



This is “Sampling Distributions”, chapter 6 from the book [Beginning Statistics \(index.html\)](#) (v. 1.0).

This book is licensed under a [Creative Commons by-nc-sa 3.0](http://creativecommons.org/licenses/by-nc-sa/3.0/) license. See the license for more details, but that basically means you can share this book as long as you credit the author (but see below), don't make money from it, and do make it available to everyone else under the same terms.

This content was accessible as of December 29, 2012, and it was downloaded then by [Andy Schmitz](http://lardbucket.org) in an effort to preserve the availability of this book.

Normally, the author and publisher would be credited here. However, the publisher has asked for the customary Creative Commons attribution to the original publisher, authors, title, and book URI to be removed. Additionally, per the publisher's request, their name has been removed in some passages. More information is available on this project's [attribution page](http://2012books.lardbucket.org/attribution.html?utm_source=header).

For more information on the source of this book, or why it is available for free, please see [the project's home page](http://2012books.lardbucket.org/). You can browse or download additional books there.

Chapter 6

Sampling Distributions

A statistic, such as the sample mean or the sample standard deviation, is a number computed from a sample. Since a sample is random, every statistic is a random variable: it varies from sample to sample in a way that cannot be predicted with certainty. As a random variable it has a mean, a standard deviation, and a probability distribution. The probability distribution of a statistic is called its **sampling distribution**¹. Typically sample statistics are not ends in themselves, but are computed in order to estimate the corresponding population parameters, as illustrated in the grand picture of statistics presented in [Figure 1.1 "The Grand Picture of Statistics"](#) in [Chapter 1 "Introduction"](#).

This chapter introduces the concepts of the mean, the standard deviation, and the sampling distribution of a sample statistic, with an emphasis on the sample mean \bar{x} .

1. The probability distribution of a sample statistic when the statistic is viewed as a random variable.

6.1 The Mean and Standard Deviation of the Sample Mean

LEARNING OBJECTIVES

1. To become familiar with the concept of the probability distribution of the sample mean.
2. To understand the meaning of the formulas for the mean and standard deviation of the sample mean.

Suppose we wish to estimate the mean μ of a population. In actual practice we would typically take just one sample. Imagine however that we take sample after sample, all of the same size n , and compute the sample mean \bar{x} of each one. We will likely get a different value of \bar{x} each time. The sample mean \bar{x} is a random variable: it varies from sample to sample in a way that cannot be predicted with certainty. We will write \bar{X} when the sample mean is thought of as a random variable, and write \bar{x} for the values that it takes. The random variable \bar{X} has a **mean**², denoted $\mu_{\bar{X}}$, and a **standard deviation**³, denoted $\sigma_{\bar{X}}$. Here is an example with such a small population and small sample size that we can actually write down every single sample.

2. The number about which means computed from samples of the same size center.
3. A measure of the variability of means computed from samples of the same size.

EXAMPLE 1

A rowing team consists of four rowers who weigh 152, 156, 160, and 164 pounds. Find all possible random samples with replacement of size two and compute the sample mean for each one. Use them to find the probability distribution, the mean, and the standard deviation of the sample mean \bar{X} .

Solution

The following table shows all possible samples with replacement of size two, along with the mean of each:

Sample	Mean	Sample	Mean	Sample	Mean	Sample	Mean
152, 152	152	156, 152	154	160, 152	156	164, 152	158
152, 156	154	156, 156	156	160, 156	158	164, 156	160
152, 160	156	156, 160	158	160, 160	160	164, 160	162
152, 164	158	156, 164	160	160, 164	162	164, 164	164

The table shows that there are seven possible values of the sample mean \bar{X} . The value $\bar{x} = 152$ happens only one way (the rower weighing 152 pounds must be selected both times), as does the value $\bar{x} = 164$, but the other values happen more than one way, hence are more likely to be observed than 152 and 164 are. Since the 16 samples are equally likely, we obtain the probability distribution of the sample mean just by counting:

\bar{x}	152	154	156	158	160	162	164
$P(\bar{x})$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

Now we apply the formulas from [Section 4.2.2 "The Mean and Standard Deviation of a Discrete Random Variable"](#) in [Chapter 4 "Discrete Random Variables"](#) for the mean and standard deviation of a discrete random variable to \bar{X} . For $\mu_{\bar{X}}$ we obtain.

$$\begin{aligned}
 \mu_{\bar{X}} &= \sum \bar{x} P(\bar{x}) \\
 &= 152 \left(\frac{1}{16} \right) + 154 \left(\frac{2}{16} \right) + 156 \left(\frac{3}{16} \right) + 158 \left(\frac{4}{16} \right) + 160 \left(\frac{3}{16} \right) \\
 &= 158
 \end{aligned}$$

For $\sigma_{\bar{X}}$ we first compute $\sum \bar{x}^2 P(\bar{x})$:

$$152^2 \left(\frac{1}{16} \right) + 154^2 \left(\frac{2}{16} \right) + 156^2 \left(\frac{3}{16} \right) + 158^2 \left(\frac{4}{16} \right) + 160^2 \left(\frac{3}{16} \right)$$

which is 24,974, so that

$$\sigma_{\bar{X}} = \sqrt{\sum \bar{x}^2 P(\bar{x}) - \mu_{\bar{X}}^2} = \sqrt{24,974 - 158^2} = \sqrt{10}$$

The mean and standard deviation of the population $\{152, 156, 160, 164\}$ in the example are $\mu = 158$ and $\sigma = \sqrt{20}$. The mean of the sample mean \bar{X} that we have just computed is exactly the mean of the population. The standard deviation of the sample mean \bar{X} that we have just computed is the standard deviation of the population divided by the square root of the sample size: $\sqrt{10} = \sqrt{20}/\sqrt{2}$. These relationships are not coincidences, but are illustrations of the following formulas.

Suppose random samples of size n are drawn from a population with mean μ and standard deviation σ . The mean $\mu_{\bar{X}}$ and standard deviation $\sigma_{\bar{X}}$ of the sample mean \bar{X} satisfy

$$\mu_{\bar{X}} = \mu \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

The first formula says that if we could take every possible sample from the population and compute the corresponding sample mean, then those numbers would center at the number we wish to estimate, the population mean μ .

The second formula says that averages computed from samples vary less than individual measurements on the population do, and quantifies the relationship.

EXAMPLE 2

The mean and standard deviation of the tax value of all vehicles registered in a certain state are $\mu = \$13,525$ and $\sigma = \$4,180$. Suppose random samples of size 100 are drawn from the population of vehicles. What are the mean $\mu_{\bar{X}}$ and standard deviation $\sigma_{\bar{X}}$ of the sample mean \bar{X} ?

