Chapter Summary: Sampling Distributions

This chapter introduced you to a key concept for inferential thinking – sampling distributions. Since we are interested in drawing conclusions about population proportions and means, it is important to know how statistics will behave in repeated random sampling. Being able to describe the sampling variability for sample statistics allows us to estimate and test claims about population parameters. This chapter provided us with some key facts about sample statistics that will help us as we begin our formal study of inference. First, statistics will vary from sample to sample. Second, if the sample size is large enough, we know that the distribution of sample statistic values will be approximately Normal. Third, the sampling distributions of \( \hat{p} \) and \( \bar{x} \) will be centered at \( p \) and \( \mu \), respectively. Finally, the variability of these sampling distributions can be computed (as long as the 10% condition is met). This variability will decrease as the sample size increases, so bigger random samples are more desirable. One important fact about sample means was revealed in this chapter. When sampling from a Normal population, the sampling distribution of \( \bar{x} \) will be Normal. However, as long as our sample size is at least 30, the shape of the sampling distribution of \( \bar{x} \) will be approximately Normal—no matter what the population distribution looks like! The central limit theorem is a powerful fact that will be revisited in the coming chapters. Now that we have a grasp of the basic concept of sampling distributions, we are ready to begin the formal study of statistical inference. In the next chapter, we will use what we have learned to estimate population proportions and means with confidence.

After You Read: What Have I Learned?
Complete the vocabulary puzzle, multiple-choice questions, and FRAPPY. Check your answers and performance on each of the learning targets. Be sure to get extra help on any targets that you identify as needing more work!

<table>
<thead>
<tr>
<th>Target</th>
<th>Got It!</th>
<th>Almost There</th>
<th>Needs Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>I can distinguish between a parameter and a statistic.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can define a sampling distribution.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can distinguish between a population distribution, sampling distribution, and distribution of sample data.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can determine whether a statistic is an unbiased estimator of a population parameter.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can describe the relationship between sample size and the variability of a statistic.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can calculate and interpret the mean and standard deviation of the sampling distribution of a sample proportion.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can check whether the 10% and Normal conditions are met in a given setting.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can use the Normal approximation to calculate probabilities involving sample proportions.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can use the sampling distribution of ( \hat{p} ) to evaluate a claim about a population proportion.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can calculate and interpret the mean and standard deviation of the sampling distribution of a sample mean.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target</td>
<td>Got It!</td>
<td>Almost There</td>
<td>Needs Work</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>--------</td>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>I can calculate probabilities involving a sample mean when the population distribution is Normal.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can explain how the shape of the sampling distribution of $\bar{x}$ is related to the shape of the population distribution.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I can use the central limit theorem to help find probabilities involving a sample mean.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Across

2. _____ distribution: the distribution of values taken by the statistic in all possible samples of the same size from the population
4. _____ distribution: the distribution of all values of a variable in the population
8. _____ of a statistic is described by the spread of the sampling distribution
12. Greek letter used for the population standard deviation
14. the Normal approximation for the sampling distribution of a sample proportion can be used when both the number of successes and failures are greater than _____
15. sampling distributions and sampling variability provide the foundation for performing _____
16. central _____ theorem tells us if the sample size is large, the sampling distribution of the sample mean is approximately Normal, regardless of the shape of the population

Down

1. a statistic is an _____ estimator if the mean of the sampling distribution is equal to the true value of the parameter being estimated.
2. a number, computed from sample data, that estimates a parameter
3. Greek letter used for the population mean
5. standard _____: measure of spread of a sampling distribution
6. sampling _____: notes the value of a statistic may be different from sample to sample
7. a number that describes a population
9. the rule of thumb for using the central limit theorem - the sample size should be greater than _____
10. when the sample size is large, the sampling distribution of a sample proportion is approximately _____
11. to draw a conclusion about a population parameter, we can look at information from a _____ sample
13. center of a sampling distribution
Section 7.1: What is a Sampling Distribution?

Before You Read: Section Summary
This section will introduce you to the big ideas behind sampling distributions. First, you will learn how to distinguish between population parameters and statistics derived from samples. Next, you will explore the fact that statistics vary from sample to sample. This simple fact is the reason we study sampling distributions. By describing the shape, center, and spread of the sampling distribution of a statistic, we can determine the critical information necessary to perform statistical inference.

