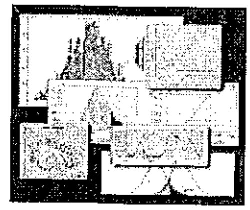
AP STATISTICS

Syllabus Fall 2014



Unit 6 SAMPLING DISTRIBUTIONS

Allen Key

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

Day	Date	Objective	Section	Assignment	
1	F 10/30	Introduction to Sampling Distributions	4th/5th Ed. 7.1 2nd Ed.9.1	Collect Data for Introductory Activity 9.1 Guided Notes	
2	M 11/2	Sampling Distributions	4th/5th Ed. 7.1 2nd Ed. 9.1	Explore Sampling Read Pages 487 - 499	Unit 6 Day 1 Practice Pg 4-6
3	T 11/3	Sampling Distributions (Proportions)	4th/5th Ed. 7.2 2nd Ed. 9.2	Reese Pieces (7) Read Pages 504-510 Pg 8-9	Sampling Proportions questions Pg. 10-11
4	W 11/4	Sampling Proportions Practice **QUIZ**	Review for Quiz (Finish Day 1-3)		
5	TH 11/5	Sample Means	4th/5th Ed. 7.3 2nd Ed. 9.3	Sample Mean Practice Read Pages 514 - 523	Unit 9 FRAPPY Sampling (Mean) Questions
6	F 11/6	Sample Means and Mixed Practice	4th/5th Ed. 7.3 2nd Ed. 9.3	Central Limit Theorem Pages: 16-25B	
7	M 11/9	Review	9.1 – 9.3	Review ALL	
8	T 11/10	Unit 6 Test	Preview Unit 7		

http://onlinestatbook.com/stat_sim/sampling_dist/index.html
http://www.rossmanchance.com/applets/Reeses/ReesesPieces.html

AP Statistics – Chapter 9 Notes: Sampling Distributions

9.1 – Sampling Distributions

Parameter – a number that describes a population (usually unknown)
Statistics – a number that describes a sample (used to estimate a parameter)

Symbols used	Sample Statistic	Population Parameter
Proportions	$\hat{p}$	p
Means	$\bar{x}$	μ

Sampling Distribution – the distribution of all values taken by a statistic in all possible samples of the same size from the same population

A statistic is called an ***unbiased estimator*** of a parameter if the mean of its sampling distribution is equal to the parameter being estimated

Important Concepts for unbiased estimators

- The mean of a sampling distribution will always equal the mean of the population for any sample size
- The spread of a sampling distribution is affected by the sample size, *not the population size*. Specifically, larger sample sizes result in smaller spread.

9.2 – Sample Proportions

Choose an SRS of size n from a large population with population proportion p having some characteristic of interest.

Let $\hat{p}$ be the proportion of the sample having that characteristic. Then the mean and standard deviation of the sampling distribution of $\hat{p}$ are

Mean: $\mu_{\hat{p}} = p$

Std. Dev.: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$

CONDITIONS FOR NORMALITY

Rule Of Thumb 1
Use the formula for the standard deviation of $\hat{p}$ only when the size of the population is at least 10 times as large as the sample size.

Rule Of Thumb 2
We will use the normal approximation to the sampling distribution of $\hat{p}$ for values of n and p that satisfy $np \geq 10$ and $n(1-p) \geq 10$.

9.3 – Sample Means

Suppose that $\bar{x}$ is the mean of a sample from a large population with mean μ and standard deviation σ. Then the mean and standard deviation of the sampling distribution of $\bar{x}$ are

Mean: $\mu_{\bar{x}} = \mu$

Std. Dev.: $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

CONDITIONS FOR NORMALITY

If an SRS is drawn from a population that has the normal distribution with mean μ and standard deviation σ, then the sample mean $\bar{x}$ will have the normal distribution $N(\mu, \sigma/\sqrt{n})$ for any sample size.

Central Limit Theorem
If an SRS is drawn from any population with mean μ and standard deviation σ, when n is large $(n \geq 30)$, the sampling distribution of the sample mean $\bar{x}$ will have the normal distribution $N(\mu, \sigma/\sqrt{n})$.

9.3 The Central Limit Theorem

Theorem: If x possesses *any* distribution with mean μ and standard deviation σ, then the sample mean $\bar{x}$ based on a random sample of size n will have a distribution that approaches the distribution of a normal random variable with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ as n increases without limit.

The theory behind this glorious theorem...

The central limit theorem is indeed surprising! It says that x can have *any* distribution whatever, but as the sample size gets larger and larger, the distribution of $\bar{x}$ will approach a *normal* distribution. From this relation, we begin to appreciate the scope and significance of the normal distribution.

In the central limit theorem, the degree to which the distribution of $\bar{x}$ values fit a normal distribution depends on both the selected value of n and the original distribution of x values. A natural question is: How large should the sample size be if we want to apply the central limit theorem? After a great deal of theoretical as well as empirical study, statisticians agree that if n is 30 or larger, the $\bar{x}$ distribution will appear to be normal and the central limit theorem will apply. However, this rule should not be applied blindly. If the x distribution is definitely not symmetrical about its mean, then the $\bar{x}$ distribution also will display a lack of symmetry. In such a case, a sample size larger than 30 may be required to get a reasonable approximation to the normal.

In practice, it is a good idea, when possible, to make a histogram of sample x values. If the histogram is approximately mound-shaped, and if it is more or less symmetrical, then we may be assured that, for all practical purposes, the $\bar{x}$ distribution will be well approximated by a normal distribution and the central limit theorem will apply when the sample size is 30 or larger. The main thing to remember is that in almost all practical applications, a sample size of 30 or more is adequate for the central limit theorem to hold. However, in a few rare applications, you may need a sample size larger than 30 to get reliable results.

AP STATISTICS
Unit 6 Introduction

SAMPLING DISTRIBUTIONS

Instructions: The results of each activity below should be:
- 3 separate histograms
- Comments following the second graph
- A paragraph description of the 3rd histogram.

This process is tedious and will require some time to do correctly. Please be patient and do a good job. These activities will be the basis of a large part of the rest of our course.

Activity 1: Label this activity: **Sampling Distribution of Proportion**
Each student should have all the recorded information and a copy of each graph in your notes.

- Use your calculator to simulate tossing a coin 10 times and record the proportion of heads that occur in the 10 tosses.
- Repeat the process until you have 10 proportions
- Create a frequency histogram of the sample proportions write down each proportion
- Add 25 proportions to the 10 you had originally
- Create a new histogram. Note differences from the 1st histogram
- Continue collecting sample proportions until you have a total of 100 samples
- Create a 3rd histogram. Comment as you did in chapter one

Activity 2: Label this activity: **Sampling Distribution of Mean**
Each student should have all the recorded information and a copy of each graph in your notes.

- Take a random sample of 5 digits from the random integers 0 – 9. Find the mean
- Repeat the process until you have 10 means
- Create a frequency histogram of the means
- Add 25 means to the 10 you had originally
- Look at the new histogram. Comment on any differences from the 1st histogram
- Continue collecting sample means until you have a total of 100 samples
- Create a 3rd histogram. Comment as you did in chapter one

Exploring Sampling Distributions

Task 1:

Go to the Sampling Distribution applet at www.ruf.rice.edu/~lane/stat_sim/sampling_dist/. Click on "Begin" on the left-hand size of the screen. From the Population pull-down menu, select one of the predefined distributions. Click on animated sample a few times and watch what happens. Then try 5 samples, 1000 samples, 10,000 samples. Adjust the sample size using the pull-down menu next to the distribution of sample means. Notice the values displayed to the left of the population distribution and the sample means distribution. Take a few more samples and watch how the values change.

The purpose of this part is for you to get familiar and comfortable with the various features of this demonstration stack. There will be nothing to write up for Task 1.

Task 2:

Select the "normal" population. Record the population mean, μ, and population standard deviation, σ. Specify sample size 2, and simulate 1000 samples of size 2. Record the sample mean, $\bar{x}$, and the sample standard deviation *s*. Compare $\bar{x}$ with μ and compare *s* with the calculated value of $\frac{\sigma}{\sqrt{n}}$. Click on the **Clear** button, and repeat the process with the new sample sizes of 9, 16, and 25. Make a table to record all of the results for later analysis.

What shapes do the sampling distributions have? As the sample size increases from 2 to 9 to 16 to 25, what is happening to the *center* of the sampling distribution? What is happening to the *spread* of the sampling distribution? What can you say about the quantity $\frac{\sigma}{\sqrt{n}}$?

Task 3:

One might suspect that sampling from a mounded (i.e. approximately normal) distribution would produce a mound-shaped sampling distribution. But if you start with a nonsymmetric distribution, would your sampling distribution be nonsymmetric, too? Select the "skewed" population, and repeat the steps of Task 2. Answer the same questions as before.

Task 4:

This time, start with a custom-made population distribution of your own choosing. Repeat the steps of Task 2 and answer the same questions.

Task 5:

Summarize what you have learned by this investigation. In particular, make a conjecture that relates the mean of the sampling distribution to the population you're drawing from. Also make a conjecture that relates the spread (i.e. standard deviation) of the sampling distribution to the host population. Finally, discuss how the shape of the sampling distribution relates to the shape of the parent population.