Solution

Since $n = 100$, the formulas yield

$$\mu_{\bar{X}} = \mu = \$13,525 \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\$4180}{\sqrt{100}} = \$418$$

KEY TAKEAWAYS

- The sample mean is a random variable; as such it is written \bar{X} , and \bar{x} stands for individual values it takes.
- As a random variable the sample mean has a probability distribution, a mean $\mu_{\bar{X}}$, and a standard deviation $\sigma_{\bar{X}}$.
- There are formulas that relate the mean and standard deviation of the sample mean to the mean and standard deviation of the population from which the sample is drawn.

EXERCISES

1. Random samples of size 225 are drawn from a population with mean 100 and standard deviation 20. Find the mean and standard deviation of the sample mean.
2. Random samples of size 64 are drawn from a population with mean 32 and standard deviation 5. Find the mean and standard deviation of the sample mean.
3. A population has mean 75 and standard deviation 12.
 - a. Random samples of size 121 are taken. Find the mean and standard deviation of the sample mean.
 - b. How would the answers to part (a) change if the size of the samples were 400 instead of 121?
4. A population has mean 5.75 and standard deviation 1.02.
 - a. Random samples of size 81 are taken. Find the mean and standard deviation of the sample mean.
 - b. How would the answers to part (a) change if the size of the samples were 25 instead of 81?

ANSWERS

1. $\mu_{\bar{X}} = 100, \sigma_{\bar{X}} = 1.33$
3.
 - a. $\mu_{\bar{X}} = 75, \sigma_{\bar{X}} = 1.09$
 - b. $\mu_{\bar{X}}$ stays the same but $\sigma_{\bar{X}}$ decreases to 0.6

6.2 The Sampling Distribution of the Sample Mean

LEARNING OBJECTIVES

1. To learn what the sampling distribution of \bar{X} is when the sample size is large.
2. To learn what the sampling distribution of \bar{X} is when the population is normal.

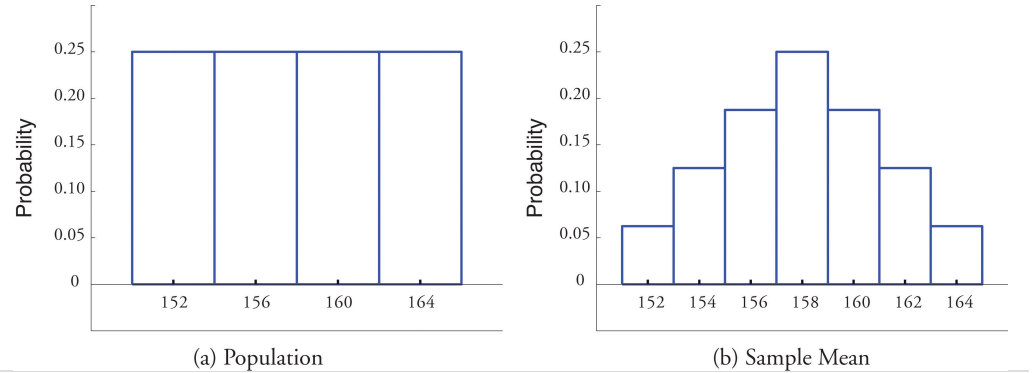
The Central Limit Theorem

In [Note 6.5 "Example 1"](#) in [Section 6.1 "The Mean and Standard Deviation of the Sample Mean"](#) we constructed the probability distribution of the sample mean for samples of size two drawn from the population of four rowers. The probability distribution is:

\bar{x}	152	154	156	158	160	162	164
$P(\bar{x})$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

[Figure 6.1 "Distribution of a Population and a Sample Mean"](#) shows a side-by-side comparison of a histogram for the original population and a histogram for this distribution. Whereas the distribution of the population is uniform, the sampling distribution of the mean has a shape approaching the shape of the familiar bell curve. This phenomenon of the sampling distribution of the mean taking on a bell shape even though the population distribution is not bell-shaped happens in general. Here is a somewhat more realistic example.

Figure 6.1 Distribution of a Population and a Sample Mean



Suppose we take samples of size 1, 5, 10, or 20 from a population that consists entirely of the numbers 0 and 1, half the population 0, half 1, so that the population mean is 0.5. The sampling distributions are:

$n = 1$:

\bar{x}	0	1
$P(\bar{x})$	0.5	0.5

$n = 5$:

\bar{x}	0	0.2	0.4	0.6	0.8	1
$P(\bar{x})$	0.03	0.16	0.31	0.31	0.16	0.03

$n = 10$:

\bar{x}	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$P(\bar{x})$	0.00	0.01	0.04	0.12	0.21	0.25	0.21	0.12	0.04	0.01	0.00

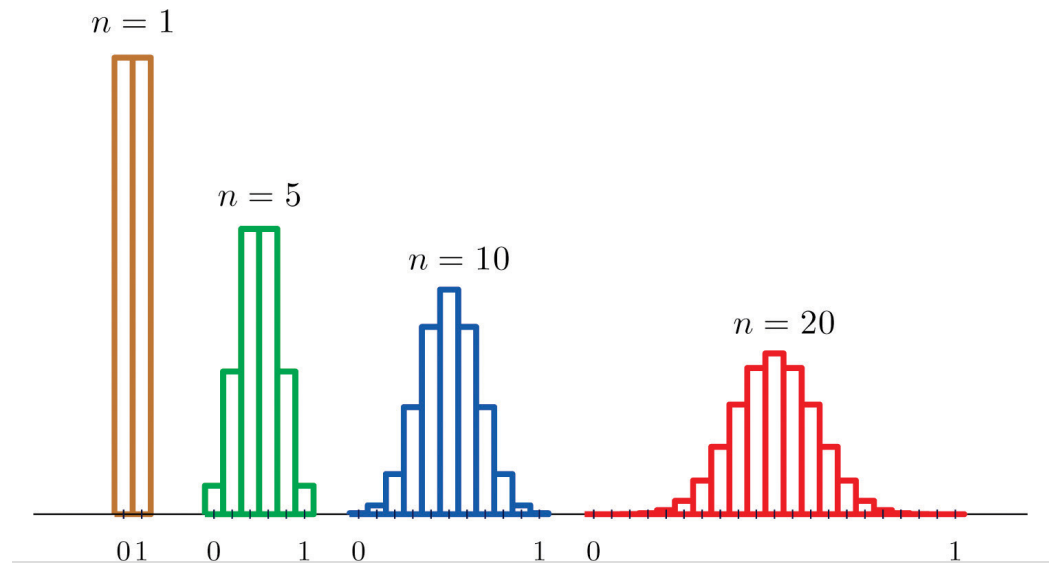
$n = 20$:

\bar{x}	0	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
$P(\bar{x})$	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.07	0.12	0.16	0.18

\bar{x}	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1
$P(\bar{x})$	0.16	0.12	0.07	0.04	0.01	0.00	0.00	0.00	0.00	0.00

Histograms illustrating these distributions are shown in [Figure 6.2 "Distributions of the Sample Mean"](#).

