Chapter 7: Sampling Distributions

Objectives: Students will:
- Define a sampling distribution.
- Contrast bias and variability.
- Describe the sampling distribution of a sample proportion (shape, center, and spread).
- Use a Normal approximation to solve probability problems involving the sampling distribution of a sample proportion.
- Describe the sampling distribution of a sample mean.
- State the central limit theorem.
- Solve probability problems involving the sampling distribution of a sample mean.

AP Outline Fit:
III. Anticipating Patterns: Exploring random phenomena using probability and simulation (20%–30%)
   D. Sampling distributions
   1. Sampling distribution of a sample proportion
   2. Sampling distribution of a sample mean
   3. Central Limit Theorem
   4. Simulation of sampling distributions

What you will learn:
A. Sampling Distributions
   1. Identify parameters and statistics in a sample or experiment.
   2. Recognize the fact of sampling variability: a statistic will take different values when you repeat a sample or experiment.
   3. Interpret a sampling distribution as describing the values taken by a statistic in all possible repetitions of a sample or experiment under the same conditions.
   4. Describe the bias and variability of a statistic in terms of the mean and spread of its sampling distribution.
   5. Understand that the variability of a statistic is controlled by the size of the sample. Statistics from larger samples are less variable.
B. Sample Proportions
   1. Recognize when a problem involves a sample proportion $\hat{p}$.
   2. Find the mean and standard deviation of the sampling distribution of a sample proportion $\hat{p}$ for an SRS of size $n$ from a population having population proportion $p$.
   3. Know that the standard deviation (spread) of the sampling distribution of $\hat{p}$ gets smaller at the rate $\sqrt{n}$ as the sample size $n$ gets larger.
   4. Recognize when you can use the Normal approximation to the sampling distribution of $\hat{p}$. Use the Normal approximation to calculate probabilities that concern $\hat{p}$.
C. Sample Means
   1. Recognize when a problem involves the mean $\bar{x}$ of a sample.
   2. Find the mean and standard deviation of the sampling distribution of a sample mean $\bar{x}$ from an SRS of size $n$ when the mean $\mu$ and standard deviation $\sigma$ of the population are known.
   3. Know that the standard deviation (spread) of the sampling distribution of $\bar{x}$ gets smaller at the rate $\sqrt{n}$ as the sample size $n$ gets larger.
   4. Understand that $\bar{x}$ has approximately a Normal distribution when the sample is large (central limit theorem). Use this Normal distribution to calculate probabilities involving $\bar{x}$.
Chapter 7: Sampling Distributions

Section 7.1: What is a Sampling Distribution?

Objectives: Students will:
Distinguish between a parameter and a statistic
Create a sampling distribution using all possible samples from a small population
Use the sampling distribution of a statistic to evaluate a claim about a parameter
Distinguish among the distribution of a population, the distribution of a sample, and the sampling distribution of a statistic
Determine if a statistic is an unbiased estimator of a population parameter
Describe the relationship between sample size and the variability of a statistic

Vocabulary:
Parameter – a number that describes some characteristic of the population
Statistic – a number that describes some characteristic of a sample
Sampling variability – different random samples of the same size from the sample population will produce different values for a statistic
Sampling Distribution (of a statistic) – the distribution of values taken by the statistic in all possible samples of the same size from the same population
Unbiased estimator – a statistic whose sampling distribution mean is equal to the true value of the parameter being estimated;
Variability (of a statistic) – a description of the spread of the statistic’s sampling distribution
Bias – the level of trustworthiness of a statistic

Key Concepts:

- Population Parameters
  - Usually unknown and are estimated by sample statistics using techniques we will learn
  - Mean: \( \mu \)
  - Standard Deviation: \( \sigma \)
  - Proportion: \( p \)
- Sample Statistics
  - Used to estimate population parameters
  - Mean: \( \bar{x} \)
  - Standard Deviation: \( s \)
  - Proportion: \( \hat{p} \)

Sampling Distribution

In other words: a sampling distribution of proportions is using the proportion of an individual sample as the data point of the samples of \( \hat{p} \) – the “bigger” sample.
Example 1: Upon entry to an airport’s customs area each passenger presses a button and either a green arrow comes on (directing the passenger on through) or a red arrow comes on (directing them to a customs agent) and they have the bags searched. Homeland Security sets the “search” parameter at 30%.

a) What type of probability distribution applies here?

b) What are the mean and standard deviation of this distribution?

c) Each of you represents a day, 8 in total, that we are going to simulate a simple random sampling of 100 passengers passing through the airport. We want to know what your individual average proportion of those who got the green arrow. This we will refer to as $p\hat{}$ or $\hat{p}$. To do this we will use our calculator.

d) We can also use our calculator to simulate this and just get the total number, which represents $p\hat{}$ or $\hat{p}$.

e) Describe the distribution below

Example 2: Which of these sampling distributions displays large or small bias and large or small variability?