Task 2: "Normal" μ = ________ σ = ________

n	$\bar{x}$	s	$\frac{\sigma}{\sqrt{n}}$	$\bar{x}$ vs. μ	s vs. $\frac{\sigma}{\sqrt{n}}$
2					
9					
16					
25					

Conclusions:

Task 3: Nonsymmetric μ = ________ σ = ________

n	$\bar{x}$	s	$\frac{\sigma}{\sqrt{n}}$	$\bar{x}$ vs. μ	s vs. $\frac{\sigma}{\sqrt{n}}$
2					
9					
16					
25					

Conclusions:

Task 4: Custom-Made μ = ________ σ = ________

n	$\bar{x}$	s	$\frac{\sigma}{\sqrt{n}}$	$\bar{x}$ vs. μ	s vs. $\frac{\sigma}{\sqrt{n}}$
2					
9					
16					
25					

Conclusions:

Summary of what you have learned...

Unit 6 Day 1 – Sampling Distributions Practice   1-4 give population, parameter sample, statistic

1. **Stop smoking!** A random sample of 1000 people who signed a card saying they intended to quit smoking were contacted nine months later. It turned out that 210 (21%) of the sampled individuals had not smoked over the past six months.

Sample 1000
Statistic 210 (21%) who stopped
population: people who signed card to stop smoking   parameter: actual number who stopped

2. **Unemployment** Each month, the Current Population Survey interviews a random sample of individuals in about 55,000 U.S. households. One of their goals is to estimate the national unemployment rate. In December 2009, 10.0% of those interviewed were unemployed. population: 55000 US households

sample # of houses selected of 55000   parameter the % of unemployment
Statistic 10% of the interview were unemployed

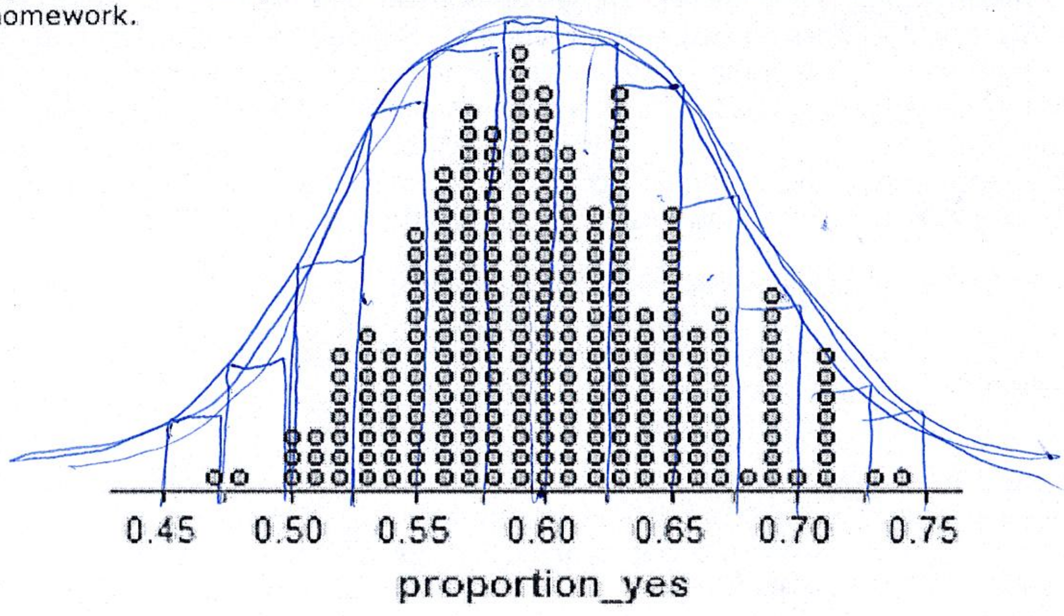
3. **Hot turkey** Tom is cooking a large turkey breast for a holiday meal. He wants to be sure that the turkey is safe to eat, which requires a minimum internal temperature of 165°F. Tom uses a thermometer to measure the temperature of the turkey meat at four randomly chosen points. The minimum reading in the sample is 170°F. population temp of turkey   sample 4 spots tested

parameter 165° min   statistic 170° is min. found

4. **Gas prices** How much do gasoline prices vary in a large city? To find out, a reporter records the price per gallon of regular unleaded gasoline at a random sample of 10 gas stations in the city on the same day. The range (maximum – minimum) of the prices in the sample is 25 cents.

pop. gas prices in large city
para max-min range in large city
sample: 10 gas stations in city   statistic: range of 25¢ in sample

5. **Doing homework** A school newspaper article claims that 60% of the students at a large high school did all their assigned homework last week. Some skeptical AP Statistics students want to investigate whether this claim is true, so they choose an SRS of 100 students from the school to interview. What values of the sample proportion To find out, we used Fathom software to simulate choosing 250 SRSs of size n = 100 students from a population in which p = 0.60. The figure below is a dotplot of the sample proportion $\hat{p}$ of students who did all their homework.



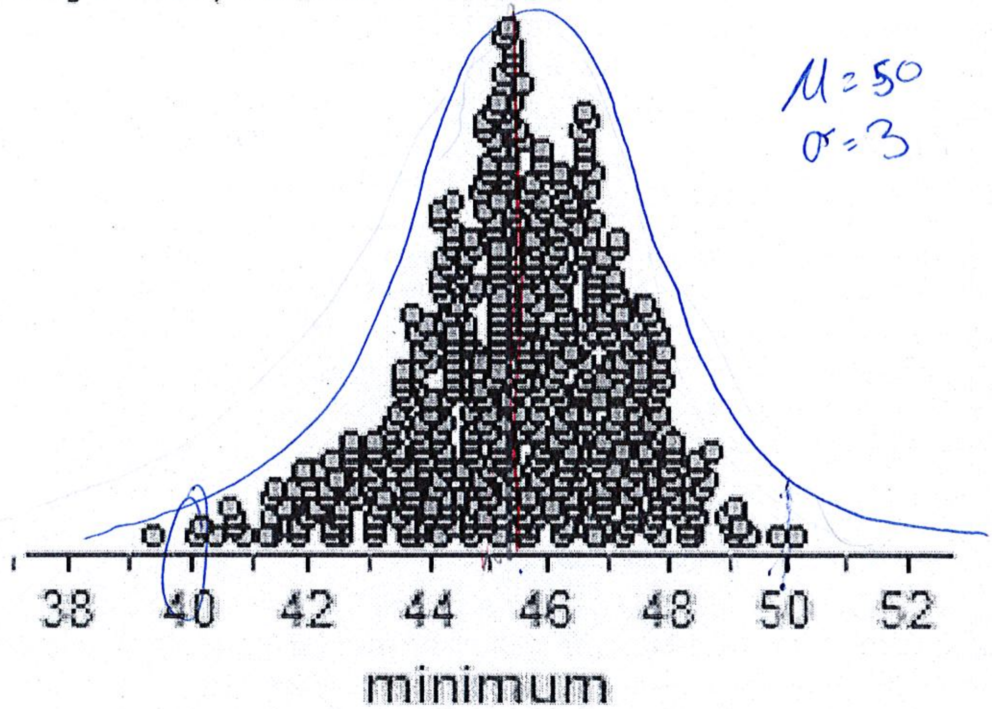
(a) Is this the sampling distribution of $\hat{p}$? Justify your answer. NO because there are more than 250 SRSs that can be taken but this is an approximation

(b) Describe the distribution. Are there any obvious outliers? SOCS

(c) Suppose that 45 of the 100 students in the actual sample say that they did all their homework last week. What would you conclude about the newspaper article's claim? Explain. That the claim was to high and should be lower. Because none of the original 250 SRS had only 45% who did all HW

4

6. **Cold cabin?** The Fathom screen shot below shows the results of taking 500 SRSs of 10 temperature readings from a population distribution that's $N(50, 3)$ and recording the sample minimum each time.



$\mu = 50$
$\sigma = 3$

SOCS
(a) Describe the approximate sampling distribution.
single peaked, no apparent outliers or abnormalities

(b) Suppose that the minimum of an actual sample is 40°F. What would you conclude about the thermostat manufacturer's claim? Explain.
That the claim minmum is too high. Only 1 or 2 of the sample minimums were at 40 or below. Since so unlikely to occur, but did, the claim could be wrong

7. **Run a mile** During World War II, 12,000 ablebodied male undergraduates at the University of Illinois participated in required physical training. Each student ran a timed mile. Their times followed the Normal distribution with mean 7.11 minutes and standard deviation 0.74 minute. An SRS of 100 of these students has mean time $\overline{x} = 7.15$ minutes. A second SRS of size 100 has mean $\overline{x} = 6.97$ minutes. After many SRSs, the values of the sample mean $\overline{x}$ follow the Normal distribution with mean 7.11 minutes and standard deviation 0.074 minute.

would indiv all 12,000 runners times

(a) What is the population? Describe the population distribution.
The 12,000 able bodied male undergrads a U of I
parameter is μ of 7.11

(b) Describe the sampling distribution of $\overline{x}$. How is it different from the population distribution? center peaked around 7.11 min
sampling distributions made up of many SRS (of 100) of the 12,000 where as the pop dist

8. **IRS audits** The Internal Revenue Service plans to examine an SRS of individual federal income tax returns from each state. One variable of interest is the proportion of returns claiming itemized deductions. The total number of tax returns in each state varies from over 15 million in California to about 240,000 in Wyoming.

(a) Will the sampling variability of the sample proportion change from state to state if an SRS of 2000 tax returns is selected in each state? Explain your answer. no because not effected by pop. size
variabity of statistic (proportion) does not depend on pop size

(b) Will the sampling variability of the sample proportion change from state to state if an SRS of 1% of all tax returns is selected in each state? Explain your answer.
yes because sample size changes based on state size
larger sample size less variability

5

9. **Predict the election** Just before a presidential election, a national opinion poll increases the size of its weekly random sample from the usual 1500 people to 4000 people.

