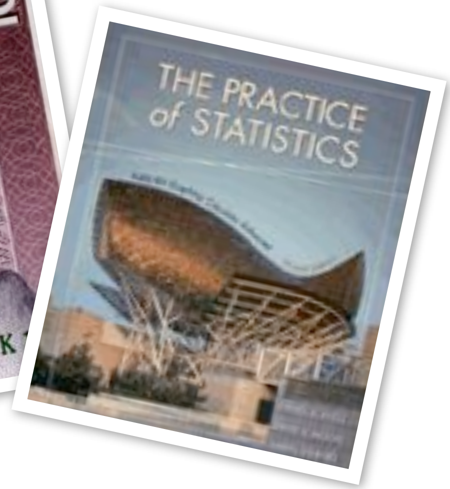
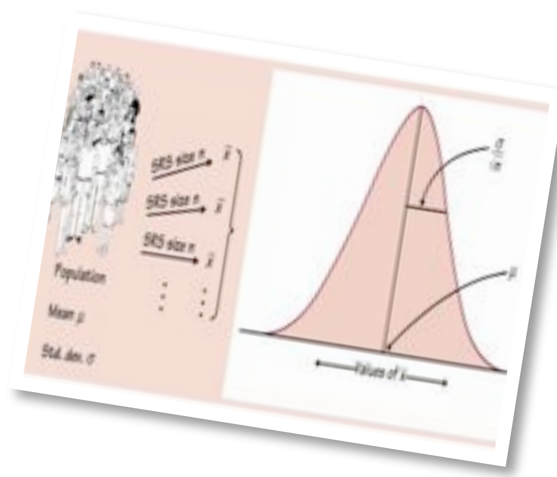


Chapter 9

Sampling Distributions

The inferential methods we will learn in the coming chapters will be based on using information from a sample to reach a conclusion about the population. In order to use this information, we must develop an understanding of how sampling information varies from sample to sample. In this chapter, we will explore the behavior of sample statistics in repeated sampling and learn one of the most important theorems in Statistics - The Central Limit Theorem.

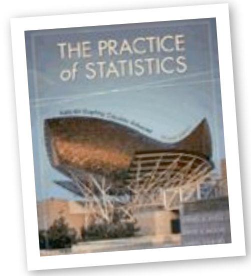


Sampling Distributions:

- 9.1: Sampling Distributions
- 9.2: Sample Proportions
- 9.3: Sample Means and The Central Limit Theorem

AP Statistics Chapter 9: Sampling Distributions

"Statistics may be defined as "a body of methods for making wise decisions in the face of uncertainty.""
~ W.A. Wallis



Tentative Lesson Guide					
Date		**Stats**	**Lesson**	**Assignment**	**Done**
Thu	1/11	9.1	Sampling Distributions	Rd 9.1 **Do 1-4, 12-15**	
Fri	1/12	9.2	Sample Proportions	Rd 9.2 **Do 19-23**	
Mon	1/15		*No School*	*Organize Semester 1 Materials*	
Tues	1/16	9.3	Sample Means	Rd 9.3 **Do 45-52**	
Wed	1/17	9.3	Central Limit Theorem	**Practice Worksheet**	
Thurs	1/18	**Rev**	**Chapter 9 Review**		
Fri	1/19	**Exam**	**Exam Chapter 9**	**Online Quiz Due**	
Mon	1/22	**Rev**	**Semester 1 Review**		
Tues	1/23	**Rev**	**Semester 1 Review**		
Wed	1/24	**Final**	***FINAL EXAM***	**50 Question MC Exam**	
Thu	1/25	**Final**	***FINAL EXAM***	**50 Question MC Exam**	

Note:
The purpose of this guide is to help you organize your studies for this chapter.
The schedule and assignments may change slightly.

Keep your homework organized and refer to this when you turn in your assignments at the end of the chapter.

Class Website:
Be sure to log on to the class website for notes, worksheets, links to our text companion site, etc.

http://web.mac.com/statsmonkey

Don't forget to take your online quiz!. Be sure to enter my email address correctly!
http://bcs.whfreeman.com/yates2e

My email address is:

jmmolesky@isd194.k12.mn.us

Chapter 9 Objectives and Skills:

These are the expectations for this chapter. You should be able to answer these questions and perform these tasks accurately and thoroughly. Although this is not an exhaustive review sheet, it gives a good idea of the "big picture" skills that you should have after completing this chapter. The more thoroughly and accurately you can complete these tasks, the better your preparation.