Learning Targets:

- I can distinguish between a parameter and a statistic.
- I can define a sampling distribution.
- I can distinguish between a population distribution, sampling distribution, and distribution of sample data.
- I can determine whether a statistic is an unbiased estimator of a population parameter.
- I can describe the relationship between sample size and the variability of a statistic.

While You Read: Key Vocabulary and Concepts

<table>
<thead>
<tr>
<th>parameter:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>statistic:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>sampling variability:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>sampling distribution:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>population distribution:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>unbiased estimator:</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>variability of a statistic:</th>
</tr>
</thead>
</table>
Concept 1: Parameters and Statistics
One of the most powerful skills we learn from statistics is the ability to answer a question about a population characteristic based on information gathered from a random sample. That is, we can use a statistic calculated from a sample to make a conclusion about a corresponding parameter in the population. However, we must note that the statistics we calculate from a sample may differ somewhat from the population characteristic we are trying to measure. Further, the statistic would likely differ from sample to sample. This sample-to-sample variability poses a problem when we try to generalize our findings to the population. However, based on what we learned in the last chapter, we can view a sample statistic as a random variable. That is, while we have no way of predicting exactly what statistic value we will get from a sample, we know how those values will behave in repeated random sampling. With the probability distribution of this random variable in mind, we can use a sample statistic to estimate the population parameter.

Check for Understanding: ___ I can distinguish between a parameter and a statistic.

For each of the following situations, identify the population of interest, the parameter, and the statistic.

a) A medical researcher is interested in exploring the effects of a new medicine on blood pressure. 500 males with high blood pressure are randomly selected and given the new drug. After two weeks, their blood pressure is measured and the average arterial pressure is calculated.

b) A study is conducted to determine whether or not the dangerous activity of texting while driving is a common practice. 1500 16- to 24-year-olds are randomly selected and asked whether or not they text while driving. Of the 1500 drivers, 12% indicate they text while driving.
Chapter 7: Sampling Distributions

Concept 2: Describing Sampling Distributions
To draw a conclusion about a population proportion \( p \), we take a random sample and calculate the sample proportion \( \hat{p} \). Likewise, to reach a conclusion about a population mean \( \mu \), we take a random sample and calculate the sample mean \( \bar{x} \). Because of chance variation in random sampling, the values of our sample statistic will vary from sample to sample. The distribution of statistic values in all possible samples of the same size from a population is called the sampling distribution of the statistic. The sampling distribution describes the sampling variability and provides a foundation for performing inference. The spread of a sampling distribution is an important attribute as all inference calculations depend upon it! When trying to estimate a parameter, we want minimum sampling variability and no bias. Random sampling helps us avoid bias while larger samples help us minimize sampling variability.

Check for Understanding: _____ I can distinguish between a population distribution, sampling distribution, and distribution of sample data.

A breakfast cereal includes marshmallow shapes in the following distribution: 10% stars, 10% crescent moons, 20% rockets, 40% astronauts, 20% planets. We are interested in examining the proportion of rockets in a random sample of 2000 marshmallows from the cereal.

(a) Sketch the population distribution of marshmallow shapes.

(b) Suppose you were to collect a random sample of 2000 marshmallow shapes. Sketch the distribution of sample data you would expect to see. How many rockets would you expect to see in your sample?

(c) Now, suppose you collected many samples of the same size. Sketch the sampling distribution of the proportion of rockets you think you would see in the samples.
Section 7.2: Sample Proportions

Before You Read: Section Summary
The objective of some statistical applications is to reach a conclusion about a population proportion \( p \). For example, we may try to estimate an approval rating through a survey, or test a claim about the proportion of defective light bulbs in a shipment based on a random sample. Since \( p \) is unknown to us, we must base our conclusion on a sample proportion, \( \hat{p} \). However, as we have noted, we know that the value of \( \hat{p} \) will vary from sample to sample. The amount of variability will depend on the size of our sample. In this section, you will learn how to describe the shape, center, and spread of the sampling distribution of \( \hat{p} \) in detail.