Figure 6.2 *Distributions of the Sample Mean*



As n increases the sampling distribution of \bar{X} evolves in an interesting way: the probabilities on the lower and the upper ends shrink and the probabilities in the middle become larger in relation to them. If we were to continue to increase n then the shape of the sampling distribution would become smoother and more bell-shaped.

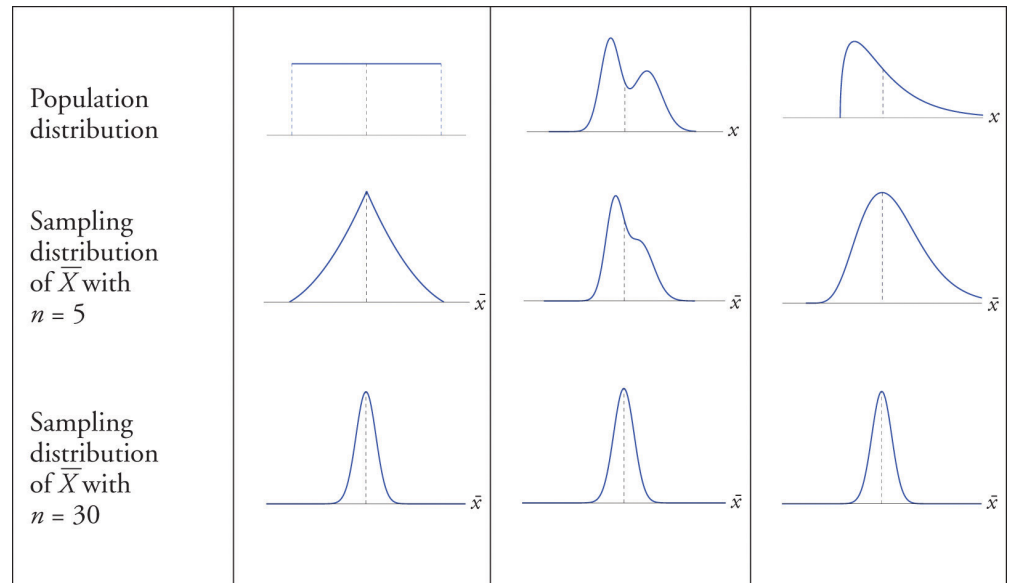
What we are seeing in these examples does not depend on the particular population distributions involved. In general, one may start with any distribution and the sampling distribution of the sample mean will increasingly resemble the bell-shaped normal curve as the sample size increases. This is the content of the Central Limit Theorem.

The Central Limit Theorem

For samples of size 30 or more, the sample mean is approximately normally distributed, with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, where n is the sample size. The larger the sample size, the better the approximation.

The Central Limit Theorem is illustrated for several common population distributions in [Figure 6.3 "Distribution of Populations and Sample Means"](#).

Figure 6.3 *Distribution of Populations and Sample Means*



The dashed vertical lines in the figures locate the population mean. Regardless of the distribution of the population, as the sample size is increased the shape of the sampling distribution of the sample mean becomes increasingly bell-shaped, centered on the population mean. Typically by the time the sample size is 30 the distribution of the sample mean is practically the same as a normal distribution.

The importance of the Central Limit Theorem is that it allows us to make probability statements about the sample mean, specifically in relation to its value in comparison to the population mean, as we will see in the examples. But to use the result properly we must first realize that there are two separate random variables (and therefore two probability distributions) at play:

1. X , the measurement of a single element selected at random from the population; the distribution of X is the distribution of the population, with mean the population mean μ and standard deviation the population standard deviation σ ;
2. \bar{X} , the mean of the measurements in a sample of size n ; the distribution of \bar{X} is its sampling distribution, with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$.

EXAMPLE 3

Let \bar{X} be the mean of a random sample of size 50 drawn from a population with mean 112 and standard deviation 40.

- Find the mean and standard deviation of \bar{X} .
- Find the probability that \bar{X} assumes a value between 110 and 114.
- Find the probability that \bar{X} assumes a value greater than 113.

Solution

- By the formulas in the previous section

$$\mu_{\bar{X}} = \mu = 112 \quad \text{and} \quad \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{40}{\sqrt{50}} = 5.65685$$

- Since the sample size is at least 30, the Central Limit Theorem applies: \bar{X} is approximately normally distributed. We compute probabilities using [Figure 12.2 "Cumulative Normal Probability"](#) in the usual way, just being careful to use $\sigma_{\bar{X}}$ and not σ when we standardize:

$$\begin{aligned} P(110 < \bar{X} < 114) &= P\left(\frac{110 - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < Z < \frac{114 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\ &= P\left(\frac{110 - 112}{5.65685} < Z < \frac{114 - 112}{5.65685}\right) \\ &= P(-0.35 < Z < 0.35) = 0.6368 - 0.3632 \end{aligned}$$

- Similarly

$$\begin{aligned}P(\bar{X} > 113) &= P\left(Z > \frac{113 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\&= P\left(Z > \frac{113 - 112}{5.65685}\right) \\&= P(Z > 0.18) \\&= 1 - P(Z < 0.18) = 1 - 0.5714 = 0.4286\end{aligned}$$

Note that if in [Note 6.11 "Example 3"](#) we had been asked to compute the probability that the value of a single randomly selected element of the population exceeds 113, that is, to compute the number $P(X > 113)$, we would not have been able to do so, since we do not know the distribution of X , but only that its mean is 112 and its standard deviation is 40. By contrast we could compute $P(\bar{X} > 113)$ even without complete knowledge of the distribution of X because the Central Limit Theorem guarantees that \bar{X} is approximately normal.

EXAMPLE 4

The numerical population of grade point averages at a college has mean 2.61 and standard deviation 0.5. If a random sample of size 100 is taken from the population, what is the probability that the sample mean will be between 2.51 and 2.71?

Solution

The sample mean \bar{X} has mean $\mu_{\bar{X}} = \mu = 2.61$ and standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 0.5/10 = 0.05$, so

$$\begin{aligned} P(2.51 < \bar{X} < 2.71) &= P\left(\frac{2.51 - \mu_{\bar{X}}}{\sigma_{\bar{X}}} < Z < \frac{2.71 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\ &= P\left(\frac{2.51 - 2.61}{0.05} < Z < \frac{2.71 - 2.61}{0.05}\right) \\ &= P(-2 < Z < 2) \\ &= P(Z < 2) - P(Z < -2) \\ &= 0.9772 - 0.0228 = 0.9544 \end{aligned}$$