Sampling Distributions

- *Parameter*: An index related to a population.
- *Statistic*: An index that is related to a sample.
- *Sampling distribution of a statistic*: The distribution of values of a statistic taken from all possible samples of a specific size.
- A statistic is unbiased if the mean of the sampling distribution is equal to the true value of the parameter being estimated.
- Note: A sample standard deviation is not an unbiased estimator of the population standard deviation.

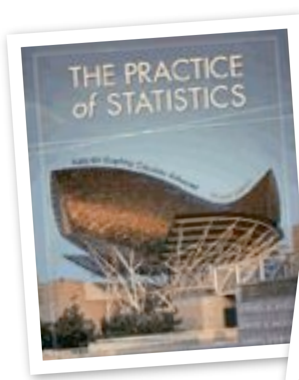
Sample Proportions

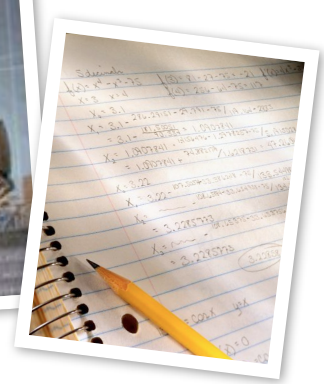
- If we choose an SRS of size n from a large population with population proportion p having some characteristic of interest, and if p(hat) is the proportion of the sample having that characteristic, then
- The sampling distribution of p(hat) is approximately normal. N(p, sqrt[p(1-p)/n])
- *It is reasonable to use the above statements when: pop>10n, np>10, n(1-p)>10*

Sample Means

THE CENTRAL LIMIT THEOREM

- Consider an SRS of size n from any population with mean μ and standard deviation s. When n is large (n>30 and pop>10n), the sampling distribution of x(bar) has the following properties:
  - (a) it is approximately normal.
  - (b) the mean of the distribution is μ.
  - (c) the standard deviation of the distribution is s/sqrt(n), where s is the standard deviation of the population.



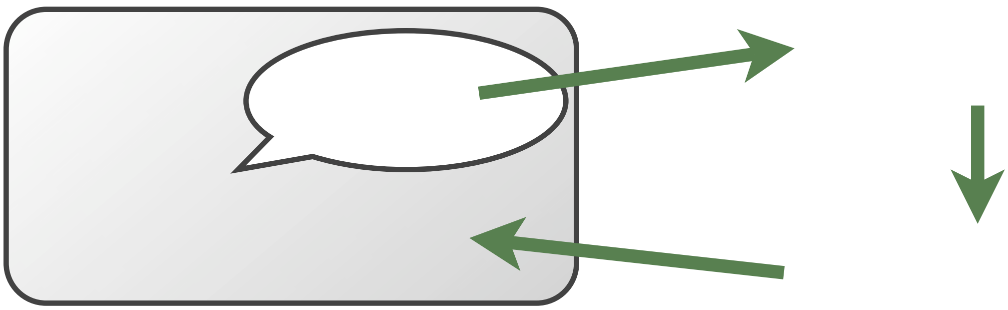


9.1: Sampling Distributions

The usual way to gain information about a population characteristic is to select a sample from the population. However, we must note that the sample information we gather may differ somewhat from the population characteristic we are trying to measure. Further, the sample information may differ from sample to sample. This sample-to-sample variability poses a problem when we try to generalize our findings to the population. In order to do so, we must gain an understanding of this variability.

- **Sample Statistic:**

- **Population Parameter:**



We can view a sample statistic as a random variable. That is, we have no way of predicting *exactly* what statistic value we will get from a sample, but, given a population parameter, we know how those values will behave in repeated sampling. If we could find *all possible* samples of a given size from a population, we could find the corresponding distribution of statistic values.

- **Sampling Distribution:**

We have no way of knowing whether or not our statistic value is equal to the parameter we are trying to estimate. We must be aware of the bias and variability of our sampling distribution. Then we can use the information about the sample to reach a conclusion about the parameter.

  - **Bias:**
  - **Variability:**

9.2: Sample Proportions

The objective of some statistical applications is to reach a conclusion about a population proportion, π. 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 π is unknown to us, we must base our conclusion on a sample proportion, p. However, as we have noted, we know that the value of p will vary from sample to sample. The amount of variability will depend on the size of our sample.

