Overview

- 7.1 Introduction to Sampling Distributions
- 7.2 Central Limit Theorem for Means
- 7.3 Central Limit Theorem for Proportions
7.1 Introduction to Sampling Distributions

Objectives:
By the end of this section, I will be able to...

1) Compute point estimates and sampling error.
2) Explain the sampling distribution of the sample mean $\bar{x}$
3) Describe the sampling distribution of the sample mean $\bar{x}$ when the population is normal.
4) Use normal probability plots to assess normality.
5) Find probabilities and percentiles for the sample mean when the population is normal.
Point Estimates

- Use known statistics to estimate unknown parameters and report a single number as the estimate

- The value of the statistic is called the **point estimate**

<table>
<thead>
<tr>
<th>Sample statistic</th>
<th>( \ldots ) estimates ( \ldots )</th>
<th>Population parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>( \bar{x} )</td>
<td>( \mu )</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>( s )</td>
<td>( \sigma )</td>
</tr>
<tr>
<td>Proportion</td>
<td>( \hat{p} )</td>
<td>( p )</td>
</tr>
</tbody>
</table>

Table 7.1 Point estimation: Use statistics to estimate unknown population parameters
Sampling error

- The distance between the point estimate and its target parameter

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Sample statistic</th>
<th>Population parameter</th>
<th>Sampling error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>$\bar{x}$</td>
<td>$\mu$</td>
<td>$</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>$s$</td>
<td>$\sigma$</td>
<td>$</td>
</tr>
<tr>
<td>Proportion</td>
<td>$\hat{p}$</td>
<td>$p$</td>
<td>$</td>
</tr>
</tbody>
</table>

Table 7.3 Sampling error for common characteristics
Example

Page 349

PRACTICING THE TECHNIQUES
For Exercises 7–12, find the sampling error.
7. \( \mu = 100, \bar{x} = 95 \)
8. \( \mu = 0, \bar{x} = 1.5 \)
9. \( \sigma = 20, s = 15 \)
10. \( \sigma = 1, s = 0.9 \)
11. \( p = 0.25, \hat{p} = 0.30 \)
12. \( p = 0.75, \hat{p} = 0.70 \)

Do problems 10 and 12
Example

Solutions:

10) \[ |s - \sigma| = |0.9 - 1.0| = 0.1 \]

12) \[ |\hat{p} - p| = |0.70 - 0.75| = 0.05 \]
Example 7.3 - Commuting times for student government members

We are interested in how long it takes the five members of the student government to commute to school. The times (in minutes) are given in Table 7.4. Since these five people are all the members of the student government, we can consider them to constitute a population.
Example 7.3 continued

Table 7.4 Commuting times for the five members of the student government

<table>
<thead>
<tr>
<th></th>
<th>Amber</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Dave</th>
<th>Emma</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10</td>
<td>20</td>
<td>5</td>
<td>30</td>
<td>15</td>
</tr>
</tbody>
</table>

a. Calculate the population mean commuting time $\mu$.

b. Take a sample of the following student government members: Amber, Brandon, and Chantal. Find the sample mean commuting time $\bar{x}$ and the sampling error $|\bar{x} - \mu|$. 
The mean commuting time of this population is

\[ \mu = \frac{\Sigma x}{N} = \frac{10 + 20 + 5 + 30 + 15}{5} = 16 \text{ minutes} \]

For Amber, Brandon, and Chantal, the sample mean commuting time is

\[ \bar{x}_1 = \frac{\Sigma x}{N} = \frac{10 + 20 + 5}{3} \approx 11.67 \text{ minutes} \]

The sampling error of this sample mean is

\[ |\bar{x}_1 - \mu| = |11.67 - 16| = 4.33 \text{ minutes}. \]
Consider previous example and find **all** possible samples of student government members, the sample means, population mean, and sampling errors (page 339):

<table>
<thead>
<tr>
<th>Sample</th>
<th>Amber</th>
<th>Brandon</th>
<th>Amber</th>
<th>Chantal</th>
<th>Amber</th>
<th>Dave</th>
<th>Amber</th>
<th>Emma</th>
<th>Amber</th>
<th>Chantal</th>
<th>Dave</th>
<th>Emma</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Dave</th>
<th>Emma</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Dave</th>
<th>Emma</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Dave</th>
<th>Emma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>5</td>
<td>5</td>
<td>30</td>
<td>5</td>
<td>5</td>
<td>30</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>30</td>
<td>15</td>
<td>30</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bar{x}$</td>
<td>11.67</td>
<td>20</td>
<td>15</td>
<td>15</td>
<td>10</td>
<td>10</td>
<td>18.33</td>
<td>18.33</td>
<td>13.33</td>
<td>21.67</td>
<td>16.67</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$</td>
<td>\bar{x} - \mu</td>
<td>$</td>
<td>4.33</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>2.33</td>
<td>2.33</td>
<td>2.67</td>
<td>5.67</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 7.5** All possible samples of size 3 from population of student government members
Sampling distribution of the sample mean $\bar{x}$

- Collection of the sample means of all possible samples of size $n$
Mean of the Sample Means

Calculate the mean of the sample means as the average of the sample means:

The mean value of the ten sample means is

\[
\frac{11.67 + 20 + 15 + 15 + 10 + 18.33 + 18.33 + 13.33 + 21.67 + 16.67}{10} = 16
\]

Note: the mean of the sample means equals the population mean in this example.
Fact 1

- The mean of the sampling distribution of the sample mean $\bar{x}$ is the value of the population mean $\mu$.