Homework: problems 1, 7, 13, 19
Chapter 7: Sampling Distributions

Section 7.2: Sample Proportions

Objectives: Students will:
- Calculate the mean and standard deviation of the sampling distribution of a sample proportion \( \hat{p} \) and interpret the standard deviation
- Determine if the sampling distribution of \( \hat{p} \) is approximately Normal
- If appropriate, use a Normal distribution to calculate probabilities involving \( \hat{p} \)

Vocabulary:
- Sample distribution of the sample proportion – \( \hat{p} \) describes the distribution of values taken by the sample proportion \( \hat{p} \) in all possible samples of the same size from the same population

Key Concepts:

Conclusions regarding the distribution of the sample proportion:
- Shape: as the size of the sample, \( n \), increases, the shape of the distribution of the sample proportion becomes approximately normal
- Center: the mean of the distribution of the sample proportion equals the population proportion, \( p \).
- Spread: standard deviation of the distribution of the sample proportion decreases as the sample size, \( n \), increases

Sampling Distribution of \( \hat{p} \)

ROT1: For a simple random sample of size \( n \) such that \( n \leq 0.10N \) (sample size is \( \leq 10\% \) of the population size)
- The mean of the sampling distribution of \( \hat{p} \) is \( \mu_{\hat{p}} = p \)
- The standard deviation of the sampling distribution of \( \hat{p} \) is \( \sigma = \sqrt{\frac{p(1-p)}{n}} \)

ROT2: The shape of the sampling distribution of \( \hat{p} \) is approximately normal provided \( np \geq 10 \) and \( n(1-p) \geq 10 \)

Sample Proportions, \( \hat{p} \)

- Remember to draw our normal curve and place the mean, \( \hat{p} \) and make note of the standard deviation

- Use normal cdf for less than values
- Use complement rule \( [1 - P(x<)] \) for greater than values

Vocabulary:
- Sample distribution of the sample proportion – \( \hat{p} \) describes the distribution of values taken by the sample proportion \( \hat{p} \) in all possible samples of the same size from the same population

Key Concepts:

Conclusions regarding the distribution of the sample proportion:
- Shape: as the size of the sample, \( n \), increases, the shape of the distribution of the sample proportion becomes approximately normal
- Center: the mean of the distribution of the sample proportion equals the population proportion, \( p \).
- Spread: standard deviation of the distribution of the sample proportion decreases as the sample size, \( n \), increases

Sampling Distribution of \( \hat{p} \)

ROT1: For a simple random sample of size \( n \) such that \( n \leq 0.10N \) (sample size is \( \leq 10\% \) of the population size)
- The mean of the sampling distribution of \( \hat{p} \) is \( \mu_{\hat{p}} = p \)
- The standard deviation of the sampling distribution of \( \hat{p} \) is \( \sigma = \sqrt{\frac{p(1-p)}{n}} \)

ROT2: The shape of the sampling distribution of \( \hat{p} \) is approximately normal provided \( np \geq 10 \) and \( n(1-p) \geq 10 \)
Chapter 7: Sampling Distributions

**Example 1:** Assume that 80% of the people taking aerobics classes are female and a simple random sample of \( n = 100 \) students is taken. What is the probability that at most 75% of the sample students are female?

**Example 2:** Assume that 80% of the people taking aerobics classes are female and a simple random sample of \( n = 100 \) students is taken. If the sample had exactly 90 female students, would that be unusual?

**Example 3:** According to the National Center for Health Statistics, 15% of all Americans have hearing trouble. In a random sample of 120 Americans, what is the probability at least 18% have hearing trouble?

**Example 4:** According to the National Center for Health Statistics, 15% of all Americans have hearing trouble. Would it be unusual if the sample above had exactly 10 having hearing trouble?

**Example 5:** We can check for undercoverage or nonresponse by comparing the sample proportion to the population proportion. About 11% of American adults are black. The sample proportion in a national sample was 9.2%. Were blacks underrepresented in the survey?

