Chapter 7: Sampling Distributions

(REQUIRED NOTES) **Section 7.1: What Is a Sampling Distribution?**

Key Vocabulary:

- parameter
- statistic
- sampling variability
- sample distribution
- sampling distribution
- population distribution
- margin of error
- biased estimator
- unbiased estimator
- bias
- variability
- variability of a statistic
- sample proportion
- sample mean
- central limit theorem

1) What is a parameter? What is a statistic? How is one related to the other?

- *S (sample and statistic) and*
- *P (population and parameter)*

Alert students that confusing these ideas is a very common source of lost credit on the AP

2) Explain the difference between μ and $\bar{x}$, between p and $\hat{p}$, between σ and s_x?

	sample statistic	*population parameter*
Mean (averages)	$\bar{x}$	μ
Proportion (%'s)	$\hat{p}$	p
Standard Deviation	s	σ

3) Identify the population, parameter(with notation), sample, and statistic(with notation),:

a. The Gallup Poll asked a random sample of 515 US adults whether or not they believed in ghosts. Of the respondents, 160 said "Yes."

- **Population** The population of all U.S. adults
- **Parameter** the parameter of interest is "p" - which is the proportion of all U.S. adults that believe in ghosts
- **Sample** random sample of 515 US adults
- **Statistic** $\hat{p}$ **=160/515 = .31**

b. A random sample of 100 female college students has a mean of 64.5 inches; which is greater than the 63 inch mean height of all adult American women.

- **Population** all adult American women
- **Parameter** μ=63 inch
- **Sample** random sample of 100 female college students
- **Statistic** $\bar{x}$ **=64.5 inches**

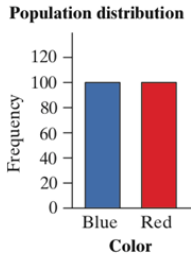
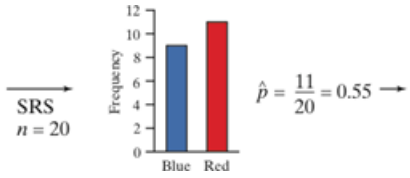
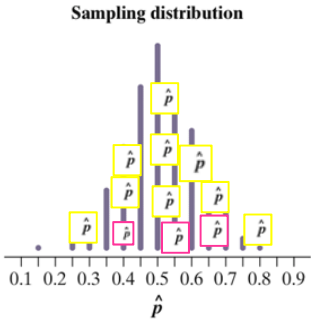
4) What is sampling variability? Why do we care?

- What is the difference between variability of the parameter and sampling variability (sample means and sample proportions)?
  - *Parameter is a constant.*
  - *For sample mean and sample proportion, the sampling variability varies from sample to sample*
- How is sampling variability related to margin of error?
  - *We need to estimate sampling variability so we know how close our estimates are to the truth—the margin of error.*

Chapter 7: Sampling Distributions

(REQUIRED NOTES) *Section 7.1: What Is a Sampling Distribution?*

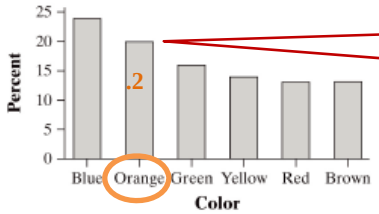
5) What is the difference between the **distribution of the population**, the **distribution of the sample**, and the **sampling distribution of a sample statistic**? *Give an example. It is helpful to sketch graphs of each! See graphs on pages 420-423.* *There are actually three distinct distributions involved when we sample repeatedly and measure a variable of interest.*

distribution of the population	distribution of the sample	sampling distribution of a sample statistic
		

- **Define** *Population Distribution; and* **sketch a graph**:
  - The **population distribution** gives the values of the variable for all the individuals in the population.
- **Define** Distribution of a sample*; and* **sketch a graph**
  - The **distribution of sample data** shows the values of the variable for all the individuals in the sample.
- **Define** Sampling distribution of a statistic*; and* **sketch a graph**:
  - The **sampling distribution** shows the statistic values from all the possible samples of the same size from the population. **It is a distribution of the statistic.**
  - **CHECK YOUR UNDERSTANDING** (page 420) complete questions 1-3

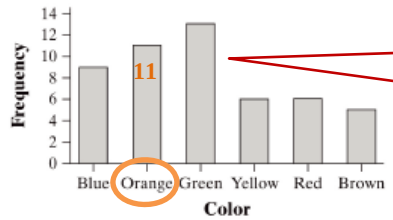
Check Your Understanding, page 420:

1. The individuals are the M&M'S® Milk Chocolate Candies, the variable is the color of the M&M and the parameter of interest is the proportion of orange M&M's.



This is the **POPULATION** distribution
p=proportion of orange M&M's
p=.20

2. For this sample there are 11 orange M&M's so $\hat{p} = \frac{11}{50} = 0.22$.



This is an example distribution for 1 **SAMPLE**
Sample Statistics
n=50 and phat=.22

3. The middle graph is the approximate sampling distribution of $\hat{p}$. The statistic measures the proportion of oranges in samples of M&M's. Assuming that the company is correct, 20% of the M&M's are orange, so the center of the distribution of $\hat{p}$ should be at approximately 0.20. The first graph shows the distribution of the colors for one sample, rather than the distribution of $\hat{p}$ from many samples, and the third graph is centered at the wrong spot.

Chapter 7: Sampling Distributions

(REQUIRED NOTES) ***Section 7.1: What Is a Sampling Distribution?***

6) Explain the difference between these 3 distributions. Why do we care sampling distributions of a statistics?