Learning Targets:

- I can calculate and interpret the mean and standard deviation of the sampling distribution of a sample proportion.
- I can check whether the 10% and Normal conditions are met in a given setting.
- I can use the Normal approximation to calculate probabilities involving sample proportions.
- I can use the sampling distribution of \( \hat{p} \) to evaluate a claim about a population proportion.

While You Read: Key Vocabulary and Concepts

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

mean of the sampling distribution of \( \hat{p} \):

standard deviation of the sampling distribution of \( \hat{p} \):

Normal approximation for \( \hat{p} \):

After You Read: Check for Understanding

Concept 1: The Sampling Distribution of \( \hat{p} \)
If we take repeated samples of the same size \( n \) from a population with a proportion of interest \( p \), the sampling distribution of \( \hat{p} \) will have the following characteristics:

1) The shape of the sampling distribution will become approximately Normal as the sample size \( n \) increases.
   - We can use Normal calculations if \( np \geq 10 \) and \( n(1-p) \geq 10 \).
2) The mean of the sampling distribution is \( \mu_{\hat{p}} = p \).
3) The standard deviation of the sampling distribution is 
\[ \sigma_p = \sqrt{\frac{p(1-p)}{n}} \]

Note: The formula for the standard deviation is exactly correct only if we are sampling from an infinite population or with replacement from a finite population. When we are sampling without replacement from a finite population, the formula is approximately correct as long as the 10% condition is satisfied. That is, the sample size must be less than or equal to 10% of the population size.

Check for Understanding: ____ I can calculate and interpret the mean and standard deviation of the sampling distribution of a sample proportion and ____ I can check whether the 10% and Normal conditions are met in a given setting.

Suppose your job at a potato chip factory is to check each shipment of potatoes for quality assurance. Further, suppose that a truckload of potatoes contains 95% that are acceptable for processing. If more than 10% are found to be unacceptable in a random sample, you must reject the shipment. To check, you randomly select and test 250 potatoes. Let \( \hat{p} \) be the sample proportion of unacceptable potatoes.

a) What is the mean of the sampling distribution of \( \hat{p} \)?

b) Check the 10% condition and calculate the standard deviation of the sampling distribution of \( \hat{p} \).

c) Check the Normal condition and sketch the sampling distribution of \( \hat{p} \). Based on this sketch, do you think it would be likely to reject the truckload based on a random sample of 250 potatoes? Why or why not?
Concept 2: Using the Normal Approximation for $\hat{p}$
When the sample size $n$ is large enough for $np$ and $n(1-p)$ to both be at least 10, the sampling distribution of $\hat{p}$ will be approximately Normal. In that case, we can use Normal calculations to determine the probability that an SRS will generate a value of $\hat{p}$ in a particular interval. This calculation is an important component of inference.

Check for Understanding: ___ I can use the Normal approximation to calculate probabilities involving sample proportions and ___ I can use the sampling distribution of $\hat{p}$ to evaluate a claim about a population proportion.

A phone company is interested in exploring marketing possibilities for a new smartphone for teenagers. They ask an SRS of 1000 high school students whether they own a smartphone. Suppose 65% of all high school students own a smartphone. What is the probability that the random sample selected by the company will result in a $\hat{p}$-value within 3 percentage points of the true population proportion? Show all your work!
Section 7.3: Sample Means

Before You Read: Section Summary

Then the goal of a statistical application is to reach a conclusion about a population mean \( \mu \) we must rely on a sample mean \( \bar{x} \). However, as we have noted, the value of \( \bar{x} \) will vary from sample to sample. As we observed with sample proportions, the amount of variability will depend on the size \( n \) of our sample. In this section, you will learn how to describe the shape, center, and spread of the sampling distribution of \( \bar{x} \) in detail.

Learning Targets:

- I can calculate and interpret the mean and standard deviation of the sampling distribution of a sample mean.
- I can calculate probabilities involving a sample mean when the population distribution is Normal.
- I can explain how the shape of the sampling distribution of \( \bar{x} \) is related to the shape of the population distribution.
- I can use the central limit theorem to help find probabilities involving a sample mean.