**Summary:** The sample proportion, \( \hat{p} \), is a random variable

- If the sample size \( n \) is sufficiently large and the population proportion \( p \) isn’t close to either 0 or 1, then this distribution is approximately normal
- The mean of the sampling distribution is equal to the population proportion \( p \)
- The standard deviation of the sampling distribution is equal to \( \sqrt{p(1-p)/n} \)

**Homework:** problems 33, 43
Chapter 7: Sampling Distributions

Section 7.3: Sample Means

Objectives: Students will:
- Calculate the mean and standard deviation of the sampling distribution of a sample mean $\bar{x}$ and interpret the standard deviation
- Explain how the shape of the sampling distribution of $\bar{x}$ is affected by the shape of the population distribution and the sample size
- If appropriate, use a Normal distribution to calculate probabilities involving $\bar{x}$

Vocabulary:
- **Sampling distribution of the sample mean** – describes the distribution of values taken by the sample mean $x$ in all possible samples of the same size from the same population
- **Central Limit Theorem** – (CLT) says that when $n$, the sample size, is large, the sampling distribution of the sample mean $\bar{x}$ is approximately Normal
- **Standard error of the mean** – standard deviation of the sampling distribution of $\bar{x}$

Key Concepts:

**Conclusions regarding the sampling distribution of X-bar:**
- Shape: normally distributed
- Center: mean equal to the mean of the population
- Spread: standard deviation less than the standard deviation of the population

**Mean and Standard Deviation of the Sampling Distribution of x-bar**
Suppose that a simple random sample of size $n$ is drawn from a large population (sample less than 5% of population) with mean $\mu$ and a standard deviation $\sigma$. The sampling distribution of $x$-bar will have a mean $\mu_{x-bar} = \mu$ and standard deviation $\sigma_{x-bar} = \sigma/\sqrt{n}$. The standard deviation of the sampling distribution of $x$-bar is called the standard error of the mean and is denoted by $\sigma_{x-bar}$.

**The shape of the sampling distribution of x-bar if X is normal**
If a random variable X is normally distributed, the distribution of the sample mean, x-bar, is normally distributed.

**Central Limit Theorem**
Regardless of the shape of the population, the sampling distribution of $x$-bar becomes approximately normal as the sample size $n$ increases. (Caution: only applies to shape and not to the mean or standard deviation)

Central Limit Theorem

Regardless of the shape of the population, the sampling distribution of $x$-bar becomes approximately normal as the sample size $n$ increases.

Caution: only applies to shape and not to the mean or standard deviation
From Sullivan: “With that said, so that we err on the side of caution, we will say that the distribution of the sample mean is approximately normal provided that the sample size is greater than or equal to 30, if the distribution of the population is unknown or not normal.”

### Summary of Distribution of $\bar{x}$

<table>
<thead>
<tr>
<th>Shape, Center and Spread of Population</th>
<th>Distribution of the Sample Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal with mean, $\mu$ and standard deviation, $\sigma$</td>
<td>Regardless of sample size, $n$, distribution of $x$-bar is normal</td>
</tr>
<tr>
<td>Population is not normal with mean, $\mu$ and standard deviation, $\sigma$</td>
<td>As sample size, $n$, increases, the distribution of $x$-bar becomes approximately normal</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shape</th>
<th>Center</th>
<th>Spread</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{x-bar} = \mu$</td>
<td>$\sigma_{x-bar} = \frac{\sigma}{\sqrt{n}}$</td>
<td>$\frac{\sigma}{\sqrt{n}}$</td>
</tr>
</tbody>
</table>
Chapter 7: **Sampling Distributions**

**Example 1:** The height of all 3-year-old females is approximately normally distributed with \( \mu = 38.72 \) inches and \( \sigma = 3.17 \) inches. Compute the probability that a simple random sample of size \( n = 10 \) results in a sample mean greater than 40 inches.

**Example 2:** We’ve been told that the average weight of giraffes is 2400 pounds with a standard deviation of 300 pounds. We’ve measured 50 giraffes and found that the sample mean was 2600 pounds. Is our data consistent with what we’ve been told?

**Example 3:** Young women’s height is distributed as a \( \text{N}(64.5, 2.5) \). What is the probability that a randomly selected young woman is taller than 66.5 inches?

**Example 4:** Young women’s height is distributed as a \( \text{N}(64.5, 2.5) \). What is the probability that an SRS of 10 young women has a mean height greater than 66.5 inches?

**Example 5:** The time a technician requires to perform preventive maintenance on an air conditioning unit is governed by the exponential distribution (similar to curve a from “in Action” slide). The mean time is \( \mu = 1 \) hour and \( \sigma = 1 \) hour. Your company has a contract to maintain 70 of these units in an apartment building. In budgeting your technician’s time should you allow an average of 1.1 hours or 1.25 hours for each unit?