A little bit of theory...

The Sampling Distribution for Proportions:

If we take repeated random samples of size *n* from a population, the sample proportion, *p*, will have the following distribution and properties:

Note: The Sampling Distribution will be approximately Normal if nπ and n(1-π) are both greater than 10.

Example: Based on Census data, we know 11% of US adults are Black. Therefore, π = 0.11 We would expect a sample to contain roughly 11% Black representation.
Suppose a sample of 1500 adults contains 138 Black individuals. Should we suspect 'undercoverage' in the sampling method?

Note, $\hat{p} = \frac{138}{1500} = .092$. Is this lower than what would be expected by chance? That is, we know it is possible that a sample could contain 9.2% Black representation…but is it likely that would happen due to natural variation in a random sampling method?

☑ Check Assumptions:

nπ>10?	1500(.11)=165	YES
n(1-π)>10?	1500(.89)=1335	YES

$\mu = .11$

$\sigma = \sqrt{\frac{.11(.89)}{1500}} = .00808$

☑ Assume the Sampling Distribution of $\hat{p}$ is approximately Normal.

☑ Calculate Probability P($\hat{p}$ ≤ .092) = P(z ≤ -2.223) = 0.0129

☑ Interpret:

Only 1.29% of samples of size 1500 would have less than 9.2% Black representation. Since this is so unlikely, we have reason to suspect possible undercoverage in this sample.

9.2: Sample Proportion Practice

Directions: Use what you know about the sampling distribution of a sample proportion to answer the following questions. Be sure to write up all four steps neatly!
1) Assumptions 2) Sampling Distribution (show sketch) 3) Probability calculation 4) Interpretation of results.

1. The development of viral hepatitis subsequent to a blood transfusion can cause serious complications for a patient. The article, "Hepatitis in Patients with Acute Nonlymphatic Leukemia" (Amer. J. of Med. (1983): 413-421) reported that in spite of careful screening for those having a hepatitis antigen, viral hepatitis occurs in 7% of blood recipients. Suppose a new treatment is believed to reduce the incidence of viral hepatitis. The treatment is given to 200 blood recipients and only 6 contract hepatitis. Does it appear that the treatment is effective? That is, is it very likely that we would observe only 6/200 contract hepatitis when 7% of the population is known to do so?

2. The article "Should Pregnant Women Move? Linking Risks for Birth Defects with Proximity to Toxic Waste Sites" (Chance (1992): 40-45) reported that in a large study carried out in the state of New York, approximately 30% of the population lived within 1 mile of a hazardous waste site. If an SRS of 400 pregnant women is selected, how likely is it that the sample proportion will be within 5% of the true population proportion? Would this probability be larger or smaller if we selected an SRS of size 500? (You don't need to do a calculation to figure this out...use common sense!)

3. The article "Thrillers" (Newsweek, Apr. 22, 1985) states, "Surveys tell us that more than half of America's college graduates are avid readers of mystery novels." Assume the true proportion is exactly 0.5. What is the probability that an SRS of 225 college graduates would give a sample proportion greater than 0.6?

4. Suppose that a particular candidate for public office is in fact favored by 48% of all registered voters in a sizable metropolitan district. A polling organization takes an SRS of 500 voters and will use the sample proportion to estimate the population parameter. What is the probability that the sample proportion will be greater than 0.5, causing the polling organization to incorrectly predict the results of the upcoming election?

5. The Gallup Poll once asked an SRS of 1540 adults, "Do you happen to jog?" Suppose 15% of all adults jog. Find the probability the poll gave a result within 2% of the actual population proportion.

6. Suppose 47% of all adult women think they do not get enough time to themselves. An opinion poll interviews 1025 randomly selected women and records the sample proportion who feel they don't get enough time for themselves. If this sample were repeated numerous times, in what range would the middle 95% of the sample results fall? What is the probability the poll gets a sample in which fewer than 45% say they do not get enough time for themselves?

7. Voter registration records show 68% of all voters in Indianapolis are registered as Republicans. A random digit dialing device is used to call 150 randomly chosen residential homes in Indianapolis. Of the 150 registered voters who are contacted, 73% are registered Republicans. Should you suspect the random digit dialing device of favoring phone numbers of Republicans?