- Denoted as $\mu_{\bar{x}} = \mu$

- Read as “the mean of the sampling distribution of $\bar{x}$ is $\mu$”
Mean of the Sample Means

For the previous example

$$\mu_x = \frac{11.67 + 20 + 15 + 15 + 10 + 18.33 + 18.33 + 13.33 + 21.67 + 16.67}{10}$$
Standard Deviation of the Population

<table>
<thead>
<tr>
<th></th>
<th>Amber</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Dave</th>
<th>Emma</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>5</td>
<td>30</td>
<td>15</td>
<td></td>
</tr>
</tbody>
</table>

\[
\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}} = \sqrt{\frac{\sum[(10 - 16)^2 + (20 - 16)^2 + (5 - 16)^2 + (30 - 16)^2 + (15 - 16)^2]}{5}} \approx 8.6023
\]
# Standard Deviation of the Sample Means

<table>
<thead>
<tr>
<th>Sample</th>
<th>Amber</th>
<th>Brandon</th>
<th>Amber</th>
<th>Brandon</th>
<th>Amber</th>
<th>Chantal</th>
<th>Amber</th>
<th>Chantal</th>
<th>Amber</th>
<th>Dave</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Amber</th>
<th>Dave</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Amber</th>
<th>Dave</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>5</td>
<td>5</td>
<td>30</td>
<td>5</td>
<td>30</td>
<td>5</td>
<td>5</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>30</td>
<td>15</td>
<td>30</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bar{x}$</td>
<td>11.67</td>
<td>20</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>10</td>
<td>18.33</td>
<td>18.33</td>
<td>13.33</td>
<td>21.67</td>
<td>16.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu$</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$</td>
<td>\bar{x} - \mu</td>
<td>_l$</td>
<td>4.33</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>2.33</td>
<td>2.33</td>
<td>2.67</td>
<td>5.67</td>
<td>0.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\sqrt{\frac{\sum (\bar{x} - \mu)^2}{10}} = \sqrt{\frac{\sum [(11.67 - 16)^2 + (20 - 16)^2 + \ldots + (16.67 - 16)^2]}{10}}
\]

= 3.5119
Fact 2

- The standard deviation of the sampling distribution of the sample mean $\bar{x}$ is

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

- $\sigma$ is the population standard deviation

- $n$ is the sample size where $n$ is assumed to be very large (if $n$ is not large, see the note on page 340)
Previous example

\[ \sigma_{\bar{x}} = \sqrt{\frac{\sum (\bar{x} - \mu)^2}{10}} = \sqrt{\frac{\sum [(11.67 - 16)^2 + (20 - 16)^2 + \ldots + (16.67 - 16)^2]}{10}} \]
34. Student Heights. The heights of a population of students have a mean of 68 inches (5 feet 8 inches) and a standard deviation of 3 inches. For each of the following sample sizes, find $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$.
   a. Sample size $n = 10$ students
   b. Sample size $n = 100$ students
   c. Sample size $n = 1000$ students
Example

Solutions

a)

\[ \mu_x = \mu = 68 \text{ inches} \]

\[ \sigma_x = \frac{\sigma}{\sqrt{n}} = 3 \text{ inches}/\sqrt{10} = 0.95 \text{ inches} \]
Example

Solutions

b) \[ \mu_x = \mu = 68 \text{ inches} \]

\[ \sigma_{x} = \frac{\sigma}{\sqrt{n}} = \frac{3 \text{ inches}}{\sqrt{100}} = 0.30 \text{ inches} \]
Example

Solutions

c) \[
\mu_x = \mu = 68 \text{ inches}
\]

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 3 \text{ inches} / \sqrt{1000} = 0.09 \text{ inches}
\]
Sampling Distribution of the Sample Mean for a Normal Population

Fact 3

- Is itself normal, regardless of sample size
Fact 4

- Distributed as normal with mean:

\[ \mu_{\bar{x}} = \mu \]

- And standard deviation:

\[ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \]
Fact 5: Standardizing a Normal Sampling Distribution for Means