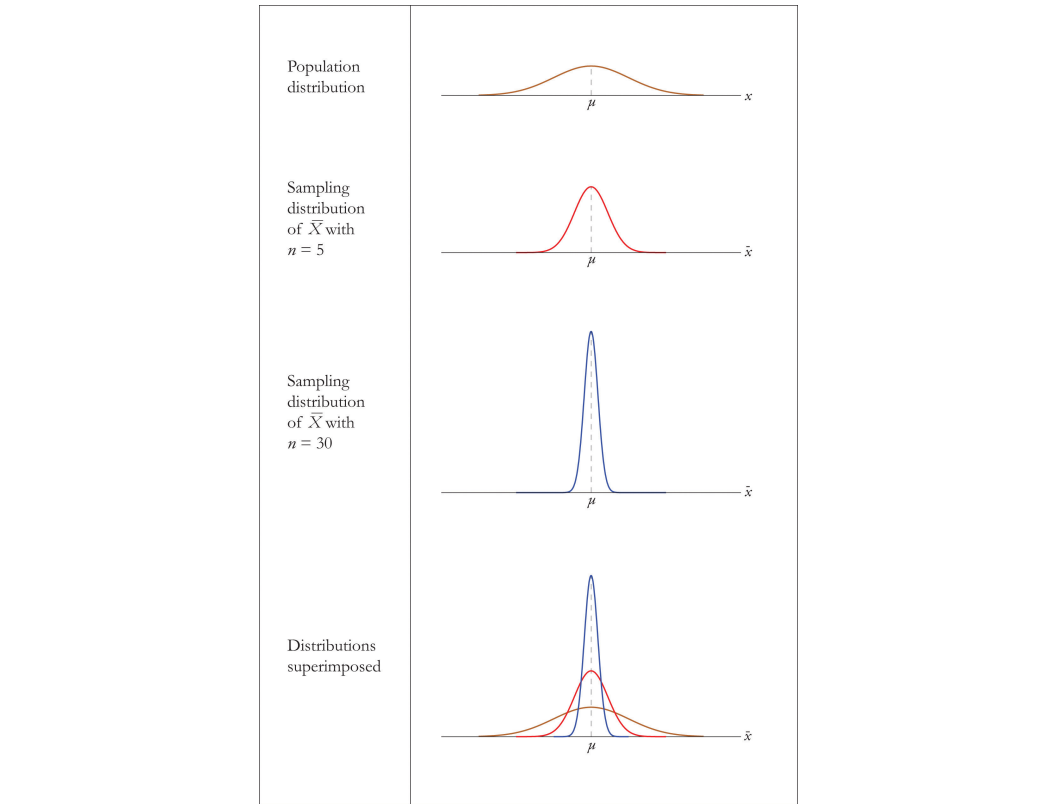
Normally Distributed Populations

The Central Limit Theorem says that no matter what the distribution of the population is, as long as the sample is “large,” meaning of size 30 or more, the sample mean is approximately normally distributed. If the population is normal to begin with then the sample mean also has a normal distribution, regardless of the sample size.

For samples of *any* size drawn from a normally distributed population, the sample mean is normally distributed, with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n}$, where n is the sample size.

The effect of increasing the sample size is shown in [Figure 6.4 "Distribution of Sample Means for a Normal Population"](#).

Figure 6.4 *Distribution of Sample Means for a Normal Population*



EXAMPLE 5

A prototype automotive tire has a design life of 38,500 miles with a standard deviation of 2,500 miles. Five such tires are manufactured and tested. On the assumption that the actual population mean is 38,500 miles and the actual population standard deviation is 2,500 miles, find the probability that the sample mean will be less than 36,000 miles. Assume that the distribution of lifetimes of such tires is normal.

Solution

For simplicity we use units of thousands of miles. Then the sample mean \bar{X} has mean $\mu_{\bar{X}} = \mu = 38.5$ and standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 2.5/\sqrt{5} = 1.11803$. Since the population is normally distributed, so is \bar{X} , hence

$$\begin{aligned} P(\bar{X} < 36) &= P\left(Z < \frac{36 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\ &= P\left(Z < \frac{36 - 38.5}{1.11803}\right) \\ &= P(Z < -2.24) = 0.0125 \end{aligned}$$

That is, if the tires perform as designed, there is only about a 1.25% chance that the average of a sample of this size would be so low.

EXAMPLE 6

An automobile battery manufacturer claims that its midgrade battery has a mean life of 50 months with a standard deviation of 6 months. Suppose the distribution of battery lives of this particular brand is approximately normal.

- On the assumption that the manufacturer's claims are true, find the probability that a randomly selected battery of this type will last less than 48 months.
- On the same assumption, find the probability that the mean of a random sample of 36 such batteries will be less than 48 months.

Solution

- Since the population is known to have a normal distribution

$$\begin{aligned} P(X < 48) &= P\left(Z < \frac{48 - \mu}{\sigma}\right) = P\left(Z < \frac{48 - 50}{6}\right) \\ &= P(Z < -0.33) = 0.3707 \end{aligned}$$

- The sample mean has mean $\mu_{\bar{X}} = \mu = 50$ and standard deviation $\sigma_{\bar{X}} = \sigma/\sqrt{n} = 6/\sqrt{36} = 1$. Thus

$$\begin{aligned} P(\bar{X} < 48) &= P\left(Z < \frac{48 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) \\ &= P\left(Z < \frac{48 - 50}{1}\right) \\ &= P(Z < -2) = 0.0228 \end{aligned}$$

KEY TAKEAWAYS

- When the sample size is at least 30 the sample mean is normally distributed.
- When the population is normal the sample mean is normally distributed regardless of the sample size.

EXERCISES

BASIC

1. A population has mean 128 and standard deviation 22.
 - a. Find the mean and standard deviation of \bar{X} for samples of size 36.
 - b. Find the probability that the mean of a sample of size 36 will be within 10 units of the population mean, that is, between 118 and 138.
2. A population has mean 1,542 and standard deviation 246.
 - a. Find the mean and standard deviation of \bar{X} for samples of size 100.
 - b. Find the probability that the mean of a sample of size 100 will be within 100 units of the population mean, that is, between 1,442 and 1,642.
3. A population has mean 73.5 and standard deviation 2.5.
 - a. Find the mean and standard deviation of \bar{X} for samples of size 30.
 - b. Find the probability that the mean of a sample of size 30 will be less than 72.
4. A population has mean 48.4 and standard deviation 6.3.
 - a. Find the mean and standard deviation of \bar{X} for samples of size 64.
 - b. Find the probability that the mean of a sample of size 64 will be less than 46.7.
5. A normally distributed population has mean 25.6 and standard deviation 3.3.
 - a. Find the probability that a single randomly selected element X of the population exceeds 30.
 - b. Find the mean and standard deviation of \bar{X} for samples of size 9.
 - c. Find the probability that the mean of a sample of size 9 drawn from this population exceeds 30.
6. A normally distributed population has mean 57.7 and standard deviation 12.1.
 - a. Find the probability that a single randomly selected element X of the population is less than 45.
 - b. Find the mean and standard deviation of \bar{X} for samples of size 16.
 - c. Find the probability that the mean of a sample of size 16 drawn from this population is less than 45.
7. A population has mean 557 and standard deviation 35.