(a) Does the larger random sample reduce the bias of the poll result? Explain.

(b) Does it reduce the variability of the result? Explain.

10. **A sample of teens** A study of the health of teenagers plans to measure the blood cholesterol levels of an SRS of 13- to 16-year-olds. The researchers will report the mean $\overline{x}$ from their sample as an estimate of the mean cholesterol level μ in this population.

(a) Explain to someone who knows no statistics what it means to say that $\overline{x}$ is an unbiased estimator of μ.

(b) The sample result $\overline{x}$ is an unbiased estimator of the population mean μ no matter what size SRS the study chooses. Explain to someone who knows no statistics why a large random sample gives more trustworthy results than a small random sample.

11. If we take a simple random sample of size n = 500 from a population of size 5,000,000, the variability of our estimate will be

(a) much less than the variability for a sample of size n = 500 from a population of size 50,000,000.

(b) slightly less than the variability for a sample of size n = 500 from a population of size 50,000,000.

(c) about the same as the variability for a sample of size n = 500 from a population of size 50,000,000.

(d) slightly greater than the variability for a sample of size n = 500 from a population of size 50,000,000.

(e) much greater than the variability for a sample of size n = 500 from a population of size 50,000,000.

The Reese's Pieces Applet

Go to www.rossmanchance.com/applets/, and look under Sampling Distribution Simulations on the left side of the screen. Click on **Reese's Pieces** (newer version). You will see a big container of colored candies: that represents the POPULATION.

Set the proportion of orange candies to 0.45 (use the first box). This is the **population parameter**. (People who have counted lots of Reese's pieces came up with this number). Change the sample size to 30 so that it would resemble samples taken by our class. Check the "count samples" box. Click on the "draw samples" button. One sample of 30 candies will be taken and the proportion of means for this sample is plotted on the graph. Repeat this again.

- Do you get the same or different values for each sample?
- How close is each **sample statistic** (proportion) to the **POPULATION PARAMETER**?

Conjecture: What happens to a sampling distribution when we increase the number of samples?

Turn off the animation (checked box that says animate) and change the number of samples to 100.
Click on draw samples, and see the distribution of sample statistics built.

- Describe its **shape, center and spread.**

Draw more samples and note the changes, if any, to the shape, center and spread of the distribution.

Conjecture: What happens to a sampling distribution when we increase the size of the samples?

First, change the sample size to 10 and draw 100 samples

- How close is each **sample statistic** (proportion) to the **POPULATION PARAMETER**?

Next, change the sample size to 100 and draw 100 samples.

- How close is each **sample statistic** (proportion) to the **POPULATION PARAMETER**?

What happens to this distribution of sample statistics as we change the number of candies in each sample (sample size).

V. Conclusions

As the number of samples increase, what happens to how well the sample statistics resemble the population parameter?

As the sample size increases, what happens to how well the sample statistics resemble the population parameter?

AP STATISTICS NAME ____________________
Unit 6 - PROPORTIONS

Suppose you have a population in which 50% of the people approve of gambling.

1. Is the value .50 a parameter or a statistic? What notation would you use for this value?

You want to take many samples of size 10 from this population to observe how the sample proportion who approve of gambling varies in repeated samples.

2. Describe the design of a simulation using the partial random digits table below to estimate the sample proportion who approve of gambling. Then carry out five trials of your simulation.

   Design:

   Results: $\hat{p}_1 =$ $\hat{p}_2 =$ $\hat{p}_3 =$ $\hat{p}_4 =$ $\hat{p}_5 =$

According to your design, mark above the numbers those that are "yes":

36009	19365	15412	39638	85453	46816
38448	48789	18338	24697	39364	42006
82739	57890	20807	47511	81676	55300
60940	72024	17868	24942	81790	90656
68417	35013	15529	72765	85089	57067

3. The sampling distribution of $\hat{p}$ is the distribution of $\hat{p}$ from all possible SRSs of size 10 from this population. What would be the mean of this distribution?

4. If you used samples of size 20 instead of size 10, which sampling distribution would give you a better estimate of the true proportion of people who approve of gambling? Explain your answer.

8

Suppose you are going to roll a fair six-sided die 70 times and record $\hat{p}$, the proportion of times that a 1 or a 2 are showing.

5. Where should the distribution of 70 $\hat{p}$-values be centered? Explain your answer.

$\mu_{\hat{p}} = p =$

6. What is the standard deviation of the sampling distribution of $\hat{p}$? Justify your answer.

7. Describe the shape of the sampling distribution of $\hat{p}$. Justify your answer.

approx normal

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 15% of sycamore trees. The power company tests the chemical on 250 sycamores. Consider these an SRS from the population of all sycamore trees.

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

$\mu_{\hat{p}} = .15$

$\sigma_{\hat{p}} = \sqrt{}$

9. What is the probability that at least 45 trees are killed? Justify your answer.

9

AP Statistics
Sample Proportions Questions

Name ______________________________

1. Which of the following is true regarding the variation of a sampling distribution of a sample proportion?

how to explain

   a) Variation depends on population size as well as sample size.
   b) The variance of a sampling distribution of a sample proportion for all samples of size 1 is 0. $n=1$ $\sigma_{\hat{p}} = 0$
   c) As the size of the sample increases, the variation of the sampling distribution approaches the variation of the population.
   d) For a given sample size, the maximum variation in the sampling distribution of a sample proportion occurs when the sample proportion is .5.
   e) None of these is true.

2. The conditions that $np \geq 10$ and $n(1 - p) \geq 10$ are imposed on a sampling distribution to protect against
   a) a sample that is not representative of the population.
   b) bias in the responses of the sample participants.
   c) skewness in the distribution.
   d) a very small population size.
   e) the conditions are not designed to protect against any of these conditions.

$p = .2$
$n = 100$
$\hat{p}_n = .25$

3. Records at a large university indicated that 20% of all freshmen are placed on academic probation at the end of their first semester. A random sample of 100 of this year's freshmen indicated that 25% of them were placed on academic probation at the end of the first semester. The results of this sample
   a) are surprising since it indicates that 5% more of these freshmen were placed on academic probation than was expected.
   b) are surprising since SAT scores have been increasing over the past few years.
   c) are not surprising since the standard deviation of the sampling distribution is ~~4%~~. .04
   d) are biased since the increase of 5% could not happen without injecting bias into the sample.

$\sigma = \sqrt{\frac{p(1-p)}{n}}$

8.6% that .25% or more will be on

4. An investigator anticipates that the proportion of red blossoms in his hybrid plants is .15. A random sample of 50 of his plants indicated that 22% of the blossoms were red. The standard deviation of the sampling distribution of the sample proportion is approximately
   a) .051
   b) .059
   c) .07
   d) .116
   e) cannot be determined.

$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = .0585$

.05049

10

Free Response Questions

For each of the following, verify assumptions for the sampling distribution of a proportion. If assumptions are not met, state why and go on to the next problem. If assumptions are met, solve the problem.

1. A manufacturer of cold medicine claims that 60% of all adults suffer at least one cold during every winter. What is the probability that a simple random sample of 200 adults will report that 65% or more of the subjects had at least one cold last winter? Would you be surprised to find such a sample result? Explain your answer.

2. A company claims that 5% of its products will be shipped in defective condition. A simple random sample of 50 products contained 5 defective products. What is the probability that 5 or more defective products could have happened by random chance?

3. A mathematics department published the claim that a minimum of 70% of students enrolled in their classes receive a final grade of C- or better in any semester. A simple random sample of 50 students from the department's classes indicated that only 65% of the students had a final grade of C- or better last semester. What is the probability that a sample of this size will have a result that differs from the claimed proportion by more than 5% (above or below)? Would such a result surprise you? State a conclusion. solve

$n \le .10N$

$50 \le .10$(all math students)

$np \ge 10$

$n(1-p) \ge 10$

$\mu_{\hat{p}} = p = .7$

$.7(50) \ge 10$

$.3(50) \ge 10$

$\sigma_{\hat{p}} = \sqrt{\frac{(.7)(.3)}{50}}$

$n = 200$

$N = 1000$

$p = .035$

4. A simple random sample of size 200 is taken from a population of 1,000 people in a professional organization regarding preferences on the issue of raising dues. What is the probability that this sample will produce a result of 10% or less favoring the raise when it is known that 3.5% of the population favors the raise?

$P(\hat{p} \le .10)$

$n \le .10N$

$200 \le .10(1000)$ - doesn't pass

$\mu_{\hat{p}} = p = .035$

cannot use

$\sigma_{\hat{p}}$ $n=?$

5. Find the size of a simple random sample needed so that the probability that its proportion differs from the population proportion by more than 2% (above or below) is .1. Assume that the population proportion is .63.