- When the sampling distribution of $\overline{x}$ is normal, we may standardize to produce the standard normal random variable $Z$ as follows:

$$Z = \frac{\overline{x} - \mu}{\sigma_{\overline{x}}} = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$$

where $\mu$ is the population mean, $\sigma$ is the population standard deviation, and $n$ is the sample size.
38. Teacher Salaries. Suppose the salaries of teachers in your city are normally distributed with a mean of $50,000 and a standard deviation of $5000. Suppose we take samples of size 25 teachers.
   a. Find the probability that the sample mean salary will exceed $52,000.
   b. Find the probability that the sample mean salary will be less than $47,000.
   c. Find the probability that the sample mean salary will be between $52,000 and $53,000.
Example

Solutions:

\[ \mu_x = \mu = \$50,000 \]

\[ \sigma_x = \sigma / \sqrt{n} = \$5000/ \sqrt{25} = \$1000 \]
Example

Solutions:

a) Z-score for sample mean of 52,000 is

\[ Z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{52000 - 50000}{1000} = 2 \]
Example

Solutions:

a) \[ P(\bar{x} > $50,000) = P(Z > 2) \]
\[ = 1 - P(Z < 2) \]
\[ = 1 - 0.9772 \]
\[ = 0.0228 \]

Look up in Table C
Example

(b) calculator

TI-83/84
Finding Areas or Probabilities for Any Normal Distribution

Step 1  Press 2nd, then DISTR (the VARS key).
Step 2  Press 2 to choose normalcdf.
Step 3  On the home screen, enter the smaller value of X, comma, the larger value of X, comma, the mean of X, comma, the standard deviation of X, then close parenthesis.
Step 4  Press ENTER.
Example

(c) calculator

\[ P(\bar{x} < \$47,000) \]
\[ = \text{normalcdf}(-10^{99}, 47, 0, 0, 50000, 1000) \]
\[ = 0.0013 \]
Example

(c) calculator

**TI-83/84**

Finding Areas or Probabilities for Any Normal Distribution

*Step 1* Press 2\(^{nd}\), then DISTR (the VARS key).

*Step 2* Press 2 to choose `normalcdf`.

*Step 3* On the home screen, enter the smaller value of \(X\), comma, the larger value of \(X\), comma, the mean of \(X\), comma, the standard deviation of \(X\), then close parenthesis. See Figure 6.51 (page 316).

*Step 4* Press ENTER.
Example

(c) calculator

\[ P($52,000 < \bar{x} < $53,000) = \text{normalcdf}(52000,53000,50000,1000) = 0.0215 \]
40. Teacher Salaries. For the teacher salaries in Exercise 38:
   a. Find the median sample mean salary. How does it compare to the population mean?
   b. Find the 95th percentile of sample mean salaries.
   c. Find the 5th percentile of sample mean salaries.
   d. Between which two values do the middle 90% of sample mean salaries lie?

Do part (a)
Example

Solution:

a) $50,000 (note: for a normal distribution, the mean and the median are equal)
Example

Page 350

40. Teacher Salaries. For the teacher salaries in Exercise 38:
   a. Find the median sample mean salary. How does it compare to the population mean?
   b. Find the 95th percentile of sample mean salaries.
   c. Find the 5th percentile of sample mean salaries.
   d. Between which two values do the middle 90% of sample mean salaries lie?

Do part (b)
Example

Solutions:

b) Find \( \bar{X}_c \) so that

\[
P(\bar{X} < \bar{X}_c) = 0.95
\]

for a normal distribution with

\[
\mu_x = \mu = \$50,000
\]

\[
\sigma_{\bar{X}} = \sigma / \sqrt{n} = \$5000 / \sqrt{25} = \$1000
\]
First find $Z_c$ so that

$$P(Z < Z_c) = 0.95$$

using the standard normal distribution. From Table C we get that:

$$Z_c = 1.655$$
Example

Use $Z_c$ to get $X_c$ by using the z-score formula:

$$Z_c = \frac{X_c - \mu}{\sigma} \quad \iff \quad 1.655 = \frac{X_c - \$50,000}{\$1000}$$

$$X_c = \$51,655$$
Example

Solutions (directly with calculator):

b)

\[ \bar{X}_C = \text{invNorm}(0.95, 50000, 1000) = 51,644.85 \]
Example

Page 350

40. Teacher Salaries. For the teacher salaries in Exercise 38:
   a. Find the median sample mean salary. How does it compare to the population mean?
   b. Find the 95th percentile of sample mean salaries.
   c. Find the 5th percentile of sample mean salaries.
   d. Between which two values do the middle 90% of sample mean salaries lie?

Do parts (c) and (d)
Example

Solutions:

c) $48,355

d) $48,355 and $51,645

\[
\bar{X}_C = \text{invNorm}(0.05, 50000, 1000) = $48,355.15
\]

$48,355 and $51,645
A normal probability plot is a scatterplot of the estimated cumulative normal probabilities (expressed as percents) against the corresponding data values in a data set.
FIGURE 7.4 Normal probability plot of normal data.
Analyzing Normal Probability Plots

- If the points in the normal probability plot either cluster around a straight line or nearly all fall within the curved bounds, then it is likely that the data set is normal.