- a. Find the mean and standard deviation of \bar{X} for samples of size 50.
 - b. Find the probability that the mean of a sample of size 50 will be more than 570.
8. A population has mean 16 and standard deviation 1.7.
- a. Find the mean and standard deviation of \bar{X} for samples of size 80.
 - b. Find the probability that the mean of a sample of size 80 will be more than 16.4.
9. A normally distributed population has mean 1,214 and standard deviation 122.
- a. Find the probability that a single randomly selected element X of the population is between 1,100 and 1,300.
 - b. Find the mean and standard deviation of \bar{X} for samples of size 25.
 - c. Find the probability that the mean of a sample of size 25 drawn from this population is between 1,100 and 1,300.
10. A normally distributed population has mean 57,800 and standard deviation 750.
- a. Find the probability that a single randomly selected element X of the population is between 57,000 and 58,000.
 - b. Find the mean and standard deviation of \bar{X} for samples of size 100.
 - c. Find the probability that the mean of a sample of size 100 drawn from this population is between 57,000 and 58,000.
11. A population has mean 72 and standard deviation 6.
- a. Find the mean and standard deviation of \bar{X} for samples of size 45.
 - b. Find the probability that the mean of a sample of size 45 will differ from the population mean 72 by at least 2 units, that is, is either less than 70 or more than 74. (Hint: One way to solve the problem is to first find the probability of the complementary event.)
12. A population has mean 12 and standard deviation 1.5.
- a. Find the mean and standard deviation of \bar{X} for samples of size 90.
 - b. Find the probability that the mean of a sample of size 90 will differ from the population mean 12 by at least 0.3 unit, that is, is either less than 11.7 or more than 12.3. (Hint: One way to solve the problem is to first find the probability of the complementary event.)

APPLICATIONS

13. Suppose the mean number of days to germination of a variety of seed is 22, with standard deviation 2.3 days. Find the probability that the mean germination time of a sample of 160 seeds will be within 0.5 day of the population mean.
14. Suppose the mean length of time that a caller is placed on hold when telephoning a customer service center is 23.8 seconds, with standard deviation 4.6 seconds. Find the probability that the mean length of time on hold in a sample of 1,200 calls will be within 0.5 second of the population mean.
15. Suppose the mean amount of cholesterol in eggs labeled “large” is 186 milligrams, with standard deviation 7 milligrams. Find the probability that the mean amount of cholesterol in a sample of 144 eggs will be within 2 milligrams of the population mean.
16. Suppose that in one region of the country the mean amount of credit card debt per household in households having credit card debt is \$15,250, with standard deviation \$7,125. Find the probability that the mean amount of credit card debt in a sample of 1,600 such households will be within \$300 of the population mean.
17. Suppose speeds of vehicles on a particular stretch of roadway are normally distributed with mean 36.6 mph and standard deviation 1.7 mph.
 - a. Find the probability that the speed X of a randomly selected vehicle is between 35 and 40 mph.
 - b. Find the probability that the mean speed \bar{X} of 20 randomly selected vehicles is between 35 and 40 mph.
18. Many sharks enter a state of tonic immobility when inverted. Suppose that in a particular species of sharks the time a shark remains in a state of tonic immobility when inverted is normally distributed with mean 11.2 minutes and standard deviation 1.1 minutes.
 - a. If a biologist induces a state of tonic immobility in such a shark in order to study it, find the probability that the shark will remain in this state for between 10 and 13 minutes.
 - b. When a biologist wishes to estimate the mean time that such sharks stay immobile by inducing tonic immobility in each of a sample of 12 sharks, find the probability that mean time of immobility in the sample will be between 10 and 13 minutes.
19. Suppose the mean cost across the country of a 30-day supply of a generic drug is \$46.58, with standard deviation \$4.84. Find the probability that the mean of a

sample of 100 prices of 30-day supplies of this drug will be between \$45 and \$50.

20. Suppose the mean length of time between submission of a state tax return requesting a refund and the issuance of the refund is 47 days, with standard deviation 6 days. Find the probability that in a sample of 50 returns requesting a refund, the mean such time will be more than 50 days.
21. Scores on a common final exam in a large enrollment, multiple-section freshman course are normally distributed with mean 72.7 and standard deviation 13.1.
 - a. Find the probability that the score X on a randomly selected exam paper is between 70 and 80.
 - b. Find the probability that the mean score \bar{X} of 38 randomly selected exam papers is between 70 and 80.
22. Suppose the mean weight of school children's bookbags is 17.4 pounds, with standard deviation 2.2 pounds. Find the probability that the mean weight of a sample of 30 bookbags will exceed 17 pounds.
23. Suppose that in a certain region of the country the mean duration of first marriages that end in divorce is 7.8 years, standard deviation 1.2 years. Find the probability that in a sample of 75 divorces, the mean age of the marriages is at most 8 years.
24. Borachio eats at the same fast food restaurant every day. Suppose the time X between the moment Borachio enters the restaurant and the moment he is served his food is normally distributed with mean 4.2 minutes and standard deviation 1.3 minutes.
 - a. Find the probability that when he enters the restaurant today it will be at least 5 minutes until he is served.
 - b. Find the probability that average time until he is served in eight randomly selected visits to the restaurant will be at least 5 minutes.

ADDITIONAL EXERCISES

25. A high-speed packing machine can be set to deliver between 11 and 13 ounces of a liquid. For any delivery setting in this range the amount delivered is normally distributed with mean some amount μ and with standard deviation 0.08 ounce. To calibrate the machine it is set to deliver a particular amount, many containers are filled, and 25 containers are randomly selected and the amount they contain is measured. Find the probability that the sample mean

will be within 0.05 ounce of the actual mean amount being delivered to all containers.

26. A tire manufacturer states that a certain type of tire has a mean lifetime of 60,000 miles. Suppose lifetimes are normally distributed with standard deviation $\sigma = 3,500$ miles.
- Find the probability that if you buy one such tire, it will last only 57,000 or fewer miles. If you had this experience, is it particularly strong evidence that the tire is not as good as claimed?
 - A consumer group buys five such tires and tests them. Find the probability that average lifetime of the five tires will be 57,000 miles or less. If the mean is so low, is that particularly strong evidence that the tire is not as good as claimed?

ANSWERS

1. a. $\mu_{\bar{X}} = 128, \sigma_{\bar{X}} = 3.67$
b. 0.9936
3. a. $\mu_{\bar{X}} = 73.5, \sigma_{\bar{X}} = 0.456$
b. 0.0005
5. a. 0.0918
b. $\mu_{\bar{X}} = 25.6, \sigma_{\bar{X}} = 1.1$
c. 0.0000
7. a. $\mu_{\bar{X}} = 557, \sigma_{\bar{X}} = 4.9497$
b. 0.0043
9. a. 0.5818
b. $\mu_{\bar{X}} = 1214, \sigma_{\bar{X}} = 24.4$
c. 0.9998
11. a. $\mu_{\bar{X}} = 72, \sigma_{\bar{X}} = 0.8944$
b. 0.0250
13. 0.9940
15. 0.9994
17. a. 0.8036
b. 1.0000
19. 0.9994
21. a. 0.2955
b. 0.8977
23. 0.9251
25. 0.9982

6.3 The Sample Proportion

LEARNING OBJECTIVES

1. To recognize that the sample proportion \hat{P} is a random variable.
2. To understand the meaning of the formulas for the mean and standard deviation of the sample proportion.
3. To learn what the sampling distribution of \hat{P} is when the sample size is large.

