Chapter 5: Sampling Distributions
Section 1: The Concept of a Sampling Distribution

5.1 The probability distribution shown here describes a population of measurements that can assume values of 0, 2, 4, and 6, each of which occurs with the same relative frequency:

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>2</th>
<th>4</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(x)</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
<td>$\frac{1}{4}$</td>
</tr>
</tbody>
</table>

a. List all the different samples of $n = 2$ measurements that can be selected from this population.
b. Calculate the mean of each different sample listed in part a.
c. If a sample of $n = 2$ measurements is randomly selected from the population, what is the probability that a specific sample will be selected?
d. Assume that a random sample of $n = 2$ measurements is selected from the population. List the different values of $\bar{x}$ found in part b and find the probability of each. Then give the sampling distribution of the sample mean $\bar{x}$ in tabular form.
e. Construct a probability histogram for the sampling distribution of $\bar{x}$.

5.3 Consider the population described by the probability distribution shown below.

<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(x)</td>
<td>.2</td>
<td>.3</td>
<td>.2</td>
<td>.2</td>
<td>.1</td>
</tr>
</tbody>
</table>

The random variable $x$ is observed twice. If these observations are independent, verify that the different samples of size 2 and their probabilities are as shown in the next column.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Probability</th>
<th>Sample</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 1</td>
<td>.04</td>
<td>3, 4</td>
<td>.04</td>
</tr>
<tr>
<td>1, 2</td>
<td>.06</td>
<td>3, 5</td>
<td>.02</td>
</tr>
<tr>
<td>1, 3</td>
<td>.04</td>
<td>4, 1</td>
<td>.04</td>
</tr>
<tr>
<td>1, 4</td>
<td>.04</td>
<td>4, 2</td>
<td>.06</td>
</tr>
<tr>
<td>1, 5</td>
<td>.02</td>
<td>4, 3</td>
<td>.04</td>
</tr>
<tr>
<td>2, 1</td>
<td>.06</td>
<td>4, 4</td>
<td>.04</td>
</tr>
<tr>
<td>2, 2</td>
<td>.09</td>
<td>4, 5</td>
<td>.02</td>
</tr>
<tr>
<td>2, 3</td>
<td>.06</td>
<td>5, 1</td>
<td>.02</td>
</tr>
<tr>
<td>2, 4</td>
<td>.06</td>
<td>5, 2</td>
<td>.03</td>
</tr>
<tr>
<td>2, 5</td>
<td>.03</td>
<td>5, 3</td>
<td>.02</td>
</tr>
<tr>
<td>3, 1</td>
<td>.04</td>
<td>5, 4</td>
<td>.02</td>
</tr>
<tr>
<td>3, 2</td>
<td>.06</td>
<td>5, 5</td>
<td>.01</td>
</tr>
<tr>
<td>3, 3</td>
<td>.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Find the sampling distribution of the sample mean $\bar{x}$.
b. Construct a probability histogram for the sampling distribution of $\bar{x}$.
c. What is the probability that $\bar{x}$ is 4.5 or larger?
d. Would you expect to observe a value of $\bar{x}$ equal to 4.5 or larger? Explain.

5.4 Refer to Exercise 5.3 and find $E(x) = \mu$. Then use the sampling distribution of $\bar{x}$ found in Exercise 5.3 to find the expected value of $\bar{x}$. Note that $E(\bar{x}) = \mu$. 
Section 2: Properties of Sampling Distributions: Unbiasedness and Minimum Variance

5.8 Consider the following probability distribution:

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>1</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>p(x)</td>
<td>1/6</td>
<td>1/3</td>
<td>1/3</td>
</tr>
</tbody>
</table>

a. Find μ and σ².

b. Find the sampling distribution of the sample mean \( \bar{x} \)
   for a random sample of \( n = 2 \) measurements from this distribution.