- Systematic deviations off the straight line are evidence against the claim that the data set is normal.
Normal Probability Plot

FIGURE 7.5 Normal probability plot of right-skewed data.
We can use sample statistics as point estimates of the unknown population parameters.

For each statistic, sampling error is the distance between the point estimate and its target parameter.

The sampling distribution of the sample mean \( \bar{x} \) for a given sample size \( n \) consists of the collection of the sample means of all possible samples of size \( n \) from the population.
Summary

- The mean of the sampling distribution of the sample mean $\bar{x}$ is the value of the population mean $\mu$ (Fact 1).

- The standard deviation of the sampling distribution of the sample mean $\bar{x}$ is $\sigma_{\bar{x}} = \sigma / \sqrt{n}$, $\sigma$ is the population standard deviation (Fact 2).

- The sampling distribution of the sample mean for a normal population is itself normal, regardless of sample size (Fact 3).
Summary

- For a normal population, the sampling distribution of the sample mean $\overline{x}$ is distributed as normal $(\mu, \sigma/\sqrt{n})$, where $\mu$ is the population mean and $\sigma$ is the population standard deviation (Fact 4).

- Normal probability plots are used to assess the normality of a data set.

- We can use Fact 4 to find probabilities and percentiles using sample means.
7.2 Central Limit Theorem for Means

Objectives:
By the end of this section, I will be able to...

1) Describe the sampling distribution of $\bar{x}$ for skewed and symmetric populations as the sample size increases.

2) Apply the Central Limit Theorem for Means to solve probability questions about the sample mean.
Main Idea

In this section we want to be able to describe the shape of the distribution of the sample means.
Symmetric Populations

For a symmetric distribution, at n=20 the sampling distribution of the mean is approximately normal.
Example

Roll a fair die once. Make up a table that represents the probability distribution of $X=$number on the die. Also, plot the probability distribution in a bar chart.
Example

Distribution (table format) is:

<table>
<thead>
<tr>
<th>$x$</th>
<th>$P(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1667</td>
</tr>
<tr>
<td>2</td>
<td>0.1667</td>
</tr>
<tr>
<td>3</td>
<td>0.1667</td>
</tr>
<tr>
<td>4</td>
<td>0.1667</td>
</tr>
<tr>
<td>5</td>
<td>0.1667</td>
</tr>
<tr>
<td>6</td>
<td>0.1667</td>
</tr>
</tbody>
</table>

$$\mu = \sum xP(x) = 3.5$$
FIGURE 7.11 Distribution of a single fair die roll is symmetric.
Example

Roll a fair die one hundred times. Take random samples of size 10 from these 100 rolls. For each sample of size 10, calculate the sample mean. Plot the distribution of the means as a histogram.
FIGURE 7.12 Sample means of size $n = 10$: already approaching normality.
Example

Roll a fair die one hundred times. Take random samples of size 20 from these 100 rolls. For each sample of size 20, calculate the sample mean. Plot the distribution of the means as a histogram.
FIGURE 7.13 Sample means of size $n = 20$: approximately normal.
FIGURE 7.14 Normal probability plot for $n = 20$: acceptable normality.
Skewed Populations

- For a skewed population, sampling distribution of the mean becomes approximately normal as the sample size approaches 30.
FIGURE 7.10 Sampling distribution of $x$ and normal probability plots for $n = 10, 20,$ and $30$. 

(a) Sample means for samples of size $n = 10$
Still very skewed

(b) Sample means for samples of size $n = 20$
Still somewhat skewed

(c) Sample means for samples of size $n = 30$
Approximately normal
Central Limit Theorem for Means

- Population with mean $\mu$

- Standard deviation $\sigma$

- The sampling distribution of the sample mean $\bar{x}$ becomes approximately normal with mean $\mu$ and standard deviation $\frac{\sigma}{\sqrt{n}}$ as the sample size gets larger

- Regardless of the shape of the population.
Rule of Thumb

- We consider $n \geq 30$ as large enough to apply the Central Limit Theorem for any population.
Three Cases for the Sampling Distribution of the Sample Mean $\bar{x}$

Case 1

- The population is normal.
- Then the sampling distribution of $\bar{x}$ is normal (Fact 3 from 7.1).
Three Cases for the Sampling Distribution of the Sample Mean $\bar{x}$ continued

Case 2

- The population is either non-normal or of unknown distribution and the sample size is at least 30.

- Apply Central Limit Theorem for Means: 
  \textit{The sampling distribution of the sample mean is approximately normal}
Three Cases for the Sampling Distribution of the Sample Mean $\bar{x}$ continued

Case 3

- The population is either non-normal or of unknown distribution and the sample size is less than 30.

- Insufficient information to conclude that the sampling distribution of the sample mean $\bar{x}$ is either normal or approximately normal.
PRACTICING THE TECHNIQUES
In Exercises 5–12, samples are taken. Determine which of the three cases for the sampling distribution of the sample mean applies.