9.3: Sample Means

When the objective of a statistical application is to reach a conclusion about a population mean, μ, we must consider a sample mean, $\bar{x}$. However, as we have noted, we know that the value of $\bar{x}$ will vary from sample to sample. The amount of variability will depend on the size of our sample.

A little bit of theory...

The Sampling Distribution for Means:

If we take repeated random samples of size *n* from a population, the sample proportion, $\bar{x}$, will have the following distribution and properties:

Note: The Sampling Distribution of $\bar{x}$ will be approximately Normal if n>30.

Example:

Let x denote the time it takes a 5th grader to complete a math problem. Suppose the mean and standard deviation are μ=2min and std dev=0.8 min respectively.

Let $\bar{x}$ be the sample average time for 9 students. Describe the sampling distribution of $\bar{x}$.

Suppose we take a SRS of 20 students. Describe the sampling distribution of $\bar{x}$. Use it to find the probability that $\bar{x}$ is greater than 2.5 min. for the sample of 20 students.

9.3: The Central Limit Theorem

When dealing with sample means from sufficiently large samples, a useful fact about the sampling distribution arises...regardless of the shape of the population distribution.

Activity: A Penny for Your Thoughts

Your teacher has a jar containing several hundred pennies. Each penny has a 'birthdate' indicated by the date on its face. Your task is to take repeated samples of pennies to determine the average birthdate of the pennies in the jar. You will take 5 samples of varying sizes and record the average birthdate on the dotplots on the board. Sketch the dotplots below and note any patterns you see emerging.

Sample Set 1: Size n=5

$\bar{x}$

Sample Set 2: Size n=15

$\bar{x}$

Sample Set 3: Size n=30

$\bar{x}$

Based on the dotplots above, what is your best guess about the average birthdate of the pennies in the jar? What conclusions can you reach about the sampling distribution of $\bar{x}$ as n gets bigger?

Note: The distribution of birthdates for the population looks like this:

The Central Limit Theorem "CLT"

If n is sufficiently large (n>30), then the sampling distribution of $\bar{x}$ has the following properties:

9.3: Sample Mean Practice

Directions: Use what you know about the sampling distribution of a sample means (ie, the Central Limit Theorem) to answer the following questions. Be sure to write up all four steps neatly!
1) Assumptions 2) Sampling Distribution (show sketch) 3) Probability calculation 4) Interpretation of results.

1. According to the article, "Song Dialects and Colonization in the House Finch" (Condor (1975): 407-422) reported the mean value of song duration for the population of house finches is 1.5 min with a standard deviation of .9 min. Suppose an SRS of 25 finches is selected. How likely is it that the average song duration of the sample will be greater than 1.7 min?

2. A soft drink bottler claims that, on average, cans contain 12 oz. of soda. Suppose the true distribution of soda volumes is normally distributed with a mean of 12 oz. and a standard deviation of .16 oz. Sixteen cans are randomly selected and their volumes are measured. What is the probability the average volume will be between 11.96 and 12.08 oz.?

3. A hot dog manufacturer claims its most popular brand of weenie has an average fat content of 18 g per hot dog. Suppose the standard deviation of the fat content of all hot dogs is 1 g. An independent testing organization selects an SRS of 36 hot dogs and finds the average fat content is 18.4 g. Does this result indicate the manufacturer's claim is incorrect?

4. The time a randomly selected individual waits for an elevator in an office building has a uniform distribution with a mean of 0.5 min and standard deviation of 0.289 min. What are the mean and standard deviation of the sampling distribution of means for SRS of size 50? Does it matter that the underlying population distribution is not normal? What is the probability a sample of 50 people will wait longer than 45 seconds for an elevator?

5. A manufacturing process is designed to produce bolts with a 0.5 in diameter. Once each day, a random sample of 36 bolts is selected and the average diameter is calculated. If the sample mean is less than 0.49 in or greater than 0.51 in, the process is shut down for adjustment. The standard deviation for the production process is .02 in. What is the probability the process will be shut down on any given day?

6. Suppose the mean value of interpupillary distance for all adult males is 65 mm and the population 5 mm. If 25 adult males are selected, what is the probability the average distance for all 25 will be between 64 and 67 mm?

7. Using the information from problem 6, answer the following questions:
   a) Approximately 95% of the time, the sample mean falls between_______ and _______

   b) Approximately .3% of the time, the sample mean is farther than _________ from the true mean.