Often sampling is done in order to estimate the proportion of a population that has a specific characteristic, such as the proportion of all items coming off an assembly line that are defective or the proportion of all people entering a retail store who make a purchase before leaving. The population proportion is denoted p and the sample proportion is denoted \hat{p} . Thus if in reality 43% of people entering a store make a purchase before leaving, $p = 0.43$; if in a sample of 200 people entering the store, 78 make a purchase, $\hat{p} = 78/200 = 0.39$.

The sample proportion is a random variable: it varies from sample to sample in a way that cannot be predicted with certainty. Viewed as a random variable it will be written \hat{P} . It has a **mean**⁴ $\mu_{\hat{P}}$ and a **standard deviation**⁵ $\sigma_{\hat{P}}$. Here are formulas for their values.

Suppose random samples of size n are drawn from a population in which the proportion with a characteristic of interest is p . The mean $\mu_{\hat{P}}$ and standard deviation $\sigma_{\hat{P}}$ of the sample proportion \hat{P} satisfy

$$\mu_{\hat{P}} = p \quad \text{and} \quad \sigma_{\hat{P}} = \sqrt{\frac{pq}{n}}$$

where $q = 1 - p$.

4. The number about which proportions computed from samples of the same size center.

5. A measure of the variability of proportions computed from samples of the same size.

The Central Limit Theorem has an analogue for the population proportion \hat{P} . To see how, imagine that every element of the population that has the characteristic of

interest is labeled with a 1, and that every element that does not is labeled with a 0. This gives a numerical population consisting entirely of zeros and ones. Clearly the proportion of the population with the special characteristic is the proportion of the numerical population that are ones; in symbols,

$$p = \frac{\text{number of 1s}}{N}$$

But of course the sum of all the zeros and ones is simply the number of ones, so the mean μ of the numerical population is

$$\mu = \frac{\sum x}{N} = \frac{\text{number of 1s}}{N}$$

Thus the population proportion p is the same as the mean μ of the corresponding population of zeros and ones. In the same way the sample proportion \hat{p} is the same as the sample mean \bar{x} . Thus the Central Limit Theorem applies to \hat{p} . However, the condition that the sample be large is a little more complicated than just being of size at least 30.

The Sampling Distribution of the Sample Proportion

For large samples, the sample proportion is approximately normally distributed, with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{pq/n}$.

A sample is large if the interval $[p - 3\sigma_{\hat{p}}, p + 3\sigma_{\hat{p}}]$ lies wholly within the interval $[0, 1]$.

In actual practice p is not known, hence neither is $\sigma_{\hat{p}}$. In that case in order to check that the sample is sufficiently large we substitute the known quantity \hat{p} for p . This means checking that the interval

$$\left[\hat{p} - 3 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + 3 \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$$

lies wholly within the interval $[0,1]$. This is illustrated in the examples.

Figure 6.5 "Distribution of Sample Proportions" shows that when $p = 0.1$ a sample of size 15 is too small but a sample of size 100 is acceptable. **Figure 6.6 "Distribution of Sample Proportions for "** shows that when $p = 0.5$ a sample of size 15 is acceptable.

Figure 6.5 *Distribution of Sample Proportions*

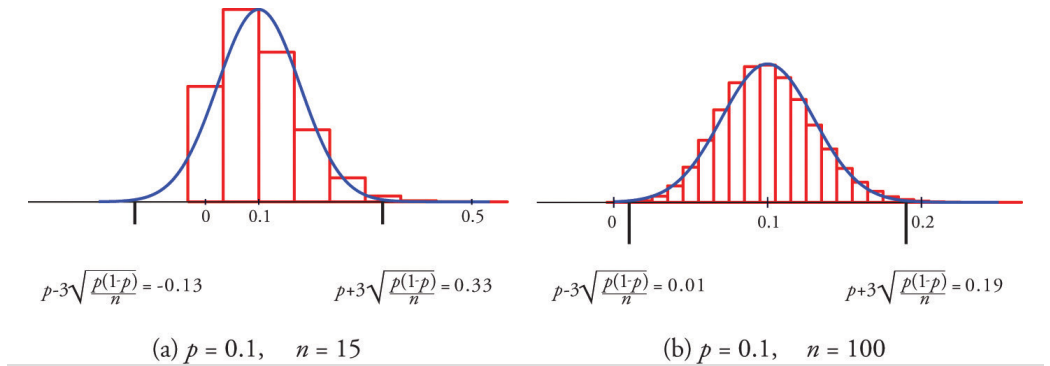
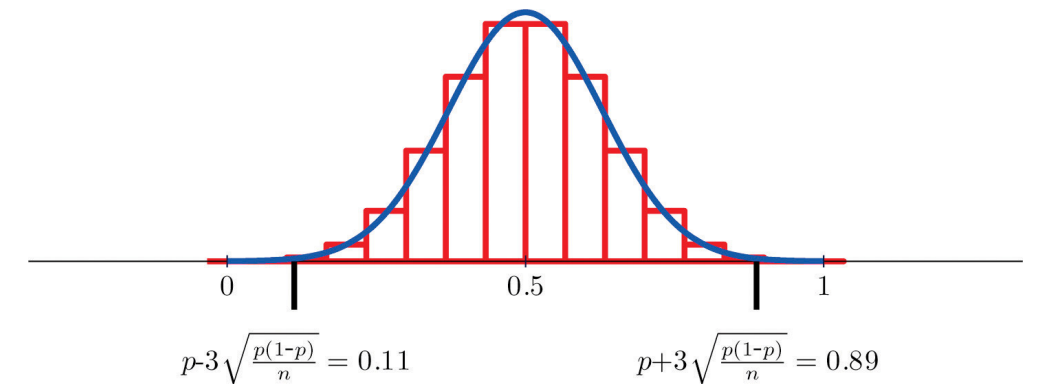


Figure 6.6 *Distribution of Sample Proportions for $p = 0.5$ and $n = 15$*



EXAMPLE 7

Suppose that in a population of voters in a certain region 38% are in favor of particular bond issue. Nine hundred randomly selected voters are asked if they favor the bond issue.

- Verify that the sample proportion \hat{P} computed from samples of size 900 meets the condition that its sampling distribution be approximately normal.
- Find the probability that the sample proportion computed from a sample of size 900 will be within 5 percentage points of the true population proportion.