6. SAT scores are not normally distributed and the sample size is large.
7. Accountant incomes are normally distributed and the sample size is small.
8. Accountant incomes are not normally distributed and the sample size is small.
9. Systolic blood pressure readings are not normally distributed and the sample size is large.
10. Systolic blood pressure readings are normally distributed and the sample size is small.
11. Body mass indices of college students are not normally distributed and the sample size is small.
12. Body mass indices of college students are not normally distributed and the sample size is large.
Example

Solutions

6) Case 2
8) Case 3
10) Case 1
In Exercises 13–20, samples are taken. Provide (a) $\mu_j$ and (b) $\sigma_j$, and (c) determine whether the sampling distribution of $\bar{x}$ is normal, approximately normal, or unknown.

16. Accountant incomes are not normally distributed, with $\mu = $60,000 and $\sigma = $10,000. A sample of size 16 is taken.
Example

Solutions

16(a) \( \mu_x = \mu = 60,000 \)

16(b) \[ \sigma_{\bar{x}} = \sigma / \sqrt{n} = 10,000 / \sqrt{16} = 2,500 \]

16(c) unknown
In Exercises 13–20, samples are taken. Provide (a) \( \mu \) and (b) \( \sigma \), and (c) determine whether the sampling distribution of \( \bar{x} \) is normal, approximately normal, or unknown.

The gas mileage for 2007 Toyota Prius hybrid automobiles is not normally distributed, with \( \mu = 50 \) miles per gallon and \( \sigma = 6 \). A sample of size 64 is taken.
Example

Solutions

20(a) \( \mu_x = \mu = 50 \text{ miles per gallon} \)

20(b)

\[
\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{64}} = 0.75 \text{ mpg}
\]

20(c) approximately normal
Example

Page 363

For Exercises 21–28, if possible find the indicated probability. If not possible, explain why not.

26. Systolic blood pressure readings are normally distributed, with $\mu = 80$ and $\sigma = 8$. A sample of size 25 is taken. Find $P(78 < \bar{x} < 82)$. 
Example

Solution:

first notice that it is possible to find the probability since the systolic blood pressure readings are normally distributed so the distribution of the sample mean is also normal (case 1)

\[ \mu_{\bar{x}} = \mu = 80 \]

\[ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{8}{\sqrt{25}} = 1.6 \]
Example

Solution:

\[ P(78 < \bar{x} < 82) = \text{normalcdf}(78,82,80,16) = 0.7887 \]
Example

Page 363

For Exercises 21–28, if possible find the indicated probability. If not possible, explain why not.

27. The pollen count distribution for Los Angeles in September is not normally distributed, with \( \mu = 8.0 \) and \( \sigma = 1.0 \). A sample of size 16 is taken. Find \( P( \bar{x} > 9.0 ) \).
Example

Solution

Not possible: since the pollen count distribution is not normally distributed and the sample size is smaller than 30, the sampling distribution of the mean of $x$ is unknown.
Example

Page 363

For Exercises 29–38, find the indicated value of $\bar{x}$. If it is not possible, explain why not.

38. The pollen count distribution for Los Angeles in September is not normally distributed, with $\mu = 8.0$ and $\sigma = 1.0$. A sample of size 64 is taken. Find the 75th percentile of sample means.
Solution:

first notice that it is possible to find the probability since even though the pollen count distribution is not normal, the sample size is at least 30, so the distribution of the sample mean is also normal (case)

\[ \mu_{\bar{x}} = \mu = 8 \]

\[ \sigma_{\bar{x}} = \sigma / \sqrt{n} = 1 / \sqrt{64} = 0.125 \]
Example

Solution (directly with calculator):

Find $\bar{X}_C$ so that $P(\bar{X} < X_C) = 0.75$

$$\bar{X}_C = \text{invNorm}(0.75, 8, 0.125) \approx 8.1$$
42. **Tennessee Temperatures.** According to the National Oceanic and Atmospheric Administration, the mean temperature for Nashville, Tennessee, in the month of January between 1872 and 2006 was 38.6°F. Assume that the standard deviation is 10 degrees and the distribution is normal. Suppose we take a sample of 25 days.

   a. Does Case 1 apply? If so, find $P(\bar{x} > 40)$. If not, explain why not.
   b. Does Case 2 apply? If so, find $P(\bar{x} > 40)$. If not, explain why not.
Solution:

(a) yes- case 1 applies

\[ \mu_{\bar{x}} = \mu = 38.6^\circ \]

\[ \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{10^\circ}{\sqrt{25}} = 2^\circ \]

\[ P(\bar{x} > 40^\circ) = 1 - P(\bar{x} < 40^\circ) \]

\[ = 1 - 0.7580 \]

\[ = 0.2420 \]
Example

Solution:

(b) case 2 does not apply since the sample size is less than 30.
In this section, we examine the behavior of the sample mean when the population is not normal.

The approximate normality of the sampling distribution of the sample mean kicks in much quicker when the original population is symmetric rather than skewed.