$p = .63$

$n =$

$P(.61 \ge \hat{p} \le .65) = .1$

± 1.645 + .0505

$n \le .10N$

$np \ge 10$

$n(1-p) \ge 10$

$\sigma_{\hat{p}} = \sqrt{\frac{(.63)(.37)}{n}}$

$-1.645 = \frac{x - .63}{\sigma}$

$1.645 = \frac{x - .63}{\sigma}$

$\sigma = .012158$

$\sigma = .01215$

$n = 1578$

11

AP Statistics
Sample Mean Practice

Name ______________________

1. Consider a normal population with mean = 150 and sd = 30. What is the probability that a randomly selected data value will have a value between 90 and 155? $N(150, 30)$ $P(90 \le x \le 155)$

2. Now suppose that a sample of size n = 20 is taken. What is the probability that this sample will have a mean value between 90 and 100? $P(90 \le \bar{x} \le 100)$

$\mu_{\bar{x}} = \mu = 150$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{30}{\sqrt{20}}$

3. Consider the approximately normal population of heights of male college-age students. Assume that the individual heights are N (69, 2.5). A random sample of 16 heights is obtained.
$n=16$
   a) Find the mean of this sampling distribution.

   $\mu_{\bar{x}} = \mu = 69$

   b) Find the standard deviation of this sampling distribution.

   $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{16}} =$

   $n \le .10N$
   $10 \le .10$(male college students)

   c) What is the shape of this sampling distribution? Why?

   approx. normal because the population is approx normal, so the sample dist. of means is also normal

   d) Find $P(\bar{x} > 70)$. $= 1 - P\left(z \le \frac{70-69}{.625}\right)$

   $= 1 - P(z \le 1.6?)$

   $= .054?$

   e) Find $P(\bar{x} \le 67)$. $= P\left(z \le \frac{67-69}{.6}\right)$

   $= P(z \le -3.?)$

12

AP Statistics
Sampling Mean Questions

Name ______________________

1. A sample of size 49 is drawn from a normal population with a mean of 63 and a standard deviation of 14. What are the mean and standard deviation of the distribution of sample means?
   a) mean = 9; standard deviation = 2.
   b) mean = 63; standard deviation = .286.
   c) mean = 63; standard deviation = 2.
   d) mean = 1.286; standard deviation = 3.5.
   e) mean = 9; standard deviation = 14.

2. A sample of size 25 is drawn from a normal population with a mean of 62. If the standard deviation of the distribution of sample means is 3.5, what is the standard deviation of the original population?
   a) .056
   b) .408
   c) 2.48
   d) 17.5
   e) 87.5

3. The distribution of SAT Math scores of students taking Calculus I at a large university is skewed left with a mean of 625 and a standard deviation of 44.5. If random samples of 100 students are repeatedly taken, which statement best describes the sampling distribution of sample means.
   a) Normal with mean of 625 and standard deviation of 44.5.
   b) Normal with mean of 625 and standard deviation of 4.45.
   c) Shape unknown with a mean of 625 and standard deviation of 44.5.
   d) Shape unknown with a mean of 625 and standard deviation of 4.45.
   e) No conclusion can be drawn since the population is not normally distributed.

4. Which of the following statements regarding the sampling distribution of sample means is *incorrect*?
   a) The sampling distribution is approximately normal when the population is normal or the sample size is sufficiently large.
   b) The mean of the sampling distribution is the mean of the population.
   c) The standard deviation of the sampling distribution is the standard deviation of the population.
   d) The sampling distribution is found by taking repeated samples of the same size from the population of interest and computing the mean of each sample.
   e) All of these are correct.

5. After repeated observations, it has been determined that the waiting time at the drive-through window of a local bank on Friday afternoons between 12:00 noon and 6:00 pm is skewed left with a mean of 3.5 minutes and a standard deviation of 1.9 minutes. A sample of 100 customers is to be taken next Friday. What is the probability that the mean of the sample will exceed 4 minutes?
   a) .0042
   b) .0396
   c) .042
   d) .396
   e) The probability cannot be determined using a normal curve approximation.

13

Free-Response Questions

In the following problems, check assumptions. If assumptions are not met, explain why and go on to the next problem. If assumptions are met, solve the problem and state a conclusion.

There is a .0153% chance a sample of 40 has a mean of 285.16, based on this I believe the average cash has not changed.

1. A college bookstore claims that data it has collected for some time indicate that the mean amount of money spent on books by each student per semester is $251.92 with a standard deviation of $58.21. A simple random sample of 40 current students found that they had spent an average of $285.16 with a standard deviation of $63.15. Find the probability that a sample of this size will have an average of $285.16 or more if the average amount spent by the population is $251.92. Do you think the average amount spent has changed? Provide appropriate statistical evidence to support your position.

$\mu = 251.92$ $\sigma = 58.21$

$P(\bar{x} \geq 285.16) = \text{normalcdf}(\text{lower } 285.16, 1e99, 251.92, 9.204)$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{58.21}{\sqrt{40}} = 9.204$ $= .000153$

$n = 40$ $\bar{x} = 285.16$ $s = 63.15$ CLT $n \geq 30$ = approx normal

2. If the mean age of retirees from the teaching profession is 57 years with a standard deviation of 5.6, find the minimum sample size so that the probability that the sample mean exceeds 55 years is .6.

$\mu = 57$ $\sigma = 5.6$ $n = ?$

$P(\bar{x} \geq 55) = .6$

$.4 = z = -.253$

$-.253 = \frac{55 - 57}{\sigma}$

$7.905 = \frac{5.6}{\sqrt{n}}$

$\sigma = 7.905$

$n < 1$

should be smaller

3. Current research indicates that the distribution of the life expectancies of a certain protozoan is normal with a mean of 45.5 days and a standard deviation of 3.4 days.
   a) Find the probability that a protozoan selected at random will live for 49 or more days.

   $P(X \geq 49 \text{ days}) = \text{normalcdf}(\text{lower } 49, \text{upper } 1e99, \text{mean } 45.5, \sigma\ 3.4) = .1516$

   b) Find the probability that a simple random sample of 50 protozoa will have a mean life expectancy of 49 or more days.

   $P(\bar{x} \geq 49)$

   $\mu_{\bar{x}} = 45.5$ $\sigma_{\bar{x}} = \frac{3.4}{\sqrt{50}} = .4808$

   $\text{normalcdf}(\text{lower } 49, \text{upper } 1e99, 45.5, .4808) = 0$

4. At a large bank, account balances are normally distributed with a mean of $1,637.52 and a standard deviation of $623.16. What is the probability that a simple random sample of 400 accounts has a mean that exceeds $1,650?

$\mu = 1637.52$ $\sigma = 623.16$ $n = 400$

$P(\bar{x} \geq 1650) = \text{normalcdf}(\text{lower } 1650, \text{upper } 1e99, 1637.52, 31.158)$ $= .3444$

$\sigma_{\bar{x}} = \frac{623.16}{\sqrt{400}} = 31.158$ $\mu_{\bar{x}} = \mu = 1637.52$

There is a 34% chance a sample of 400 will have a mean 1650 or over.

5. At a large factory, it is found that the mean hourly wage is $16.50 with a standard deviation of $1.50. What is the probability that a random sample of 25 workers will have a mean hourly wage of less than $15.75?

$\mu = 16.50$ $\sigma = 1.50$ $n = 25$

$P(\bar{x} \leq 15.75)$

don't know if normal

sample less than 30 so CLT does not apply

6. SAT Math scores are normally distributed with a mean of 500 and a standard deviation of 100.
   a) What is the probability that a random sample of 100 students has a mean score between 490 and 510?

   $P(490 \leq \bar{x} \leq 510)$ $n = 100$ $\sigma = \frac{100}{\sqrt{100}} = \frac{100}{10} = 10$ $\approx .6827$

   b) If a random sample of 100 students had a mean SAT Math score of 525, would you describe these students as exceptional? Explain your answer.

   $n = 100$

   $P(\bar{x} \geq 525)$ $= \text{normalcdf}: \text{lower } 525, \text{upper } 1e99, 500, 10$

   $= .006$

   Being that a sample of 100 has a .6% chance of having a mean of 525, yes I would say they are exceptional as it is unlikely

14

AP STATISTICS/Kohler
Chapter 9 "FRAPPY" {Free Response AP Problem...Yay!}

NAME ______ DATE ______

Big Town Fisheries recently stocked a new lake in a city park with 2,000 fish of various sizes. The distribution of the lengths of these fish is approximately normal.

Scoring:

E P I

a) Big Town Fisheries claims that the mean length of the fish is 8 inches. If the claim is true, which of the following would be more likely?

- A random sample of 15 fish having a mean length that is greater than 10 inches, OR
- A random sample of 50 fish having a mean length that is greater than 10 inches.

Justify your answer.

$\mu_{\bar{x}} = \mu = 8$ $\quad P(\bar{x} > 10)$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$:

Due to sample size being different, and a larger sample decreases variability (spread), meaning a smaller standard deviation, there is a better chance (more likely) a sample of 15 will result in a mean greater than 10 in.

E P I

b) Suppose the standard deviation of the sampling distribution of the sample mean for random samples of size 50 is 0.3 inches. If the mean length of the fish is 8 inches, use the normal distribution to compute the probability that a random sample of 50 fish will have a mean length less than 7.5 inches.

$n = 50$ $\quad \sigma_{\bar{x}} = .3$

$\mu_{\bar{x}} = \mu = 8$

$P(\bar{x} < 7.5 \text{ in}) =$ normalcdf(lower -1e99, upper 7.5, mean 8, std. .3)

$= .0478$

E P I

c) Suppose the distribution of fish lengths in this lake was non-normal but had the same mean and standard deviation. Would it still be appropriate to use the normal distribution to compute the probability in part b? Justify your answer.

Yes, because of the central limits theorem stating if samples are large ($n \geq 30$) then the distribution of sample means is normal.