- ***Definition of "distribution" is always the same****: describes the possible values and how often the values occur.* You must always describe the center, spread, and shape of the distribution. ***Always specify which distribution you are talking about!***
  - *The* ***population distribution*** *describes the values of the variable for all individuals* ***in the population****.*
  - *The* ***sample distribution*** *describes the values of the variable for all individuals* ***in a sample*** *.*
  - *The* ***sampling distribution*** ***of a statistic*** *describes the values of the* ***statistic in all possible samples of the same size from the same population.***
  - *Advise students of the dangers of ambiguous statement "when the sample size increases, the variability decreases"—variability of* ***which distribution****??*

- ***Definition****:* ***Sampling distribution*** *of a statistic is the distribution of values taken by the statistic in* ***all*** *possible samples of the same size from the same population.*
  - **The fact that statistics from random samples have definite sampling distributions allows us to answer the question:** "How trustworthy is a statistic as an estimator of the parameter?"

7) What is an unbiased estimator? What is a biased estimator? Why do we care?

- **Define**: Unbiased Estimator - ***A statistic used to estimate a parameter is an unbiased estimator if the mean of its sampling distribution is equal to the true value of the parameter being estimated.***

- Explain the difference between Biased and Unbiased Estimators
  - *Unbiased doesn't mean perfect! Unbiased means not consistently too high or consistently too low when taking many random samples.*
  - *Biased means statistic is consistently higher or lower than the parameter.*
  - *Do not confuse with a survey sampling process (undercoverage, response, non-response) which produces biased data. We cannot correct for this poor survey design and we should not use this statistic is an estimator for the population.*

- When is a statistic considered an unbiased estimator?

  The sample proportion (or the same mean) is an **unbiased estimator for the population parameter they are equal (e.g. BIAS → statistic ≠ parameter)**

- **CHECK YOUR UNDERSTANDING** (page 426) complete questions 1-3

Check Your Understanding, page 426:

1. The median does not appear to be an unbiased estimator of the population median. The mean of the 500 sample medians is 73.5 whereas the median of the population is 75.

2. With larger samples, the spread of the sampling distribution is smaller, so increasing the sample size from 10 to 20 will decrease the spread of the sampling distribution.

3. The sampling distribution is skewed to the left. This means that, in general, underestimates of the population median will be greater than overestimates.

Chapter 7: Sampling Distributions

(REQUIRED NOTES) *Section 7.1: What Is a Sampling Distribution?*

8) What is the variability of a statistic? Why do we care?

- **Definition**: The variability of a statistic is described by the spread of its sampling distribution.
- **Why do we care?** This spread is determined primarily by the size of the random sample.

- **How can you reduce the variability of a statistic**?
  - By taking a larger sample. Larger samples give smaller spread.
  - Also by better design, such as stratified sampling.

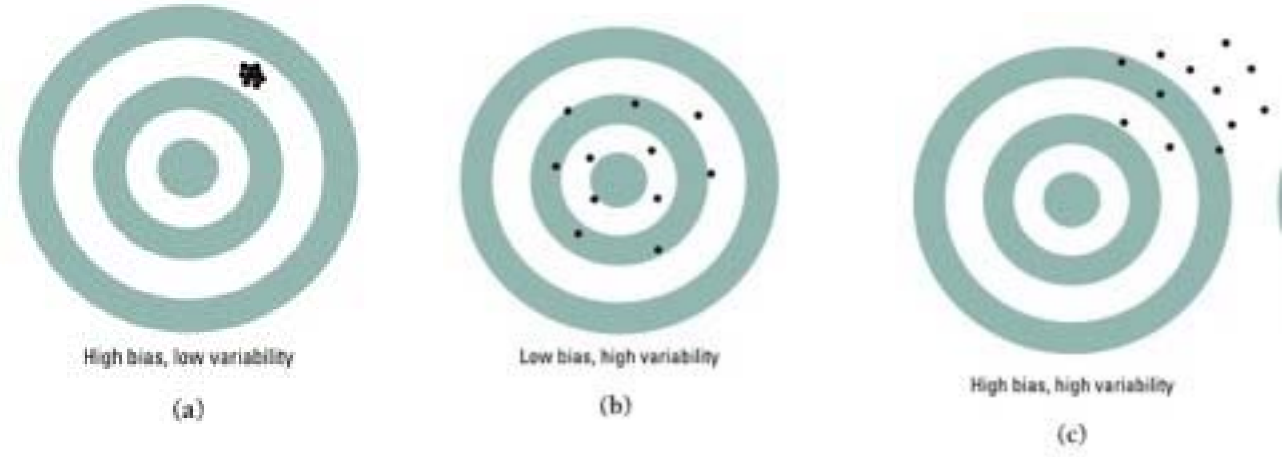
- **What effect does the size of the population have on the variability (spread) of a statistic**?
  - The spread of the sampling distribution does not depend on the size the population as long as the sample meets the 10% condition *(the population is at least 10 times larger than the sample)*

9) What is the difference between accuracy and precision? How does this relate to bias and variability?

- Accurate = unbiased
- Precise = low variability

10) Explain the difference between bias and variability. Sketch the 4 bull's eyes on page 426 and clearly explain their bias and variability.

- **Bias means** that our aim is off and we consistently miss the bull's-eye in the same direction. Our sample values do not center on the population value.
- **High variability means** that repeated shots are widely scattered on the target. Repeated samples do not give very similar results.



- What is the ideal estimator?
  - ***No or low bias***
  - ***Minimal variability***

The Ideal