The Central Limit Theorem is one of the most important results in statistics and is stated as follows:
Summary

- Given a population with mean $\mu$ and standard deviation $\sigma$, the sampling distribution of the sample mean $\overline{x}$ becomes approximately normal $(\mu, \sigma / \sqrt{n})$ as the sample size gets larger, regardless of the shape of the population.

- This approximation applies for smaller sample sizes when the original distribution is more symmetric.
7.3 Central Limit Theorem for Proportions

Objectives:
By the end of this section, I will be able to...

1) Explain the sampling distribution of the sample proportion $\hat{p}$.
2) Describe the sampling distribution of the sample proportion $\hat{p}$ for extreme and moderate values of $p$.
3) Apply the Central Limit Theorem for Proportions to solve probability questions about the sample proportion.
Sample Proportion $\hat{p}$

- Suppose each individual in a population either has or does not have a particular characteristic.
- For sample of size $n$ sample proportion $\hat{p}$ (read “$p$-hat”) is

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

- $x$ represents the number of individuals in the sample that have the particular characteristic.
- Use $\hat{p}$ to estimate the unknown value of the population proportion $p$. 
Suppose we are interested in the proportion of female members of the student government. Since 3 of the 5 members of the student government are female, the population proportion of females is \( p = \frac{3}{5} = 0.6 \).

<table>
<thead>
<tr>
<th></th>
<th>Amber</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Dave</th>
<th>Emma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td>Male</td>
<td>Female</td>
<td></td>
</tr>
</tbody>
</table>

Suppose we take all possible samples of size 3. For example, the first sample of Amber, Brandon, and Chantal contains 2 females and 1 male, so the sample proportion of females is \( \hat{p} = \frac{2}{3} \). The second sample of Amber, Brandon, and Dave contains 1 female and 2 males, so \( \hat{p} = \frac{1}{3} \).

a. Make a table of the sample proportion \( \hat{p} \) and the sampling error \( |\hat{p} - p| \) for all possible samples of size \( n = 3 \).

b. Construct a dot plot of the values of \( \hat{p} \) for all possible samples of size \( n = 3 \).
Table 7.6 All possible samples of size 3 from population of student government members

<table>
<thead>
<tr>
<th>Sample</th>
<th>Amber</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Amber</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Dave</th>
<th>Amber</th>
<th>Chantal</th>
<th>Dave</th>
<th>Brandon</th>
<th>Chantal</th>
<th>Dave</th>
<th>Chantal</th>
<th>Dave</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>F</td>
</tr>
<tr>
<td>$\hat{p}$</td>
<td>2/3</td>
<td>1/3</td>
<td>2/3</td>
<td>2/3</td>
<td>3/3</td>
<td>2/3</td>
<td>1/3</td>
<td>2/3</td>
<td>1/3</td>
<td>2/3</td>
<td>1/3</td>
<td>2/3</td>
<td>2/3</td>
<td>2/3</td>
<td>2/3</td>
</tr>
</tbody>
</table>
Example 7.15 - Mean of sample proportions

Calculate the mean of the ten sample proportions \( \hat{p} \) from Table 7.6 page 367.
Example 7.15 continued

Solution

- Mean is

\[
\left( \frac{2}{3} + \frac{1}{3} + \frac{2}{3} + \frac{2}{3} + \frac{3}{3} + \frac{2}{3} + \frac{1}{3} + \frac{2}{3} + \frac{1}{3} + \frac{2}{3} \right) \div 10 = 0.6
\]

- \( \hat{p} \) equals the population proportion of females for the original population, 
  \( p = \frac{3}{5} = 0.6 \).
Fact 6: Mean of the Sampling Distribution of the Sample Proportion \( \hat{p} \)

- The value of the population proportion \( \hat{p} \)
- Denoted as \( \mu_{\hat{p}} \)
- Where \( \mu_{\hat{p}} = p \)
- Read as “the mean of the sampling distribution of \( p \) is \( p \)”
Fact 7 - Standard Deviation of the Sampling Distribution of the Sample Proportion $\hat{p}$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}}$$

where \( q = 1 - p \)

- \( p \) is the population proportion
- \( n \) is the sample size
Example

Page 379

PRACTICING THE TECHNIQUES
In Exercises 7–12, samples are taken. Find (a) \( \mu_{\hat{p}} \) and (b) \( \sigma_{\hat{p}} \), and (c) determine whether the sampling distribution of \( \hat{p} \) is approximately normal or unknown.