While You Read: Key Vocabulary and Concepts

- Sampling distribution of \( \bar{x} \):
- Mean of the sampling distribution of \( \bar{x} \):
- Standard deviation of the sampling distribution of \( \bar{x} \):
- Central Limit Theorem:
- Normal condition for sample means:

After You Read: Check for Understanding

Concept 1: Sampling Distribution of \( \bar{x} \)

We take repeated random samples of the same size \( n \) from a population with mean \( \mu \), the sampling distribution of \( \bar{x} \) will have the following characteristics:

1) The shape of the sampling distribution depends upon the shape of the population distribution. If the population is Normally distributed, the sampling distribution of \( \bar{x} \) will be Normally distributed. If the population distribution is non-Normal, the sampling distribution of \( \bar{x} \) will become more and more Normal as \( n \) increases.
2) The mean of the sampling distribution is \( \mu_x = \mu \).

3) The standard deviation of the sampling distribution is \( \sigma_x = \frac{\sigma}{\sqrt{n}} \).

Note: The formula for the standard deviation is exactly correct only if we are sampling from an infinite population or with replacement from a finite population. When we are sampling without replacement from a finite population, the formula is approximately correct as long as the 10% condition is satisfied. That is, the sample size must be less than or equal to 10% of the population size.

**Check for Understanding:** ___ I can calculate and interpret the mean and standard deviation of the sampling distribution of a sample mean and ___ I can calculate probabilities involving a sample mean when the population distribution is Normal.

The times it takes 5th graders to complete a particular mathematics problem are Normally distributed with mean 2 minutes and standard deviation 0.8 minutes.

Find the probability that a randomly chosen 5th grader will take more than 2.5 minutes to complete the problem. Show your work.

Suppose you give the problem to an SRS of 20 students. Sketch the sampling distribution of \( \bar{x} \). Then use this distribution to determine the probability that the mean time to complete the problem for the SRS of students is greater than 2.5 minutes. Show your work.
Concept 2: The Central Limit Theorem

When the population is Normally distributed, we know that the sampling distribution of \( \bar{x} \) will be Normally distributed, so we can use Normal calculations. However, most population distributions are not Normally distributed. If our sampling distribution is skewed or non-Normal in some other way we cannot use Normal calculations to answer questions. Thankfully, a pretty remarkable fact about sample means helps us out: when the sample size \( n \) is large, the shape of the sampling distribution of \( \bar{x} \) will be approximately Normal no matter what the shape of the population distribution may be! For our purposes, we’ll define “large” to be any sample that is at least 30. So, if \( n \geq 30 \), we can be safe assuming that the sampling distribution of \( \bar{x} \) will be approximately Normal and we can proceed to perform Normal calculations. If \( n < 30 \), we must consider the shape of the population distribution.

Check for Understanding: ____ I can use the central limit theorem to help find probabilities involving a sample mean.

The blood cholesterol level of adult men has mean 188 mg/dl and standard deviation 41 mg/dl. A SRS of 250 men is selected and the mean blood cholesterol level in the sample is calculated.

Sketch the sampling distribution of \( \bar{x} \) and calculate the probability that the sample mean will be greater than 193.
Chapter 7 Multiple Choice Practice

Directions. Identify the choice that best completes the statement or answers the question. Check your answers and note your performance when you are finished.

1. The variability of a statistic is described by
   (A) the spread of its sampling distribution.
   (B) the amount of bias present.
   (C) the vagueness in the wording of the question used to collect the sample data.
   (D) probability calculations.
   (E) the stability of the population it describes.

2. Below are dot plots of the values taken by three different statistics in 30 samples from the same population. The true value of the population parameter is marked with an arrow.

   ![Dot Plots]

   The statistic that has the largest bias among these three is
   (A) statistic A.
   (B) statistic B.
   (C) statistic C.
   (D) A and B have similar bias, and it is larger than the bias of C.
   (E) B and C have similar bias, and it is larger than the bias of A.

