Section 4.4

Sampling Distribution of $\bar{x}$ and $\hat{p}$

A **parameter** is a number that describes a population.

A **statistic** is a number that describes a sample. A statistic can be a mean, proportion, standard deviation, etc. Often, a statistic is used to estimate an unknown parameter.

A **sampling distribution** of a statistic is the distribution of values taken by the statistic in all possible samples of the same sizes from the same population. The **sample distribution of** $\bar{x}$ is the distribution of all sample means in all possible samples of the population.

Example 1: A local university has a large number of commuter students. The mean time it takes for the commuters to get to school is 16 minutes. Suppose we take 5 commuter students from this population and are interested in how long it takes these five students to commute to school. The information is listed below.

<table>
<thead>
<tr>
<th>Xerxes</th>
<th>Yaura</th>
<th>Zoe</th>
<th>Annie</th>
<th>Becca</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>20</td>
<td>5</td>
<td>30</td>
<td>15</td>
</tr>
</tbody>
</table>

Take samples of size 3 and calculate the mean, $\bar{x}$, of each sample.

a. How many possible samples of size 3 exist?

b. Calculate the sample mean, $\bar{x}$, for each sample.
If our original population has a normal distribution, the sample mean’s distribution is also normally distributed.

Suppose that \( \bar{x} \) is the mean of a simple random sample of size \( n \) drawn from a large population. If the population mean is \( \mu \) and the population standard deviation is \( \sigma \), then

- the mean of the sampling distribution of \( \bar{x} \) is \( \mu_{\bar{x}} = \mu \)
- and
- the standard deviation of the sampling distribution is \( \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \), which is called the standard error.

As \( n \) gets larger the standard deviation gets smaller, which results in more precise estimation.

What if our original population is not normal? For example, skewed left or skewed right.

- As the sample size increases, the distribution becomes approximately normal.

How large does it have to be? In general it depends on the degree of symmetry, or skewness, of the population. Actually as the sample size \( n \) increases, the sample mean becomes approximately normal.

Rule of thumb would be if \( n \geq 30 \), then we can assume normality as stated by the following theorem.

The Central Limit Theorem states that if we draw a simple random sample of size \( n \) from any population with mean \( \mu \) and standard deviation \( \sigma \), when \( n \) is large the sampling distribution of the sample mean \( \bar{x} \) is close to the normal distribution \( N(\mu, \sigma / \sqrt{n}) \).

If the population is either not normal or unknown and \( n < 30 \), then we have insufficient information to conclude normality or approximately normal.
Example 2: State tax officials claim that the amount of money claimed by all the taxpayers within the state for charitable deductions during 2010 is normally distributed with an average amount of $968 with a standard deviation of $102. Many samples of size 64 are taken. Find the mean of these samples and the standard error.

Command: Answer:

Example 3: The distribution of weight of a certain brand of crackers is normal with a mean of 12.2 ounces and a standard deviation of 1.5 ounces. A sample of size 16 is taken.

a. What is the sample mean?

b. What is the standard deviation of the sample mean?

Command: Answer:

c. What is the probability that the mean of a sample of 16 such crackers is less than 11.5 ounces?

Command: Answer:
d. What is the probability that the mean of a sample of 16 such crackers is more than 13 ounces?

Command: ____________________________ Answer: ____________________________

e. What is the probability that the mean of a sample of 16 such crackers is between 11.7 and 13.5?

Command: ____________________________ Answer: ____________________________

Example 4: Suppose that a random sample of size 64 is to be selected from a population with mean 42 and standard deviation 9. What is the approximate probability that the sample mean will be within 0.5 of the population mean?

Command: ____________________________ Answer: ____________________________

Section 4.4 – **Sampling Distribution of** $\bar{x}$ **and** $\hat{p}$
Sampling Proportions

Next we’ll study the sampling distribution of the sample proportion called \( \hat{p} \) (p-hat).

Let’s say that from a large population we take a sample of size \( n \), and each element either possesses or does not possess a certain characteristic. The proportion that represents the number of elements in the sample that possess the particular characteristic is given by \( \frac{X}{n} \).

Like the sampling distribution of the sample mean, the sampling distribution of the sample proportion is not always normal. To achieve normality the following conditions must be satisfied.

- \( np \geq 10 \)
- \( n(1 - p) \geq 10 \)

The sampling distribution of \( \hat{p} \) has:

- a mean of \( \mu_{\hat{p}} = p \) (the proportion itself)
- a standard deviation of \( \sigma_{\hat{p}} = \sqrt{\frac{p(1 - p)}{n}} \)

Example 5: A certain surveying company reported that 25% of telephone users no longer use landlines and only use their cell phones. Suppose we take samples of size 40. Can we achieve normality? If so, proceed to answering the next questions.

a. Find the mean and standard error of the proportion of telephone users who no longer use landlines.
b. Find the probability that the sampling proportion is less than 0.26

Example 6: 43% of the voters in the 1992 Presidential election voted for Bill Clinton. Suppose you take a simple random sample of 500 voters from this population.

a. Is 43% a parameter or a statistic?

b. If we can achieve normality then determine the probability that the sample proportion of Clinton voters falls between 40% and 46%.

Command: Answer:
Example 7: Power companies kill trees growing near their lines to avoid power failures due to falling limbs in storms. Applying a chemical to slow the growth of the trees is cheaper than trimming, but the chemical kills some of the trees. Suppose that one such chemical would kill 20% of sycamore trees. The power company tests the chemical on 250 sycamores. Consider these a SRS from the population of all sycamore trees.

a. What are the mean and standard deviation of the proportion of trees that are killed?

b. What is the probability that at least 60 trees are killed?

Command: Answer: