative probability. The probability of \bar{x} being in the interval from 51,300 to 52,300 is then given by $.7532 - .2468 = .5064$. We note that this result is slightly different from the probability obtained using the table, because in using the normal table we rounded to two decimal places of accuracy when computing the z value. The result obtained using NORM.DIST is thus more accurate.

The preceding computations show that a simple random sample of 30 EAI employees has a .5064 probability of providing a sample mean \bar{x} that is within \$500 of the population mean. Thus, there is a $1 - .5064 = .4936$ probability that the sampling error will be more than \$500. In other words, a simple random sample of 30 EAI employees has roughly a 50–50 chance of providing a sample mean within the allowable \$500. Perhaps a larger sample size should be considered. Let us explore this possibility by considering the relationship between the sample size and the sampling distribution of \bar{x} .

The sampling distribution of \bar{x} can be used to provide probability information about how close the sample mean \bar{x} is to the population mean μ .

Relationship Between the Sample Size and the Sampling Distribution of \bar{x}

Suppose that in the EAI sampling problem we select a simple random sample of 100 EAI employees instead of the 30 originally considered. Intuitively, it would seem that with more data provided by the larger sample size, the sample mean based on $n = 100$ should provide a better estimate of the population mean than the sample mean based on $n = 30$. To see how much better, let us consider the relationship between the sample size and the sampling distribution of \bar{x} .

First note that $E(\bar{x}) = \mu$ regardless of the sample size. Thus, the mean of all possible values of \bar{x} is equal to the population mean μ regardless of the sample size n . However, note that the standard error of the mean, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, is related to the square root of the sample size. Whenever the sample size is increased, the standard error of the mean $\sigma_{\bar{x}}$ decreases. With $n = 30$, the standard error of the mean for the EAI problem is 730.3. However, with the increase in the sample size to $n = 100$, the standard error of the mean is decreased to

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4000}{\sqrt{100}} = 400$$

The sampling distributions of \bar{x} with $n = 30$ and $n = 100$ are shown in Figure 7.9. Because the sampling distribution with $n = 100$ has a smaller standard error, the values of \bar{x} have less variation and tend to be closer to the population mean than the values of \bar{x} with $n = 30$.

We can use the sampling distribution of \bar{x} for the case with $n = 100$ to compute the probability that a simple random sample of 100 EAI employees will provide a sample mean that is within \$500 of the population mean. In this case the sampling distribution is normal with a mean of 51,800 and a standard deviation of 400 (see Figure 7.10). Again, we could compute the appropriate z values and use the standard normal probability distribution table to make this probability calculation. However, Excel's NORM.DIST function is easier to use and provides more accurate results. Entering the formula =NORM.DIST(52300,51800,400,TRUE) into a cell of an Excel worksheet provides the cumulative probability corresponding to $\bar{x} = 52,300$. The value provided

25. The mean preparation fee H&R Block charged retail customers in 2012 was \$183 (*The Wall Street Journal*, March 7, 2012). Use this price as the population mean and assume the population standard deviation of preparation fees is \$50.
- What is the probability that the mean price for a sample of 30 H&R Block retail customers is within \$8 of the population mean?
 - What is the probability that the mean price for a sample of 50 H&R Block retail customers is within \$8 of the population mean?
 - What is the probability that the mean price for a sample of 100 H&R Block retail customers is within \$8 of the population mean?
 - Which, if any, of the sample sizes in parts (a), (b), and (c) would you recommend to have at least a .95 probability that the sample mean is within \$8 of the population mean?
26. To estimate the mean age for a population of 4000 employees, a simple random sample of 40 employees is selected.
- Would you use the finite population correction factor in calculating the standard error of the mean? Explain.
 - If the population standard deviation is $\sigma = 8.2$ years, compute the standard error both with and without the finite population correction factor. What is the rationale for ignoring the finite population correction factor whenever $n/N \leq .05$?
 - What is the probability that the sample mean age of the employees will be within ± 2 years of the population mean age?

7.6

Sampling Distribution of \bar{p}

The sample proportion \bar{p} is the point estimator of the population proportion p . The formula for computing the sample proportion is

$$\bar{p} = \frac{x}{n}$$

where

- x = the number of elements in the sample that possess the characteristic of interest
- n = sample size

As noted in Section 7.4, the sample proportion \bar{p} is a random variable and its probability distribution is called the sampling distribution of \bar{p} .

SAMPLING DISTRIBUTION OF \bar{p}

The sampling distribution of \bar{p} is the probability distribution of all possible values of the sample proportion \bar{p} .

To determine how close the sample proportion \bar{p} is to the population proportion p , we need to understand the properties of the sampling distribution of \bar{p} : the expected value of \bar{p} , the standard deviation of \bar{p} , and the shape or form of the sampling distribution of \bar{p} .

Expected Value of \bar{p}

The expected value of \bar{p} , the mean of all possible values of \bar{p} , is equal to the population proportion p .

EXPECTED VALUE OF \bar{p}

$$E(\bar{p}) = p \quad (7.4)$$