3. According to a recent poll, 27% of Americans prefer to read their news in a physical newspaper instead of online. Let's assume this is the parameter value for the population. If you take a simple random sample of 25 Americans and let \( \hat{p} \) = the proportion in the sample who prefer a newspaper, is the shape of the sampling distribution of \( \hat{p} \) approximately Normal?
   (A) No, because \( p < 0.50 \)
   (B) No, because \( np = 6.75 \)
   (C) Yes, because we can reasonably assume there are more than 250 individuals in the population.
   (D) Yes, because we took a simple random sample.
   (E) Yes, because \( n(1-p) = 18.25 \)

4. The time it takes students to complete a statistics quiz has a mean of 20.5 minutes and a standard deviation of 15.4 minutes. What is the probability that a random sample of 40 students will have a mean completion time greater than 25 minutes?
   (A) 0.9678
   (B) 0.0322
   (C) 0.0344
   (D) 0.3851
   (E) 0.6149

5. A fair coin (one for which both the probability of heads and the probability of tails are 0.5) is tossed 60 times. The probability that more than 1/3 of the tosses are heads is closest to
   (A) 0.9951.
   (B) 0.33.
   (C) 0.109.
   (D) 0.09.
   (E) 0.0049.
3. The histogram below was obtained from data on 750 high school basketball games in a regional athletic conference. It represents the number of three-point baskets made in each game.

What is the range of sample sizes a researcher could take from this population without violating conditions required for performing Normal calculations with the sampling distribution of \( \bar{x} \)?

(A) \( 0 \leq n \leq 30 \)
(B) \( 30 \leq n \leq 50 \)
(C) \( 30 \leq n \leq 75 \)
(D) \( 30 \leq n \leq 750 \)
(E) \( 75 \leq n \leq 750 \)

4. The incomes in a certain large population of college teachers have a normal distribution with mean $60,000 and standard deviation $5000. Four teachers are selected at random from this population to serve on a salary review committee. What is the probability that their average salary exceeds $65,000?

(A) 0.0228
(B) 0.1587
(C) 0.8413
(D) 0.9772
(E) essentially 0

5. A random sample of size 25 is to be taken from a population that is Normally distributed with mean 60 and standard deviation 10. The mean \( \bar{x} \) of the observations in our sample is to be computed. The sampling distribution of \( \bar{x} \)

(A) is Normal with mean 60 and standard deviation 10.
(B) is Normal with mean 60 and standard deviation 2.
(C) is approximately Normal with mean 60 and standard deviation 2.
(D) has an unknown shape with mean 60 and standard deviation 10.
(E) has an unknown shape with mean 60 and standard deviation 2.

6. The scores of individual students on a college entrance examination have a left-skewed distribution with mean 18.6 and standard deviation 6.0. At Millard North High School, 36 seniors take the test. The sampling distribution of mean scores for random samples of 36 students is

(A) approximately Normal.
(B) symmetric and mound-shaped, but non-Normal.
(C) skewed right.
(D) neither Normal nor non-normal. It depends on the particular 36 students selected.
(E) exactly Normal.

7. The distribution of prices for home sales in Minnesota is skewed to the right with a mean of $290,000 and a standard deviation of $145,000. Suppose you take a simple random sample of 100 home sales from this (very large) population. What is the probability that the mean of the sample is above $325,000?

(A) 0.0015
(B) 0.0027
(C) 0.0079
(D) 0.4046
(E) 0.4921
FRAPPY! Free Response AP® Problem, Yay!

The following problem is modeled after actual Advanced Placement Statistics free response questions. Your task is to generate a complete, concise response in 15 minutes. After you generate your response, view two example solutions and determine whether you feel they are "complete", "substantial", "developing" or "minimal". If they are not "complete", what would you suggest to the student who wrote them to increase their score? Finally, you will be provided with a rubric. Score your response and note what, if anything, you would do differently to increase your own score.

A television producer must schedule a selection of paid advertisements during each hour of programming. The lengths of the advertisements are Normally distributed with a mean of 28 seconds and standard deviation of 5 seconds. During each hour of programming, 45 minutes are devoted to the program and 15 minutes are set aside for advertisements. To fill in the 15 minutes, the producer randomly selects 30 advertisements.

a) Describe the sampling distribution of the sample mean length for random samples of 30 advertisements.

b) If 30 advertisements are randomly selected, what is the probability that the total time needed to air them will exceed the 15 minutes available? Show your work.