TOTAL: ____/4

15

AP STATISTICS
SAMPLE MEAN PRACTICE

NAME ____________________________

1. The heights of 18-year old men are approximately normally distributed with mean 68 inches and standard deviation 3 inches.
   a) What is the probability that an 18-year old man selected at random is between 67 and 69 inches tall? = normalcdf(67, 69, 68, 3) =
   b) If a random sample of nine 18-year-old men is selected, what is the probability that the mean height is between 67 and 69 inches? normalcdf(67, 69, 68, $\frac{3}{\sqrt{9}}$) =
   c) Compare your answers for parts (a) and (b). Explain the reason for the difference.

2. Let x be a random variable that represents level of glucose in the blood after a 12-hour fast. Assume for people under 50 years old that x has a distribution that is approximately normal with mean 85 and an estimated standard deviation of 25. A test result less than 40 is an indication of severe excess insulin, and medication is usually prescribed.
   a) What is the probability that on a single test x < 40? normalcdf(-1e99, 40, 85, 25)
   b) Suppose that a doctor uses the average for two tests taken about a week apart. What can we say about the probability distribution of $\bar{x}$?
   c) What is the probability that $\bar{x} < 40$? normalcdf(-1e99, 40, 85, $\frac{25}{\sqrt{2}}$)
   d) Repeat (c) for 3 tests taken a week apart.
   e) Repeat (c) for 5 tests taken a week apart. change

3. Let x be a random variable that represents weights in kilograms of healthy adult female deer in December in Mesa Verde National Park. Then x has a distribution that is approximately normal with mean 63 kg and standard deviation 7.1. Suppose a doe that weighs less than 54 kg is considered undernourished.
   a) What is the probability that a single doe captured at random in December is undernourished? normalcdf(-1e99, 54, 63, 7.1) =
   b) If the park has about 2200 does, what number do you expect to be undernourished in December? answer to (a) · 2200
   c) To estimate the health of the December doe population, park rangers use the rule that a sample of 50 does should be more than 60 kg. If the average weight is less than 60 kg, it is thought the entire population of does might be undernourished. What is the probability that the average weight of a sample of 50 does is less than 60 kg? normalcdf(-1e99, 60, 63, $\frac{7.1}{\sqrt{50}}$)
   n=50
   d) Suppose park rangers captured, weighed, and released 50 does in December, and the average weight was 64.2. Do you think the doe population is undernourished or not? Explain. normalcdf(64.2, 1e99, 63, $\frac{7.1}{\sqrt{50}}$)

4. n=50

AP Statistics
More Sample Mean Practice

Name ______________________________

1. Consider a variable X that is known to be normally distributed with mean 100 and standard deviation 20. What is the probability that a randomly selected data value will have a value between 90 and 100?

2. Referring to problem #1, what if a sample of size 16 is taken. What is the probability that this sample will have a mean value between 90 and 100?

3. Kindergarten children have heights that are approximately normally distributed with mean 39 inches and standard deviation 2 inches.
   a) If an individual kindergartner is selected at random, what is the probability that s/he has a height between 38 and 40 inches?

   b) A randomly selected classroom of 30 of these children is used as a sample. What is the probability that the class mean is between 38 and 40?

   c) If an individual kindergartner is selected at random, what is the probability that s/he is taller than 40 inches?

   d) A randomly selected classroom of 30 of these children is used as a sample. What is the probability that the mean is greater than 40 inches?

4. Consider the experiment of taking the STAT ("Statistics Aptitude Test"). The random variable X is the raw score received. Assume that X is N(720, 60).
   a) An individual student takes the exam. What is the probability that she scores less than 700? How rare an event would this be?
   b) A randomly selected group of 100 students takes the exam. What is the probability that the mean score of the group is less than 700? How rare of an event would this be?

5. IQ's of male college-age students are known to have mean 102 and standard deviation 111.
   a) One student is chosen at random. What is the probability that his IQ will be over 105?

   b) An SRS of size 45 is chosen from the college, and its mean is calculated. What is the probability that this mean IQ will be over 105?

6. From a sample of 50 employees, taken in a random manner from all of the employees of a large firm, the mean weekly earnings was $295.30. Given the current wage structure, it has been estimated in labor negotiations that the population standard deviation of is $34.10. What is the probability of the sample outcome or less, if the population mean is really $275.00, as the negotiator insists? Justify your calculations and conclusion.

7. Tirza's Tires, a local manufacturer of automobile parts, claims that – based on years of experience with a particular brand of tires – the mean mileage for its popular "Sevilla Radial" is 35,000 miles and the standard deviation is 5000. A consumer agency randomly buys 100 of these tires and finds a mean of 33,000.
   a) If the consumer agency did the sample again, would they get the same mean?

   b) According to the Central Limit Theorem, how is $\bar{x}$ distributed?

   c) What would be the probability of a <u>single tire</u> having a lifetime of 33,000 miles or worse?

   d) Assuming that Tirza is telling the truth, calculate the probability of the event happening as described – the <u>sample mean</u> being 33,000 miles or worse.

   e) What conclusion do you draw based on your answer to (d)?

8. A trucking firm delivers appliances for a large retail operation. The packages (or crates) have a mean weight of 300 lbs. and a standard deviation of 50 lbs. If the truck has a capacity of 8 tons, what is the probability that it will be able to carry an entire load of 50 crates? (Note: 1 ton = 2000 lbs).

AP Statistics
Sampling Distribution Practice

Name ____________________

Use what you know about sampling distributions to answer the following questions. Be sure to write all four steps neatly!
1) Assumptions 2) Sampling Distribution (show sketch) 3) Probability calculation
4) Interpretation of results in context of the problem (Write a sentence)

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?

3. 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?

4. 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 that the average volume will be between 11.96 and 12.08 oz?

5. The article "Thrillers" (Newsweek, Apr.22, 1985) states, "Surveys tell us that more than half of American'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?

6. A hot dog manufacturer claims its most popular brand of weenie has an average fat content of 18g 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 18.4g. Does this result indicate the manufacturer's claim is incorrect?

19

7. Suppose that a particular candidate for public office is in fact favored by 48% of all registered voters in a sizable metropolitan district. A poling 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?

8. The Gallup Poll once asked an SRS of 1540 adults, "Do you happen to jog?" Suppose 15% of all adults job. Find the probability the poll gave a result within 2% of the actual population proportion. $P(.13 < \hat{p} < .17)$

9. 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 that a sample of 50 people will wait longer than 45 seconds for an elevator?

no CLT $P(\bar{x} > 45)$

10. 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?

11. 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 feels 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?

12. 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?

$\mu_{\bar{x}} = 65$

13. 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 if farther than _______ from the true mean.

2.5 + 97.5

14. 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?

$P = .68$ $n = 150$ $\hat{p} = .73$

$P(\hat{p} > .73)$

Unit 6 Review

$\sigma_{\bar{x}}$ .10 $(n \le .1N)$ $np \ge 10$ and $n(1-p) \ge 10$ $\hat{p}$

1. Kindergarten children have heights that are approximately Normally distributed with $\mu = 39''$ and $\sigma = 2''$. **Do the drawings for each of these.**

a. If an individual kindergartener is selected at random, what is the probability that s/he has a height between 38 and 40 inches?

$P(38 \le X \le 40)$ = normalcdf: (lower 38, upper 40, mean 39, std 2)

$\mu = 39$ $\sigma = 2$ = .3829

b. A randomly selected classroom of 30 of these children is used as a sample. What is the probability that the class mean $\bar{x}$ is between 38 and 40? $n = 30$ $\mu_{\bar{x}} = \mu = 39$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{30}} = .36514$

$P(38 \le \bar{x} \le 40)$ = normalcdf (lower 38, upper 40, mean 39, std .36514)

= .9938

c. If an individual kindergartener is selected at random, what is the probability that s/he is taller than 40 inches? $P(x > 40)$ = normalcdf (lower 40, upper 1e99, mean 39, std 2)

= .3085

d. A randomly selected classroom of 30 of these children is used as a sample. What is the probability that $\bar{x}$ is greater than 40 inches?

$P(\bar{x} > 40)$ = normalcdf (lower 40 upper 1e99, mean 39, std .36514)

= .0031

2. IQ's of male college-age students are known to be mean $\mu = 102$ and $\sigma = 11$.

a. One student is chosen at random. What is the probability that his IQ will be over 105?

$P(x > 105) = 1 - P(z \le \frac{105-102}{11})$ = 1 − .6074

$= 1 - P(z \le .27)$ = .3925

b. An SRS of size n=45 is chosen from the college, and its mean calculated. What is the probability that this mean IQ will be over 105? $\mu_{\bar{x}} = \mu = 102$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{11}{\sqrt{45}} = 1.64$

$P(\bar{x} > 105) = 1 - P(z \le \frac{105-102}{1.64})$ = 1 − .9663

$= 1 - P(z \le 1.83)$ = .0337

3. From a sample of 50 employees, taken in a random manner from all of the employees of a large firm, the mean weekly earnings $\bar{x}$ was \$259.30. Given the current wage structure, it has been estimated in labor negotiations that the standard deviation is \$34.10 (use this for σ). What is the probability of the sample outcome or less, if the population mean is really \$275.00, as the management negotiator insists? That is, if μ = \$275.00, what is Pr($\bar{x}$ <259.30)? (Notice that this problem doesn't tell us that the population of individual earnings is Normally distributed! Why are we allowed to use the Normal distribution when answering this question about $\bar{x}$?)