Solution

- The information given is that $p = 0.38$, hence $q = 1 - p = 0.62$. First we use the formulas to compute the mean and standard deviation of \hat{P} :

$$\mu_{\hat{p}} = p = 0.38 \quad \text{and} \quad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.38)(0.62)}{900}} = 0.01618$$

$$\text{Then } 3\sigma_{\hat{p}} = 3(0.01618) = 0.04854 \approx 0.05 \quad \text{so}$$

$$[p - 3\sigma_{\hat{p}}, p + 3\sigma_{\hat{p}}] = [0.38 - 0.05, 0.38 + 0.05] = [0.33, 0.43]$$

which lies wholly within the interval $[0, 1]$, so it is safe to assume that \hat{P} is approximately normally distributed.

- To be within 5 percentage points of the true population proportion 0.38 means to be between $0.38 - 0.05 = 0.33$ and $0.38 + 0.05 = 0.43$. Thus

$$\begin{aligned}P(0.33 < \hat{P} < 0.43) &= P\left(\frac{0.33 - \mu_{\hat{P}}}{\sigma_{\hat{P}}} < Z < \frac{0.43 - \mu_{\hat{P}}}{\sigma_{\hat{P}}}\right) \\&= P\left(\frac{0.33 - 0.38}{0.01618} < Z < \frac{0.43 - 0.38}{0.01618}\right) \\&= P(-3.09 < Z < 3.09) \\&= P(3.09) - P(-3.09) \\&= 0.9990 - 0.0010 = 0.9980\end{aligned}$$

EXAMPLE 8

An online retailer claims that 90% of all orders are shipped within 12 hours of being received. A consumer group placed 121 orders of different sizes and at different times of day; 102 orders were shipped within 12 hours.

- Compute the sample proportion of items shipped within 12 hours.
- Confirm that the sample is large enough to assume that the sample proportion is normally distributed. Use $p = 0.90$, corresponding to the assumption that the retailer's claim is valid.
- Assuming the retailer's claim is true, find the probability that a sample of size 121 would produce a sample proportion so low as was observed in this sample.
- Based on the answer to part (c), draw a conclusion about the retailer's claim.

Solution

- The sample proportion is the number x of orders that are shipped within 12 hours divided by the number n of orders in the sample:

$$\hat{p} = \frac{x}{n} = \frac{102}{121} = 0.84$$

- Since $p = 0.90$, $q = 1 - p = 0.10$, and $n = 121$,

$$\sigma_{\hat{p}} = \sqrt{\frac{(0.90)(0.10)}{121}} = 0.027$$

hence

$$[p - 3\sigma_{\hat{p}}, p + 3\sigma_{\hat{p}}] = [0.90 - 0.08, 0.90 + 0.08] = [0.82, 0.98]$$

Because

$$[0.82, 0.98] \subset [0, 1],$$

it is appropriate to use the normal distribution to compute probabilities related to the sample proportion \hat{P} .

- c. Using the value of \hat{P} from part (a) and the computation in part (b),

$$\begin{aligned} P(\hat{P} \leq 0.84) &= P\left(Z \leq \frac{0.84 - \mu_{\hat{P}}}{\sigma_{\hat{P}}}\right) \\ &= P\left(Z \leq \frac{0.84 - 0.90}{0.027}\right) \\ &= P(Z \leq -2.20) = 0.0139 \end{aligned}$$

- d. The computation shows that a random sample of size 121 has only about a 1.4% chance of producing a sample proportion as the one that was observed, $\hat{p} = 0.84$, when taken from a population in which the actual proportion is 0.90. This is so unlikely that it is reasonable to conclude that the actual value of p is less than the 90% claimed.

KEY TAKEAWAYS

- The sample proportion is a random variable \hat{P} .
- There are formulas for the mean $\mu_{\hat{P}}$ and standard deviation $\sigma_{\hat{P}}$ of the sample proportion.
- When the sample size is large the sample proportion is normally distributed.

EXERCISES

BASIC

1. The proportion of a population with a characteristic of interest is $p = 0.37$. Find the mean and standard deviation of the sample proportion \hat{P} obtained from random samples of size 1,600.
2. The proportion of a population with a characteristic of interest is $p = 0.82$. Find the mean and standard deviation of the sample proportion \hat{P} obtained from random samples of size 900.
3. The proportion of a population with a characteristic of interest is $p = 0.76$. Find the mean and standard deviation of the sample proportion \hat{P} obtained from random samples of size 1,200.
4. The proportion of a population with a characteristic of interest is $p = 0.37$. Find the mean and standard deviation of the sample proportion \hat{P} obtained from random samples of size 125.
5. Random samples of size 225 are drawn from a population in which the proportion with the characteristic of interest is 0.25. Decide whether or not the sample size is large enough to assume that the sample proportion \hat{P} is normally distributed.
6. Random samples of size 1,600 are drawn from a population in which the proportion with the characteristic of interest is 0.05. Decide whether or not the sample size is large enough to assume that the sample proportion \hat{P} is normally distributed.
7. Random samples of size n produced sample proportions \hat{p} as shown. In each case decide whether or not the sample size is large enough to assume that the sample proportion \hat{P} is normally distributed.
 - a. $n = 50, \hat{p} = 0.48$
 - b. $n = 50, \hat{p} = 0.12$
 - c. $n = 100, \hat{p} = 0.12$
8. Samples of size n produced sample proportions \hat{p} as shown. In each case decide whether or not the sample size is large enough to assume that the sample proportion \hat{P} is normally distributed.
 - a. $n = 30, \hat{p} = 0.72$

- b. $n = 30, \hat{p} = 0.84$
c. $n = 75, \hat{p} = 0.84$
9. A random sample of size 121 is taken from a population in which the proportion with the characteristic of interest is $p = 0.47$. Find the indicated probabilities.
- a. $P(0.45 \leq \hat{P} \leq 0.50)$
b. $P(\hat{P} \geq 0.50)$
10. A random sample of size 225 is taken from a population in which the proportion with the characteristic of interest is $p = 0.34$. Find the indicated probabilities.
- a. $P(0.25 \leq \hat{P} \leq 0.40)$
b. $P(\hat{P} \leq 0.35)$
11. A random sample of size 900 is taken from a population in which the proportion with the characteristic of interest is $p = 0.62$. Find the indicated probabilities.
- a. $P(0.60 \leq \hat{P} \leq 0.64)$
b. $P(0.57 \leq \hat{P} \leq 0.67)$
12. A random sample of size 1,100 is taken from a population in which the proportion with the characteristic of interest is $p = 0.28$. Find the indicated probabilities.
- a. $P(0.27 \leq \hat{P} \leq 0.29)$
b. $P(0.23 \leq \hat{P} \leq 0.33)$

APPLICATIONS

13. Suppose that 8% of all males suffer some form of color blindness. Find the probability that in a random sample of 250 men at least 10% will suffer some form of color blindness. First verify that the sample is sufficiently large to use the normal distribution.
14. Suppose that 29% of all residents of a community favor annexation by a nearby municipality. Find the probability that in a random sample of 50 residents at

least 35% will favor annexation. First verify that the sample is sufficiently large to use the normal distribution.