7. \( p = 0.5, n = 100 \)
8. \( p = 0.5, n = 5 \)
9. \( p = 0.01, n = 100 \)
10. \( p = 0.01, n = 500 \)
11. \( p = 0.9, n = 40 \)
12. \( p = 0.9, n = 50 \)

Do 8 and 10 parts (a) and (b)
Example

Solutions

8(a) \( \mu_{\hat{p}} = p = 0.5 \)

8(b) \( q = 1 - p = 0.5 \)

\[
\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.5) \cdot (0.5)}{5}} = 0.2236
\]
Example

Solutions

10(a) \( \mu_{\hat{p}} = p = 0.01 \)

10(b) \( q = 1 - p = 0.99 \)

\[
\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.01) \cdot (0.99)}{500}} = 0.0044
\]
Fact 8 - Conditions for Approximate Normality for the Sampling Distribution of the Sample Proportion $\hat{p}$

- May be considered approximately normal only if both the following conditions hold:

$\begin{align*}
(1) \ np & \geq 5 \ \text{and} \ (2) \ n(1 - p) \geq 5
\end{align*}$
Example

Page 379

PRACTICING THE TECHNIQUES
In Exercises 7–12, samples are taken. Find (a) \( \mu_{\hat{p}} \) and (b) \( \sigma_{\hat{p}} \), and (c) determine whether the sampling distribution of \( \hat{p} \) is approximately normal or unknown.
7. \( p = 0.5, n = 100 \)
8. \( p = 0.5, n = 5 \)
9. \( p = 0.01, n = 100 \)
10. \( p = 0.01, n = 500 \)
11. \( p = 0.9, n = 40 \)
12. \( p = 0.9, n = 50 \)

Do part (c)
Example

Solutions

8(c)

\[ np = 5 \cdot (0.5) = 2.5 < 5 \]

\[ n(1 - p) = nq = 5 \cdot (0.5) = 2.5 < 5 \]

Unknown (we cannot conclude that the sampling distribution of the proportion is normal in this case)
Example

Solutions

10(c)

\[ np = 500 \cdot (0.01) = 5 \]

\[ n(1 - p) = nq = 500 \cdot (0.99) = 495 > 5 \]

We conclude that the sampling distribution of the proportion is approximately normal in this case.
Fact 8 continued

- Given a value for $p$, the minimum sample size required to produce approximate normality in the sampling distribution of the proportion can be found by solving each of these for $n$:

$$np = 5 \quad \text{and} \quad nq = 5$$

and choosing the next largest integer value that is at least as large as both of these values for $n$. 
In Exercises 13–18, find the minimum sample size that produces a sampling distribution of \( \hat{p} \) that is approximately normal.

13. \( p = 0.5 \)
14. \( p = 0.25 \)
15. \( p = 0.1 \)
16. \( p = 0.05 \)
17. \( p = 0.01 \)
18. \( p = 0.001 \)

Do problem 16
Example

Solutions

16)

\[ np = n \cdot (0.05) = 5 \quad \Rightarrow \quad n = 100 \]

\[ nq = n \cdot (0.95) = 5 \quad \Rightarrow \quad n = 5.26 \]

The minimum sample size is 100
Fact 9 - Standardizing a Normal Sampling Distribution for Proportions

- When the sampling distribution of $\hat{p}$ is normal (or approximately normal), we can standardize to produce the standard normal $Z$:

$$Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{\hat{p} - p}{\sqrt{\frac{p \cdot (1-p)}{n}}}$$

- where $p$ is the population proportion of successes and $n$ is the sample size.
Example

Page 379

For Exercises 29–34, if possible find the indicated probability. If it is not possible, explain why not.

29. \( p = 0.5, n = 100, P(\hat{\rho} > 0.55) \)
30. \( p = 0.5, n = 5, P(\hat{\rho} > 0.55) \)
31. \( p = 0.01, n = 100, P(\hat{\rho} > 0.011) \)
32. \( p = 0.01, n = 500, P(\hat{\rho} > 0.011) \)
33. \( p = 0.9, n = 40, P(0.88 < \hat{\rho} < 0.91) \)
34. \( p = 0.9, n = 50, P(0.88 < \hat{\rho} < 0.91) \)

Do problem 30
Solutions

30) First check that we can assume a normal distribution:

\[ np = 5 \cdot (0.5) = 2.5 < 5 \]

\[ n(1 - p) = nq = 5 \cdot (0.5) = 2.5 < 5 \]

We cannot conclude that the sampling distribution of the proportion is normal in this case and we cannot find the probability.
Example

Do problem 32
Example

Solutions

First check that we can assume a normal distribution for $\hat{p}$:

\[
np = 500 \cdot (0.01) = 5
\]

\[
n(1 - p) = nq = 500 \cdot (0.99) = 495 > 5
\]
Example

Solutions

For normal distribution use

mean: \( \mu = p = 0.01 \)

Standard deviation:

\[
\sigma = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.01) \cdot (0.99)}{500}} = 0.0044
\]
Example

Solutions

z-score method (Table)

\[ Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{0.011 - 0.01}{0.0044} \approx 0.23 \]
Example

Solutions

using standard normal distribution and table lookup:

\[ P(Z > 0.23) = 1 - P(Z < 0.23) \]
\[ = 1 - 0.5910 \]
\[ = 0.4090 \]

From Table T-10
Example

Solutions

It is more accurate to compute $Z$ as

\[
Z = \frac{\hat{p} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \frac{0.011 - 0.01}{\sqrt{(0.01) \cdot (0.99)/500}} \approx 0.22
\]

so that:

\[
P(Z > 0.22) = 1 - P(Z < 0.22) = 1 - 0.5871 = 0.4129
\]
Example

For Exercises 35–40, find the indicated value of $\hat{p}$. If it is not possible, explain why not.

