Section 6.3: The Central Limit Theorem

A sampling distribution of sample means is a distribution using the means computed from all possible random samples of a specific size taken from a population.

Sampling error is the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population.

Properties of the Distribution of Sample Means

1. The mean of the sample means will be the same as the population mean.

2. The standard deviation of the sample means will be smaller than the standard deviation of the population, and it will be equal to the population standard deviation divided by the square root of the sample size.

Exercise 14. Suppose a professor gave an 8-point quiz to a small class of four students. The results of the quiz were 2, 6, 4, and 8. Assume that the four students constitute the population.

\[
\mu = \frac{2 + 6 + 4 + 8}{4} = 5
\]

\[
\sigma^2 = \frac{(2-5)^2 + (6-5)^2 + (4-5)^2 + (8-5)^2}{4} = \frac{9 + 1 + 1 + 9}{4} = \frac{20}{4} = 5
\]

\[
\sigma = \sqrt{5} \approx 2.236
\]

Take samples of size 2 with replacement and find the mean of each sample.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>2, 2</td>
<td>2</td>
</tr>
<tr>
<td>2, 4</td>
<td>3</td>
</tr>
<tr>
<td>2, 6</td>
<td>5</td>
</tr>
<tr>
<td>2, 8</td>
<td>5</td>
</tr>
<tr>
<td>4, 2</td>
<td>3</td>
</tr>
<tr>
<td>4, 4</td>
<td>4</td>
</tr>
<tr>
<td>4, 6</td>
<td>5</td>
</tr>
<tr>
<td>4, 8</td>
<td>6</td>
</tr>
<tr>
<td>6, 2</td>
<td>4</td>
</tr>
<tr>
<td>6, 4</td>
<td>5</td>
</tr>
<tr>
<td>6, 6</td>
<td>6</td>
</tr>
<tr>
<td>6, 8</td>
<td>7</td>
</tr>
<tr>
<td>8, 2</td>
<td>5</td>
</tr>
<tr>
<td>8, 4</td>
<td>6</td>
</tr>
<tr>
<td>8, 6</td>
<td>7</td>
</tr>
<tr>
<td>8, 8</td>
<td>8</td>
</tr>
</tbody>
</table>

Frequency Distribution of Sample Means

<table>
<thead>
<tr>
<th>Sample Mean</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
</tbody>
</table>

Histogram

Histogram of Sample Means

\[
M_X = \frac{2 + 6 + 12 + 20 + 18 + 14 + 8}{16} = 5
\]

\[
\sigma_{X} = \sqrt{\frac{(2-5)^2 + (3-5)^2 + (4-5)^2 + (5-5)^2 + (6-5)^2 + (7-5)^2 + (8-5)^2}{16}}
\]

\[
\approx [1.58] \triangleleft \frac{2.236}{\sqrt{2}} \approx 1.581
\]
The Central Limit Theorem

As the sample size \( n \) increases without limit, the shape of the distribution of the sample means taken with replacement from a population with mean \( \mu \) and standard deviation \( \sigma \) will approach a normal distribution. As previously shown, this distribution will have a mean \( \mu \) and a standard deviation \( \frac{\sigma}{\sqrt{n}} \).

Note: If the sample size is sufficiently large, the central limit theorem can be used to answer questions about sample means in the same manner that a normal distribution can be used to answer questions about individual values. The only difference is that a new formula must be used for the \( z \) values. It is

\[
z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}
\]

It's important to remember two things when you use the central limit theorem:

a) When the original variable is normally distributed, the distribution of the sample means will be normally distributed, for any sample size \( n \).

b) When the distribution of the original variable is not normal, a sample size of 30 or more is needed to use a normal distribution to approximate the distribution of the sample means. The larger the sample, the better the approximation will be.

Exercise 15. A. C. Neilsen reported that children between the ages of 2 and 5 watch an average of 25 hours of television per week. Assume the variable is normally distributed and the standard deviation is 3 hours. If 20 children between the ages of 2 and 5 are randomly selected, find the probability that the mean of the number of hours they watch television will be greater than 26.3 hours.

\[
\mu = 25, \quad \sigma = 3, \quad n = 20
\]

1. \( \bar{X} = 25 \)
2. Convert \( \bar{X} \) to \( z \)
\[
z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{26.3 - 25}{\frac{3}{\sqrt{20}}} = 1.94
\]
3. \( A(1.94) = .9738 \)
   \( 1 - .9738 = .0262 \)
   Prob. that sample mean is greater than 26.3 hrs is .0262.
Exercise 16. The average age of a vehicle registered in the United States is 8 years, or 96 months. Assume the standard deviation is 16 months. If a random sample of 36 vehicles is selected, find the probability that the mean of their age is between 90 and 100 months.

\[ \mu = 96 \text{ months} \]
\[ \sigma = 16 \text{ months} \]
\[ n = 36 \]

\[ z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} \]

(1) \[ \bar{x} = 90 \]
\[ z = \frac{90 - 96}{\frac{16}{\sqrt{36}}} = -2.25 \]

(2) \[ \bar{x} = 100 \]
\[ z = \frac{100 - 96}{\frac{16}{\sqrt{36}}} = 1.50 \]

(3) \[ A(-2.25) = .0122 \]
\[ A(1.50) = .9332 \]
\[ .9332 - .0122 = .9210 \]

Note: To gain information about an individual data value obtained from the population, use formula \[ z = \frac{X - \mu}{\sigma} \]. To gain information about a sample mean, use formula \[ z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \].

Exercise 17. The average time spent by construction workers who work on weekends is 7.93 hours (over 2 days). Assume the distribution is approximately normal and has a standard deviation of 0.8 hour.

a) Find the probability that an individual who works at that trade works fewer than 8 hours on the weekend.

\[ \mu = 7.93 \]
\[ \sigma = 0.8 \]

\[ z = \frac{X - \mu}{\sigma} \]
\[ = \frac{8 - 7.93}{0.8} \approx 0.09 \]

\[ A(0.09) = .5359 \]

Area between \( z = 0.09 \) and \( 8 \) hrs on a weekend is .5359
b) If a sample of 40 construction workers is randomly selected, find the probability that the mean of the sample will be less than 8 hours.

\[ \overline{X} = 8, \quad n = 40 \]

\[ Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{8 - 7.93}{\frac{0.8}{\sqrt{40}}} \approx 0.55 \]

\[ A(0.55) = 0.7088 \]

Prob. of getting a sample mean of less than 8 hrs when sample size is 40 is 0.7088

The probability of selecting an individual construction worker who works less than 8 hours on a weekend is __.5359__. The probability of selecting a random sample of 40 construction workers with a mean of less than 8 hours per week is __.7088___. This difference of __.1729___ is due to the fact that the distribution of sample means is much less variable than the distribution of individual data values. The reason is that as the sample size increases, the standard deviation of the means decreases.