15. Suppose that 2% of all cell phone connections by a certain provider are dropped. Find the probability that in a random sample of 1,500 calls at most 40 will be dropped. First verify that the sample is sufficiently large to use the normal distribution.
16. Suppose that in 20% of all traffic accidents involving an injury, driver distraction in some form (for example, changing a radio station or texting) is a factor. Find the probability that in a random sample of 275 such accidents between 15% and 25% involve driver distraction in some form. First verify that the sample is sufficiently large to use the normal distribution.
17. An airline claims that 72% of all its flights to a certain region arrive on time. In a random sample of 30 recent arrivals, 19 were on time. You may assume that the normal distribution applies.
 - a. Compute the sample proportion.
 - b. Assuming the airline's claim is true, find the probability of a sample of size 30 producing a sample proportion so low as was observed in this sample.
18. A humane society reports that 19% of all pet dogs were adopted from an animal shelter. Assuming the truth of this assertion, find the probability that in a random sample of 80 pet dogs, between 15% and 20% were adopted from a shelter. You may assume that the normal distribution applies.
19. In one study it was found that 86% of all homes have a functional smoke detector. Suppose this proportion is valid for all homes. Find the probability that in a random sample of 600 homes, between 80% and 90% will have a functional smoke detector. You may assume that the normal distribution applies.
20. A state insurance commission estimates that 13% of all motorists in its state are uninsured. Suppose this proportion is valid. Find the probability that in a random sample of 50 motorists, at least 5 will be uninsured. You may assume that the normal distribution applies.
21. An outside financial auditor has observed that about 4% of all documents he examines contain an error of some sort. Assuming this proportion to be accurate, find the probability that a random sample of 700 documents will contain at least 30 with some sort of error. You may assume that the normal distribution applies.
22. Suppose 7% of all households have no home telephone but depend completely on cell phones. Find the probability that in a random sample of 450

households, between 25 and 35 will have no home telephone. You may assume that the normal distribution applies.

ADDITIONAL EXERCISES

23. Some countries allow individual packages of prepackaged goods to weigh less than what is stated on the package, subject to certain conditions, such as the average of all packages being the stated weight or greater. Suppose that one requirement is that at most 4% of all packages marked 500 grams can weigh less than 490 grams. Assuming that a product actually meets this requirement, find the probability that in a random sample of 150 such packages the proportion weighing less than 490 grams is at least 3%. You may assume that the normal distribution applies.
24. An economist wishes to investigate whether people are keeping cars longer now than in the past. He knows that five years ago, 38% of all passenger vehicles in operation were at least ten years old. He commissions a study in which 325 automobiles are randomly sampled. Of them, 132 are ten years old or older.
 - a. Find the sample proportion.
 - b. Find the probability that, when a sample of size 325 is drawn from a population in which the true proportion is 0.38, the sample proportion will be as large as the value you computed in part (a). You may assume that the normal distribution applies.
 - c. Give an interpretation of the result in part (b). Is there strong evidence that people are keeping their cars longer than was the case five years ago?
25. A state public health department wishes to investigate the effectiveness of a campaign against smoking. Historically 22% of all adults in the state regularly smoked cigars or cigarettes. In a survey commissioned by the public health department, 279 of 1,500 randomly selected adults stated that they smoke regularly.
 - a. Find the sample proportion.
 - b. Find the probability that, when a sample of size 1,500 is drawn from a population in which the true proportion is 0.22, the sample proportion will be no larger than the value you computed in part (a). You may assume that the normal distribution applies.
 - c. Give an interpretation of the result in part (b). How strong is the evidence that the campaign to reduce smoking has been effective?
26. In an effort to reduce the population of unwanted cats and dogs, a group of veterinarians set up a low-cost spay/neuter clinic. At the inception of the clinic

a survey of pet owners indicated that 78% of all pet dogs and cats in the community were spayed or neutered. After the low-cost clinic had been in operation for three years, that figure had risen to 86%.

- a. What information is missing that you would need to compute the probability that a sample drawn from a population in which the proportion is 78% (corresponding to the assumption that the low-cost clinic had had no effect) is as high as 86%?
 - b. Knowing that the size of the original sample three years ago was 150 and that the size of the recent sample was 125, compute the probability mentioned in part (a). You may assume that the normal distribution applies.
 - c. Give an interpretation of the result in part (b). How strong is the evidence that the presence of the low-cost clinic has increased the proportion of pet dogs and cats that have been spayed or neutered?
27. An ordinary die is “fair” or “balanced” if each face has an equal chance of landing on top when the die is rolled. Thus the proportion of times a three is observed in a large number of tosses is expected to be close to $1/6$ or $0.\overline{16}$. Suppose a die is rolled 240 times and shows three on top 36 times, for a sample proportion of 0.15.
- a. Find the probability that a fair die would produce a proportion of 0.15 or less. You may assume that the normal distribution applies.
 - b. Give an interpretation of the result in part (b). How strong is the evidence that the die is not fair?
 - c. Suppose the sample proportion 0.15 came from rolling the die 2,400 times instead of only 240 times. Rework part (a) under these circumstances.
 - d. Give an interpretation of the result in part (c). How strong is the evidence that the die is not fair?

ANSWERS

1. $\mu_{\hat{p}} = 0.37, \sigma_{\hat{p}} = 0.012$
3. $\mu_{\hat{p}} = 0.76, \sigma_{\hat{p}} = 0.012$
5. $p \pm 3\sqrt{\frac{pq}{n}} = 0.25 \pm 0.087$, yes
7.
 - a. $\hat{p} \pm 3\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.48 \pm 0.21$, yes
 - b. $\hat{p} \pm 3\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.12 \pm 0.14$, no
 - c. $\hat{p} \pm 3\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.12 \pm 0.10$, yes
9.
 - a. 0.4154
 - b. 0.2546
11.
 - a. 0.7850
 - b. 0.9980
13. $p \pm 3\sqrt{\frac{pq}{n}} = 0.08 \pm 0.05$
and
 $[0.03, 0.13] \subset [0, 1], 0.1210$
15. $p \pm 3\sqrt{\frac{pq}{n}} = 0.02 \pm 0.01$
and
 $[0.01, 0.03] \subset [0, 1], 0.9671$
17.
 - a. 0.63
 - b. 0.1446
19. 0.9977
21. 0.3483
23. 0.7357
25.
 - a. 0.186
 - b. 0.0007

- c. In a population in which the true proportion is 22% the chance that a random sample of size 1500 would produce a sample proportion of 18.6% or less is only 7/100 of 1%. This is strong evidence that currently a smaller proportion than 22% smoke.
- 27.
- a. 0.2451
 - b. We would expect a sample proportion of 0.15 or less in about 24.5% of all samples of size 240, so this is practically no evidence at all that the die is not fair.
 - c. 0.0139
 - d. We would expect a sample proportion of 0.15 or less in only about 1.4% of all samples of size 2400, so this is strong evidence that the die is not fair.