35. $p = 0.5$, $n = 100$, value of $\hat{p}$ larger than 90% of all values of $\hat{p}$
36. $p = 0.5$, $n = 400$, value of $\hat{p}$ larger than 90% of all values of $\hat{p}$
37. $p = 0.9$, $n = 64$, 95th percentile of values of $\hat{p}$
38. $p = 0.9$, $n = 144$, 95th percentile of values of $\hat{p}$
39. $p = 0.1$, $n = 64$, 10th percentile of values of $\hat{p}$
40. $p = 0.1$, $n = 144$, 10th percentile of values of $\hat{p}$

Do problem 36
Example

Solutions

First check that we can assume a normal distribution for $\hat{p}$:

\[
np = 400 \cdot (0.5) = 200
\]

\[
n(1 - p) = nq = 400 \cdot (0.50) = 200
\]

Yes- both are greater than 5
Example

Solutions

Find the value: \( \hat{p}_c \)

So that

\[
P(\hat{p} < \hat{p}_c) = 0.90
\]

\( \hat{p}_c \) is the 90\textsuperscript{th} percentile of values of \( \hat{p} \)
Example

Solutions

For normal distribution use

mean: \( \mu_\hat{p} = p = 0.5 \)

standard deviation:

\[
\sigma_\hat{p} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.5) \cdot (0.5)}{400}} = 0.025
\]
Example

Solutions

Using calculator gives:

\[ \hat{p}_c = \text{invNorm}(0.90, 0.5, 0.025) \approx 0.532 \]
Central Limit Theorem for Proportions

- the sampling distribution of the sample proportion $\hat{p}$ follows an approximately normal distribution with
- mean $\mu_{\hat{p}} = p$
- standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{pq}{n}}$

when the following conditions are satisfied: $np \geq 5$ and $n(1 - p) \geq 5$
42. Women’s Radio Preferences. In their 2001 study *What Women Want: Five Secrets to Better Ratings*, the Arbitron company reported that “Music I Like” is the biggest factor in deciding which radio station to tune to, chosen by 87% of women.

a. Find the minimum sample size $n^*$ that produces a sampling distribution of $\hat{p}$ that is approximately normal.

b. Confirm that this sample size satisfies the conditions in Fact 8.

c. Describe the sampling distribution of $\hat{p}$ if we use this minimum sample size. Which fact allows us to say this?

d. Find $\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$ for $n = 50$.

e. Find the probability that, in a sample of 50 women, more than 45 chose “Music I Like” as the reason they decide which radio station to tune to.
Example

Solutions

(a) Take \( p = 0.87 \) and choose smallest integer value of \( n \) so that both of the following are true:

\[
np = n(0.87) \geq 5
\]
\[
n(1 - p) = nq = n \cdot (0.13) \geq 5
\]

Answer: \( n = n^* = 39 \)
Example

Solutions

(b) Check that we can assume a normal distribution for $\hat{p}$ with a sample size of $n=39$

$$np = 39 \cdot (0.87) = 33.93$$

$$n(1 - p) = nq = 39 \cdot (0.13) = 5.07$$

Yes- both larger than 5
Example

Solutions

(c) The Central Limit Theorem tells us that the sampling distribution of $\hat{p}$ is approximately normal when $n=39$
(d) The Central Limit Theorem gives that the sampling distribution of $\hat{p}$ is approximately normal with

$$\mu_{\hat{p}} = p = 0.87$$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.87 \cdot 0.13}{50}} \approx 0.0476$$
Example

(d) For \( n=50 \), find

\[
P\left( \hat{p} > \frac{45}{50} \right) = P \left( \Phi > 0.90 \right)
\]

\[
= \text{normalcdf}(0.90,10^{99},0.87,0.0476) \approx 0.2643
\]
Summary

- The sampling distribution of the sample proportion \( \hat{p} \) for a given sample size \( n \) consists of the collection of the sample proportions of all possible samples of size \( n \) from the population.

- The approximate normality of the sampling distribution of the sample proportion kicks in much quicker when the population proportion is moderate rather than extreme.
Summary

- According to the Central Limit Theorem for Proportions, the sampling distribution of the sample proportion \( \hat{p} \) follows an approximately normal distribution with mean \( \mu_{\hat{p}} = p \) and standard deviation \( \sigma_{\hat{p}} = \sqrt{p \cdot (1-p) / n} \) if the following conditions are satisfied: (1) \( np \geq 5 \) and (2) \( n(1 - p) \geq 5 \).

- We can use Fact 9 to find probabilities and percentiles for sample proportions.