**Summary:** The sample mean is a random variable with a distribution called the sampling distribution
- If the sample size \( n \) is sufficiently large (30 or more is a good rule of thumb), then this distribution is approximately normal
- The mean of the sampling distribution is equal to the mean of the population
- The standard deviation of the sampling distribution is equal to \( \sigma / \sqrt{n} \)

**Homework:** problems 53, 57, 65
Chapter 7: Review

Objectives: Students will be able to:
- Summarize the chapter
- Define a sampling distribution
- Contrast bias and variability
- Describe the sampling distribution of a sample proportion (shape, center, and spread)
- Use a Normal approximation to solve probability problems involving the sampling distribution of a sample proportion
- Describe the sampling distribution of a sample mean
- State the central limit theorem
- Solve probability problems involving the sampling distribution of a sample mean

Define the vocabulary used
Know and be able to discuss all sectional knowledge objectives
Complete all sectional construction objectives
Successfully answer any of the review exercises

Vocabulary: None new

Khan Academy Chapter Test

Homework: T7.1 – T7.10
Chapter 7: Sampling Distributions

Review Problems:

1. Based on a simple random sample of size 100, a researcher calculated the standard deviation associated with a sample proportion to be 0.08. If she increases the sample size to 400, what will be the new standard deviation associated with the sample proportion?

2. We know that \( \hat{p} \) is a/an ______________ statistic because the mean of the sampling distribution of \( \hat{p} \) is equal to the true population proportion \( p \).

3. According to the manufacturer’s specifications, the mean time required for a particular anesthetic drug to produce unconsciousness is 7.5 minutes with a standard deviation of 1.8 minutes. A random sample of 36 patients is to be selected and the average time for the drug to work will be computed for the sample. Find the probability that
   (a) the mean time for the sample will be less than 7.0 mins
   (b) a randomly selected patient requires less than 7.0
   (c) If more random samples of size 36 were selected, the middle 95% of the sample means should fall between ____________ minutes and ____________ minutes.

4. As we have discussed in class, a one-pound (16 ounce) box of sugar generally weighs more than 1 lb. According to some state laws, producers will be fined if the mean of 5 randomly selected boxes is less than 1 lb. If the packaging equipment delivers individual weights that are \( N (\mu, 0.4) \) ounces, what setting should be used for \( \mu \) so the probability of being fined is 0.01? Provide a sketch to support your answer.

5. According to the __________________________, when a simple random sample of size \( n \) is drawn from any population with mean \( \mu \) and standard deviation \( \sigma \), if \( n \) is sufficiently large the sampling distribution of the sample mean is approximately normal.

6. Place the word “true” or “false” in the blank at the end of each of the following sentences.
   (a) If the underlying population is skewed, the distribution of \( \bar{x} \) will be normal for \( n = 2 \). _________________
   (b) If the underlying population is skewed, the distribution of \( \bar{x} \) will be normal for \( n = 100 \). _________________
   (c) If the underlying population is normal, the distribution of \( \bar{x} \) will be normal for \( n = 2 \). _________________
   (d) If the underlying population is normal, the distribution of \( \bar{x} \) will be normal for \( n = 100 \). _________________

7. We know that 60% of the students in a large state university are male.
   (a) Determine the mean and standard deviation of the sampling distribution of the sample proportion of males (p-hat) when samples of 400 students are randomly selected from this population.
   (b) Verify that the formula you used for your standard deviation computation is valid in this situation. State the condition(s) that must be satisfied and convince me that all necessary conditions are met.
   (c) What is the probability that a simple random sample of 400 students will contain more than 65% males?

8. The weight of eggs produced by a certain breed of hen is \( N (60, 4) \). What is the probability that the weight of a dozen (12) randomly selected eggs falls between 700 grams and 725 grams.
Chapter 7: Sampling Distributions

5 Minute Reviews

Section 7-1:
Answer the following

1. Mean and standard deviation of a sample are denoted by

2. Mean and standard deviation of a population are denoted by

3. Rule of thumb on sample size compared to population

4. As sample sizes increase, the variability of the mean ___________.

5. Which is better, low bias, high variability or high bias low variability?

Section 7-2:

1. What is the unbiased estimator of the population proportion?

2. What is the distribution of population proportion random variable?

3. What are the two rules-of-thumb mentioned in last lesson?

4. What distribution do these ROT allow us to use?

5. If the sample size goes up by a factor of 4, what effect does it have on the sample portion standard deviation?

6. Assume that 80% of the people taking aerobics classes are female and a simple random sample of $n = 100$ students is taken. Find $P(\text{females} > 90)$.

Section 7-3:

1. What is the unbiased estimate of the population mean?

2. If the sample size is expanded nine-fold, what effect does that have on the sample standard deviation of $x$-bar?

3. When can we use a Normal distribution to answer questions involving the distribution of sample means?

4. What does the Central Limit Theorem say?

5. Young women’s height is distributed with a mean of 64.5 and a standard deviation of 2.5. What is the probability that an SRS of 50 young women’s mean height is greater than 65 inches?