c. Show that \( \bar{x} \) is an unbiased estimator of \( \mu \). [Hint: Show that \( E(\bar{x}) = \sum x p(x) = \mu \).]
Section 3: The Sampling Distribution of the Sample Mean and the Central Limit Theorem

5.15 Will the sampling distribution of $\bar{x}$ always be approximately normally distributed? Explain.

5.16 Suppose a random sample of $n = 25$ measurements is selected from a population with mean $\mu$ and standard deviation $\sigma$. For each of the following values of $\mu$ and $\sigma$, give the values of $\mu_X$ and $\sigma_X$.
   a. $\mu = 10, \sigma = 3$
   b. $\mu = 100, \sigma = 25$
   c. $\mu = 20, \sigma = 40$
   d. $\mu = 10, \sigma = 100$

5.18 A random sample of $n = 64$ observations is drawn from a population with a mean equal to 20 and a standard deviation equal to 16.
   a. Give the mean and standard deviation of the (repeated) sampling distribution of $\bar{x}$.
   b. Describe the shape of the sampling distribution of $\bar{x}$.
      Does your answer depend on the sample size?
   c. Calculate the standard normal $z$-score corresponding to a value of $\bar{x} = 15.5$.
   d. Calculate the standard normal $z$-score corresponding to $\bar{x} = 23$.

5.19 Refer to Exercise 5.18. Find the probability that
   a. $\bar{x}$ is less than 16.
   b. $\bar{x}$ is greater than 23.
   c. $\bar{x}$ is greater than 25.
   d. $\bar{x}$ falls between 16 and 22.
   e. $\bar{x}$ is less than 14.

5.24 Salary of a travel management professional. According to a National Business Travel Association (NBTA) 2010 survey, the average salary of a travel management professional is $96,850. Assume that the standard deviation of such salaries is $30,000. Consider a random sample of 50 travel management professionals and let $\bar{x}$ represent the mean salary for the sample.
   a. What is $\mu_x$?
   b. What is $\sigma_x$?
   c. Describe the shape of the sampling distribution of $\bar{x}$.
   d. Find the $z$-score for the value $\bar{x} = 89,500$.
   e. Find $P(\bar{x} > 89,500)$.

5.26 Critical-part failures in NASCAR vehicles. Refer to The Sport Journal (Winter 2007) analysis of critical-part failures at NASCAR races, Exercise 4.144 (p. 248). Recall that researchers found that the time $x$ (in hours) until the first critical-part failure is exponentially distributed with $\mu = .10$ and $\sigma = .10$. Now consider a random sample of $n = 50$ NASCAR races and let $\bar{x}$ represent the sample mean time until the first critical-part failure.
   a. Find $E(\bar{x})$ and $\text{Var}(\bar{x})$.
   b. Although $x$ has an exponential distribution, the sampling distribution of $\bar{x}$ is approximately normal. Why?
   c. Find the probability that the sample mean time until the first critical-part failure exceeds 0.13 hour.
5.27 **Tomato as a taste modifier.** Miraculin is a protein naturally produced in a rare tropical fruit that can convert a sour taste into a sweet taste. Refer to the *Plant Science* (May 2010) investigation of the ability of a hybrid tomato plant to produce miraculin, Exercise 4.97 (p. 233). Recall that the amount $x$ of miraculin produced in the plant had a mean of 105.3 micrograms per gram of fresh weight with a standard deviation of 8.0. Consider a random sample of $n = 64$ hybrid tomato plants and let $\bar{x}$ represent the sample mean amount of miraculin produced. Would you expect to observe a value of $\bar{x}$ less than 103 micrograms per gram of fresh weight? Explain.