CLT $n \ge 30$ $\mu = 275.00$ $\sigma = 34.10$ $\mu_{\bar{x}} = \mu = 275$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{34.10}{\sqrt{50}} = 4.823$

$P(\bar{x} \le 259.30)$ = normalcdf (lower −1e99, upper 259.30, mean 275, std 4.823)

= .000566

4. Clay is rolling dice – six at a time! Let X be the number of 1's that he gets. Describe the distribution of X, and find the probability that he gets at least two 1's in a single roll of the six dice. (Yes, this is a binomial problem; don't be alarmed – it's for review.)

0 - (0 1 2 3 4 5 6) $P(X \ge 2)$ = 1 − binomcdf: (n=6, p=1/6, 1) = .2632

5. A mathematics department published the claim that a minimum of 70% of students enrolled in their classes received a final grade of C- or better in any semester. A simple random sample of 50 students from the department's classes indicated that only 65% of the students had a final grade of C- or better last semester. What is the probability that a sample of this size will have a result that differs from the claimed proportion by more than 5% (above or below)? Would such a result surprise you? State a conclusion.

There is a .4404 chance a sample of 50 will have proportion differing from claimed proportion by 5% or more. Being such a high chance, the result does not surprise me

$n = 50$
$p = .7$

$P(65\% \ge \hat{p} \ge 75\%)$ = .4404

$\mu_{\hat{p}} = \mu = .7$ $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{.0042}$

21

200

6. A company ~~claims that 5%~~ of its products will be shipped in defective condition. A simple random sample of ~~50~~ products contained ~~5~~ defective products. What is the probability that ~~5~~ or more defective products could have happened by random chance?

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 an SRS from the population of all sycamore trees. What is the probability that at least 60 trees (24% of the sample) are killed? (Remember to follow the steps discussed in class).

= .0569

8. According to government data, 22% of American children under the age of 6 live in households with incomes less than the official poverty level. A study in early childhood chooses an SRS of 300 children. What is the probability that more than 20% of the sample are from poverty households? (Again, remember to follow the steps discussed in class.)

= .798

22

Worksheet 9.3A AP Statistics Name:

The weights of newborn children in the United States vary according to the normal distribution with mean 7.5 pounds and standard deviation 1.25 pounds. The government classifies a newborn as having low birth weight if the weight is less than 5.5 pounds.

1. What is the probability that a baby chosen at random weighs less than 5.5 pounds at birth?

You choose three babies at random and compute their mean weight, $\bar{x}$.

2. What are the mean and standard deviation of the mean weight $\bar{x}$ of the three babies?

3. What is the probability that their average birth weight is less than 5.5 pounds?

4. Would your answers to 1, 2, or 3 be affected if the distribution of birth weights in the population were distinctly nonnormal?

Worksheet 9.3C AP Statistics Name: ______________

The distribution of actual weights of 8-ounce chocolate bars produced by a certain machine is normal with mean 8.1 ounces and standard deviation 0.1 ounces. Company managers do not want the weight of a chocolate bar to fall below 7.85 ounces, for fear that consumers will complain.

1. Find the probability that the weight of a randomly selected candy bar is less than 7.85 ounces.

Four candy bars are selected at random and their mean weight, $\bar{x}$, is computed.

2. Describe the center, shape, and spread of the sampling distribution of $\bar{x}$.

3. Find the probability that the mean weight of the four candy bars is less than 7.85 ounces.

4. Would your answers to 1, 2, or 3 be affected if the weights of chocolate bars were distinctly nonnormal?

Central Limit Problems

1. Consider a normal population with mean 100 and standard deviation 20. What is the probability that a randomly selected data value will be between 90 and 100?
   P(90 < X < 100)

   normalcdf(90, 100, 100, 20) =

2. Now suppose that a sample of size n = 16 is taken from the same population. What is the probability that this sample will have a mean value between 90 and 100?
   P(90 < $\bar{x}$ < 100)

   normal

   = normalcdf(90, 100, 100, $\frac{20}{\sqrt{16}}$) =

3. Kindergarten children have heights that are approximately normally distributed with a mean of 39 inches and a standard deviation of 2 inches.
   a) If an individual kindergartener is selected at random, what is the probability that s/he has a height between 38 and 40 inches?

      P(38 < x < 40) = normalcdf(38, 40, 39, 2) =

   b) A randomly selected classroom of 30 of these children is used as a sample. What is the probability that the class mean is between 38 and 40 inches?

      P(38 < $\bar{x}$ < 40) = normalcdf(38, 40, 39, $\frac{2}{\sqrt{30}}$) =

   c) If an individual kindergartener is selected at random, what is the probability that s/he is taller than 40 inches?

      P(X > 40) = normalcdf(40, 1E99, 39, 2) =

   d) A randomly selected classroom of 30 of these children is used as a sample. What is the probability that the class mean is greater than 40 inches?

      P($\bar{x}$ > 40) normalcdf(40, 1E99, 39, $\frac{2}{\sqrt{30}}$)

4. Consider the experiment of taking a standardized mathematics exam. The random variable "X" is the raw score received. Assume that X is N(720, 60).

   a) An individual student takes the exam. What is P(she scores less than 700)? How rare of an event would this be?

      P(X < 700) = normalcdf(-1E99, 700, 720, 60)

   b) A randomly selected group of 100 students takes the exam. What is the probability that the mean score of the group is less than 700? How rare of an event would this be?

      P($\bar{X}$ < 700) = normalcdf(-1E99, 700, 720, $\frac{60}{\sqrt{100}}$) = 0

      very rare

25

5. Consider the approximately normal population of heights of male college-age students. Assume that the individual heights are N(69, 2.5). A random sample of 16 heights is obtained.

   a) Find the mean of this sampling distribution.

   $\mu_{\bar{x}} = \mu = 69$

   b) Find the standard deviation of this sampling distribution.

   $\sigma_{\bar{x}} = \frac{2.5}{\sqrt{16}} = .625$ 10% 16 ≤ .1 (male college)

   c) What is the shape of this sampling distribution? Why?

   Pop normal so sample is normal (describe normal)

   d) Find P($\bar{x}$ > 70).

   = .0548

   e) Find P ($\bar{x}$ ≤ 67).

   .000687

6. From a sample of 50 employees, taken in a random manner from all of the employees of a large firm, the mean weekly earnings (the sample mean) was \$259.30. Given the current wage structure, it has been estimated in labor negotiations that the standard deviation is \$34.10 (use this for sigma). What is the probability of the sample outcome or less, if the population mean is really \$275.00, as the management negotiator insists? Justify your process.

   $P(\bar{x} \le 259.30)$ = normalcdf(, 259.30, 275, $\frac{34.10}{\sqrt{50}}$)

   $\mu_{\bar{x}} = 275$ = .000566 unlikely

   $\sigma_{\bar{x}} = \frac{34.10}{\sqrt{50}}$

7. Tarnowski's Tires, a local manufacturer claims, based on years of experience with a particular brand of tires, that its mean mileage is 35,000 and the standard deviation is 5000. A consumer agency randomly buys 100 of these tires and finds a mean of 31,000. Should the agency believe Tarnowski's claim?

   μ :

   σ -

   $\mu_{\bar{x}} = 35000$ $P(\bar{x} \le 31000)$

   = normalcdf (-1e99, 31000, 35000, $\frac{5000}{\sqrt{100}}$) = 0

   $\sigma_{\bar{x}} = \frac{5000}{\sqrt{100}}$ — very unlikely to happen

   claim is too high

8. A trucking firm delivers appliances for a large retail operation. The packages (or crates) have a mean weight of 300 lbs and a variance of 2500 lbs. If the truck has a capacity of 4 tons, what is the probability that it will be able to carry an entire load of 25 crates?

   $\mu_{\bar{x}} = 300$ $\sigma^2 = 2500$ $\sigma = 50$ $\sigma_{\bar{x}} = \frac{50}{\sqrt{25}}$