5.31 **Is exposure to a chemical in Teflon-coated cookware hazardous?** Perfluorooctanoic acid (PFOA) is a chemical used in Teflon-coated cookware to prevent food from sticking. The EPA is investigating the potential risk of PFOA as a cancer-causing agent (*Science News Online*, August 27, 2005). It is known that the blood concentration of PFOA in people in the general population has a mean of $\mu = 6$ parts per billion (ppb) and a standard deviation of $\sigma = 10$ ppb. *Science News Online* reported on tests for PFOA exposure conducted on a sample of 326 people who live near DuPont’s Teflon-making Washington (West Virginia) Works facility.

a. What is the probability that the average blood concentration of PFOA in the sample is greater than 75 ppb?

b. The actual study resulted in $\bar{x} = 300$ ppb. Use this information to make an inference about the true mean $\mu$ PFOA concentration for the population of people who live near DuPont’s Teflon facility.

5.33 **Motivation of drug dealers.** Refer to the *Applied Psychology in Criminal Justice* (Sep. 2009) investigation of the personality characteristics of drug dealers, Exercise 2.80 (p. 85). Convicted drug dealers were scored on the Wanting Recognition (WR) Scale—a scale that provides a quantitative measure of a person’s level of need for approval and sensitivity to social situations. (Higher scores indicate a greater need for approval.) Based on the study results, we can assume that the WR scores for the population of convicted drug dealers has a mean of 40 and a standard deviation of 5. Suppose that in a sample of 100 people, the mean WR scale score is $\bar{x} = 42$. Is this sample likely to have been selected from the population of convicted drug dealers? Explain.
Section 4: The Sampling Distribution of the Sample Proportion

5.37 Suppose a random sample of $n = 500$ measurements is selected from a binomial population with probability of success $p$. For each of the following values of $p$, give the mean and standard deviation of the sampling distribution of the sample proportion, $\hat{p}$.
   a. $p = .1$
   b. $p = .5$
   c. $p = .7$

5.39 A random sample of $n = 250$ measurements is drawn from a binomial population with probability of success $.85$.
   a. Find $E(\hat{p})$ and $\sigma_{\hat{p}}$.
   b. Describe the shape of the sampling distribution of $\hat{p}$.
   c. Find $P(\hat{p} < .9)$.

5.43 Paying for music downloads. According to a recent Pew Internet & American Life Project Survey (October 2010), 67% of adults who use the Internet have paid to download music. (See Exercise 2.4, p. 48.) In a random sample of $n = 1,000$ adults who use the Internet, let $\hat{p}$ represent the proportion who have paid to download music.
   a. Find the mean of the sampling distribution of $\hat{p}$.
   b. Find the standard deviation of the sampling distribution of $\hat{p}$.
   c. What does the Central Limit Theorem say about the shape of the sampling distribution of $\hat{p}$?
   d. Compute the probability that $\hat{p}$ is less than .75.
   e. Compute the probability that $\hat{p}$ is greater than .50.

5.44 Working on summer vacation. According to an Adweek/ Harris (July 2011) poll of U.S. adults, about 45% work during their summer vacation. (See Exercise 3.14, p. 141.) Assume that the true proportion of all U.S. adults that work during summer vacation is $p = .45$. Now consider a random sample of 500 U.S. adults.
   a. What is the probability that between 40% and 50% of the sampled adults work during summer vacation?
   b. What is the probability that over 60% of the sampled adults work during summer vacation?

5.48 Hotel guest satisfaction. Refer to the results of the 2009 North American Hotel Guest Satisfaction Index Study referenced in Exercise 4.48 (p. 210). Recall that 66% of hotel guests were aware of the hotel’s “green” conservation program; of these guests, 72% actually participated in the program by reusing towels and bed linens. In a random sample of 100 hotel guests, find the probability that fewer than 42 were aware of and participated in the hotel’s conservation efforts.

5.49 Fingerprint expertise. Refer to the Psychological Science (August 2011) study of fingerprint identification, Exercise 4.53 (p. 210). Recall that when presented with prints from the same individual, a fingerprint expert will correctly identify the match 92% of the time. Consider a forensic database of 1,000 different pairs of fingerprints, where each pair is a match.
   a. What proportion of the 1,000 pairs would you expect an expert to correctly identify as a match?
   b. What is the probability that an expert will correctly identify fewer than 900 of the fingerprint matches?