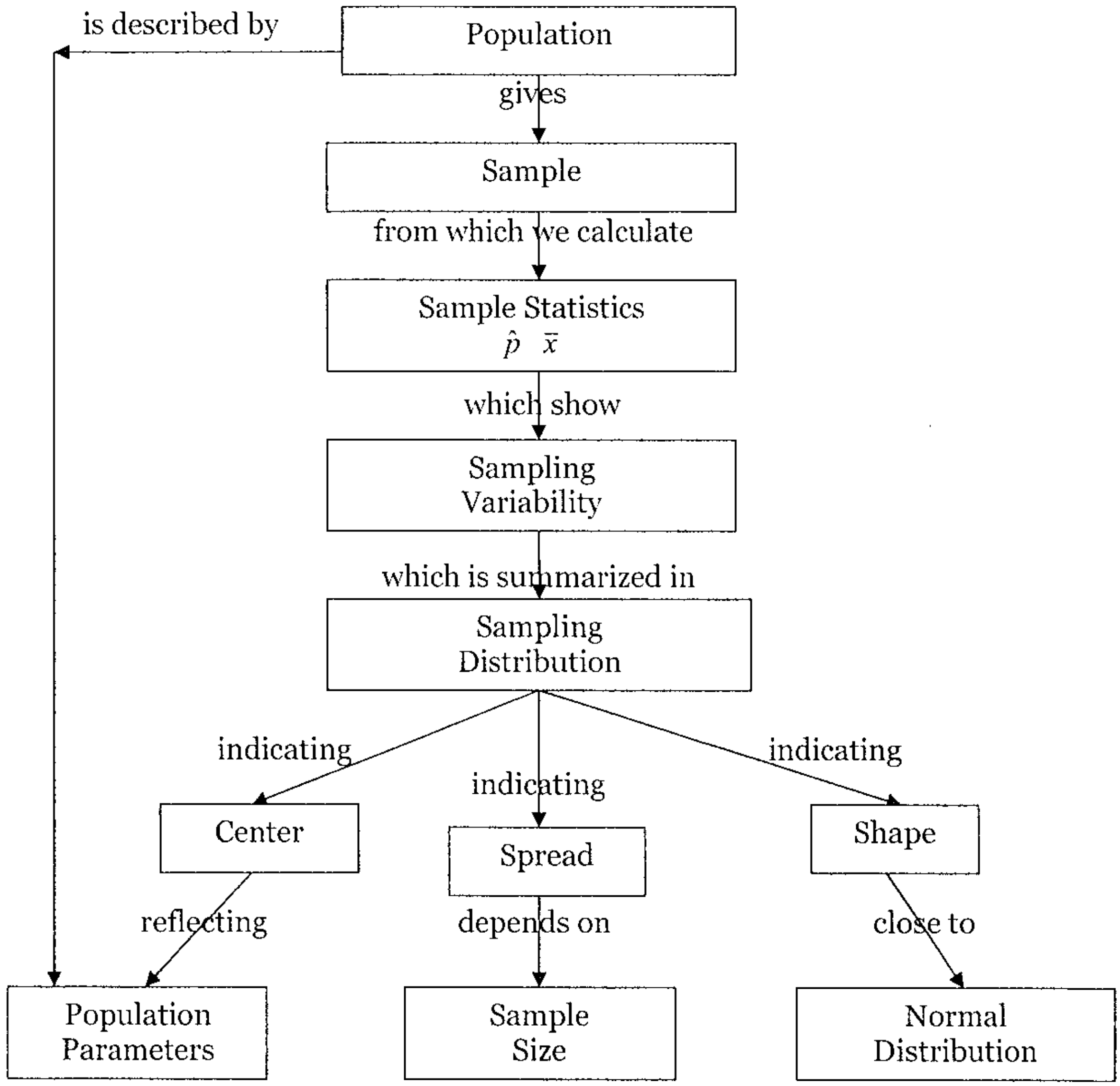
   $P(\bar{x} > 320)$ = normalcdf(320, 1e99, 300, $\frac{50}{\sqrt{25}}$)

   = .0228 unlikely to be over weight

   so yes it can carry entire load

25P

Concept Map for Sampling Distributions



Fill in each blank below:

________________ are selected from populations.

Populations are described by _________________, which are constant in value.

Samples are described by ____________________, which are variable quantities.

The distribution of a sample statistic is known as a __________________________.

The sampling distribution of the sample statistic is characterized by _______________, _________________________ and ___________________________.

The spread of the sampling distribution is related to ____________________________.

The sampling distribution is centered at ______________________________________.

26

Chapter 9 Practice Multiple Choice

1. A phone-in poll conducted by a newspaper reported that 73% of those who called in likely business tycoon Donald Trump. The unknown true percentage of American citizens who like Donald Trump is a
   C
   a. Statistic
   b. Sample
   c. Parameter
   d. Population
   e. Sampling Distribution

2. A survey conducted by Black Flag asked whether or not the action of a certain type of roach disk would be effective in killing roaches. 79% of the respondents agreed that the roach disk would be effective. The number 79% is a
   C
   a. Parameter.
   b. Population.
   c. Statistic.
   d. Sample.
   e. Sampling Distribution.

3. The distribution of values taken by a statistic in all possible samples of the same size from the same population is the
   D
   a. Probability that the statistic is obtained.
   b. Population parameter.
   c. Variance of the values.
   d. Sampling distribution of the statistic.
   e. Same shape as the population distribution.

4. The sampling distribution of a statistic is
   C
   a. The probability that we obtain the statistic in repeated random samples.
   b. The mechanism that determines whether or not randomization was effective.
   c. The distribution of values taken by a statistic in all possible samples of the same size from the same population.
   d. The extent to which the sample results differ systematically from the truth.
   e. Approximately normal.

5. I flip a coin ten times and record the proportion of heads I obtain. I then repeat this process of flipping the coin ten times and recording the proportion of heads obtained by many, many times. When done, I make a histogram of my results. This histogram represents
   D
   a. The bias, if any, that is present.
   b. The true population parameter.
   c. Simple random sampling.
   d. The sampling distribution of the proportion of heads in ten flips of the coin.
   e. The population distribution.

6. A statistic is said to be unbiased if
   B
   a. The survey used to obtain the statistic was designed so as to avoid even the hint of racial or sexual prejudice.
   b. The mean of its sampling distribution is equal to the true value of the parameter being estimated.
   c. Both the person who calculated the statistic and the subjects whose responses make up the statistic were truthful.

27

d. It is used for only honest purposes.
e. It shows little variability.

7. The number of undergraduates at Johns Hopkins University is approximately 2000, while the number at Ohio State University is approximately 40,000. At both schools a simple random sample of about 3% of the undergraduates is taken. We conclude that
   a. The sample from Johns Hopkins has less sampling variability than that from Ohio State.
   b. The sample from Johns Hopkins has more sampling variability than that from Ohio State.
   c. The sample from Johns Hopkins has almost the same sampling variability as that from Ohio State.
   d. The sample from Johns Hopkins has exactly the same sampling variability as that from Ohio State.
   e. It is impossible to make any statements about the sampling variability of the two samples because the students surveyed were different.

Use the following to answer questions 8-10:

A survey asks a random sample of 1500 adults in Ohio if they support an increase in the state sales tax from 5% to 65, with the additional revenue going to education. Let *p* denote the proportion in the sample that say they support the increase. Suppose that 40% of *all* adults in Ohio support the increase.

8. The mean μ_p of *p* is
   a. 5%
   b. 0.06
   c. 40% $\pm$ 5%
   d. 0.40
   e. 600

9. The standard deviation σ_p of *p* is
   a. 0.000327
   b. 0.00613
   c. 0.0126
   d. 0.00016
   e. 0.0056

10. How large a sample would be needed to guarantee that the standard deviation σ_p of *p* is less than 0.01?
   a. 100
   b. 564
   c. 1500
   d. ~~2400~~
   e. 2401

E

11. Suppose you are going to roll a die 60 times and record *p*, the proportion of times that a 1 or a 2 is showing. The sampling distribution of *p* should be centered about
   a. 1/6
   b. 1/3
   c. 1/2
   d. 20
   e. 30

12. The law of large numbers states that, as the number of observations drawn at random from a population with finite mean μ increases, the mean J of the observed values
    a. Gets larger and larger.
    b. Gets smaller and smaller.
    c. Gets closer and closer to the population mean μ.
    d. Fluctuates steadily between one standard deviation above and one standard deviation below the mean.
    e. Varies randomly about μ.

Use the following to answer question 13:

A factory produces plate glass with a mean thickness of 4 mm and a standard deviation of 1.1 mm. A simple random sample of 100 sheets of glass is to be measured, and the sample mean thickness of the 100 sheets J is to be computed.

13. We knew the random variable J has approximately a normal distribution because of the
    a. Law of large numbers
    b. Central Limit Theorem
    c. Law of Proportions
    d. Fact that probability is the long run proportion of times an event occurs.
    e. Normality of the population distribution.

14. The incomes in a certain large population of college teachers have a normal distribution with mean \$35,000 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 \$40,000?
    a. 0.0228
    b. 0.1587
    c. 0.8413
    d. 0.9772
    e. Essentially 0

15. The Central Limit Theorem says that when a simple sample of size n is drawn from any population with mean μ and standard deviation σ, then when n is sufficiently large, the
    a. Standard deviation of the sample mean is $\frac{\sigma^2}{n}$.
    b. Distribution of the population is exactly normal.
    c. Distribution of the sample mean is approximately normal.
    d. Distribution of the sample mean is exactly normal.
    e. Mean of the sampling distribution of J is μ.

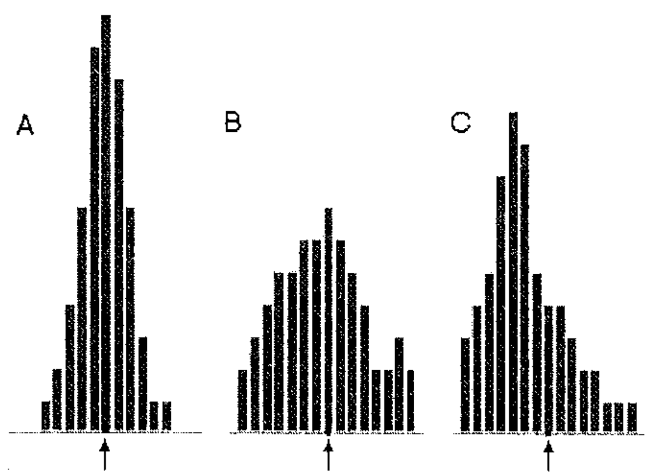
In items 1–3, classify each underlined number as a parameter or statistic. Give the appropriate notation for each.

1. Forty-two percent of today's 15-year-old girls will get pregnant in their teens.

2. A 1993 survey conducted by the *Richmond Times-Dispatch* one week before Election Day asked voters which candidate for the state's Attorney General they would vote for. Thirty-seven percent of the respondents said they would vote for the Democratic candidate. On Election Day, 41% actually voted for the Democratic candidate.

3. The National Center for Health Statistics reports that the mean systolic blood pressure for males 35 to 44 years of age is 128 and the standard deviation is 15. The medical director of a large company looks at the medical records of 72 executives in this age group and finds that the mean systolic blood pressure for these executives is 126.07.

Below are histograms of the values taken by three sample statistics in several hundred samples from the same population. The true value of the population parameter is marked on each histogram.



4. Which statistic has the largest bias among these three? Justify your answer.

5. Which statistic has the lowest variability among these three?

6. Based on the performance of the three statistics in many samples, which is preferred as an estimate of the parameter? Why?

AP Statistics   Name Key
Sampling Distribution Questions

1. Which of the following best describes a sampling distribution of a statistic?
   a) It is the probability that the sample statistic equals the parameter of interest.
   b) It is the probability distribution of all the values that are contained in all possible samples of the same size.
   c) It is the distribution of all of the statistics calculated from all possible samples of the same size.
   d) It is the histogram of sample statistics from all possible samples of the same size.
   e) It is none of these.

C

2. A simple random sample of 50 adults were asked to reveal their gross annual incomes. The variance of this sample:
   a) is always smaller than the variance of the population.
   b) ~~cannot be computed since the population size is not given.~~
   c) equals the variance of the population.
   d) is an estimate of the variance in the sampling distribution of the means of the gross annual incomes of all possible samples.
   e) is an estimate of the variance of the population but may differ from the variance of the population.

D

3. A simple random sample of 100 high school seniors in a certain suburb reveals that 65% of them have at least part-time jobs in addition to school. If the expected value of this proportion is equal to the proportion of high school seniors who have at least part-time jobs for the entire suburb, then we say that the sample proportion is
   a) a true value.
   b) an unbiased estimator of the population proportion.
   c) equal to the population proportion.
   d) an estimate whose variance equals the variance of data in the population.
   e) less than the population proportion since only 100 students were sampled.

B

$\mu_{\hat{p}} = p$
sample = pp

Use the following information for questions 4 and 5. Suppose you roll a die 10 times and record the proportion of sixes. Suppose you then conduct a simulation of this experiment, first 100 times, then 1,000 times, and draw one histogram of the proportion of sixes found after 100 simulations and a second histogram of the proportions of sixes found after 1,000 simulations.

$p = 1/6$

4. Which of the following is true regarding the mean of the proportions of sixes from each simulation?
   a) The mean of the proportion of sixes for the 100 simulations will equal the mean of the proportion of sixes for the 1,000 simulations.
   b) The mean of the proportion of sixes for the 1,000 simulations will be a better estimator of the theoretical probability of rolling a six than the mean of the proportion of sixes for the 100 simulations.
   c) ~~The mean of the proportion of sixes for the 100 simulations will be less than the mean of the proportion of sixes for the 1,000 simulations.~~
   d) ~~The mean of the proportions of sixes for both simulations will not estimate the theoretical probability of rolling a six since they are finite samples from an infinite population.~~
   e) ~~None of these is true.~~

31

5. Which of the following statements are true regarding the histograms of the results from the two simulations?
   ~~I.~~ The histogram from both simulations will be skewed left since a fair die does not exist in nature.
   II. The histograms from both simulations will be mound-shaped and symmetric.
   III. The histogram from the experiment that has 1,000 simulations will tend to be more mound-shaped and symmetric than the histogram from the experiment that has 100 simulations.
   a) I only
   b) I and II
   c) ~~II and III~~
   d) III only
   e) None of these statements is true.

3/11

6. Two simple random samples of 50 undergraduates each from two universities are taken to determine the proportion of students who approve of the food service at their respective schools. The first university has an enrollment of 5,000 undergraduates while the second university has an enrollment of 35,000 undergraduates. Which of the following is the most accurate statement regarding these samples?
   a) The variability of the sample from the larger university will be greater than the variability of the sample from the smaller university.
   b) The proportion of students who approve of the food service will be the same since the sample sizes are the same.
   c) The enrollment figures from the two universities are not relevant to whether the sample statistics obtained are unbiased estimates of the parameters of the two populations.
   d) If a university with 100,000 undergraduates conducted a simple random sample of 50 of its students, the results would be less accurate than either sample referenced above.
   e) None of these is an accurate statement.

FREE-RESPONSE QUESTION

Consider the following experiment: Draw a single card out of a box containing 3 red cards and 2 blue cards 10 times with replacement. Calculate the proportion of red cards that are drawn out of the 10 trials. Describe what you would expect regarding center, variability, and the histogram if the experiment is repeated:
   a) 20 times.
   b) 100 times.
   c) For all possible combinations.

6/10

more variability on both side
center would be around $\frac{3}{5}$ (.6)
center would be .6
variability would be between 0 ~ 1
histogram mounded single peaked

($\frac{3}{5}$)

$p = \frac{3}{5}$

$(\frac{3}{5})(10)$

$\hat{p} = \frac{6}{10}$

32

Name: ______________________________

AP Statistics Unit 6 Review

Be sure to check the rules of thumb before using the formula for the standard deviation or before using a normal approximation to calculate a probability!

1. For each of the following statements, identify the number in boldface type as a parameter or a statistic. Give the appropriate notation for each.
   (a) A department store reports that **84%** of all customers who use the store's credit plan pay their bills on time.
   $p = .84$ parameter
   (b) A sample of 100 students at a large university had a mean age of **24.1** years. $\bar{x}_{10} = 24.1$
   (c) The Department of Motor Vehicles reports that **22%** of all vehicles registered in NC are imports.
   $p = .22$ parameter pop is All NC cars
   (d) A hospital reports that based on the ten most recent cases, the mean length of stay for surgical patients is **6.4** days.
   $\bar{x}_{10} = 6.4$

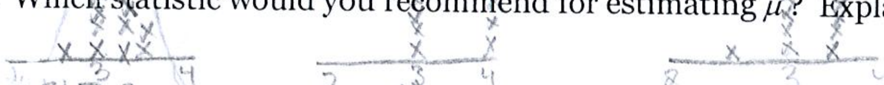
2. Consider the following population: {2, 3, 3, 4, 4}. The value of μ is 3.2, but suppose that this is not known to an investigator, who therefore wants to estimate μ from sample data. Three possible statistics for estimating μ are:

   Statistic #1: the sample mean, $\bar{x}$
   Statistic #2: the sample median, M
   Statistic #3: the average of the largest and the smallest values in the sample.

   A random sample of size 3 will be selected without replacement. Provided that we disregard the order in which observations are selected, there are ten possible samples that might result (writing 3 and 3*, 4 and 4* to distinguish the two 3's and the two 4's in the population):

2.67, 3, 2.5	3, 3, 3	3, 3, 3	3, 3, 3	3, 3, 3
2, 3, 3*	2, 3, 4	2, 3, 4*	2, 3*, 4	2, 3*, 4*
2, 4, 4*	3, 3*, 4	3, 3*, 4*	3, 4, 4*	3*, 4, 4*
3.3, 4, 3	3.3, 3, 3.5	3.3, 3, 3.5	3.67, 4, 3.5	3.67, 4, 3.5

   For each of these ten samples, compute statistics 1, 2, and 3. Construct the sampling distribution of each of these statistics. Which statistic would you recommend for estimating μ? Explain the reasons for your choice.



   mean $\bar{x}$, spread 7 from 2.7–3.3

3. Let x denote the time in minutes that it takes a fifth-grade student to read a certain passage. Suppose that the mean value and standard deviation of x are $\mu = 2$ min and $\sigma = 0.8$ min.
   (a) If $\bar{x}$ is the sample average time for a random sample of 9 students, where is the $\bar{x}$ distribution centered? $\mu_{\bar{x}} = \mu = 2$
   (b) How much does the sampling distribution spread out about the center (as described by the standard deviation)? $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{.8}{\sqrt{9}} = \frac{.8}{3} = .267$

   $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

   (c) Find the mean and standard deviation of the sampling distributions of sample sizes $n = 20$ and $n = 100$. How do the centers and spreads of the three distributions compare to one another? $\mu_{\bar{x}_{20}} = \mu_{\bar{x}_{100}} = 2$ $\sigma_{\bar{x}_{20}} = \frac{.8}{\sqrt{20}} = .1789$ $\sigma_{\bar{x}_{100}} = \frac{.8}{\sqrt{100}} = .08$
   (d) Which sample size would be most likely to result in a $\bar{x}$ value close to μ, and why?
   n = 100, less variability among the spread

   The larger the sample size the smaller the spread
   $\mu_{\bar{x}} = \mu$ regardless of sample size because $\bar{x}$ is an unbiased estimator

33

4. An airplane with room for 100 passengers has a total baggage limit of 6000 pounds. Suppose that the total weight of the baggage checked by an individual passenger is a random variable x with mean value 50 lb and standard deviation 20 lb. If 100 passengers board a flight, what is the approximate probability that the total weight of their baggage will exceed the limit? (Hint: With $n = 100$, the total weight exceeds the limit when the average weight exceeds 6000/100.)

5. A manufacturing process is designed to produce bolts with a .5 in diameter. Once each day, a random sample of 36 bolts is selected and the diameters recorded. If the resulting sample mean is less than .49 in or greater than .51 in, the process is shut down for adjustment. The standard deviation for diameter is .02 in. What is the probability that the manufacturing line will be shut down unnecessarily? (Hint: Find the probability of observing an $\bar{x}$ in the shut-down range when the true process mean really is .5 in.)

6. Suppose that a particular candidate for public office is in fact favored by 48% of all registered voters in the district. A polling organization takes a random sample of 500 voters and will use $\hat{p}$, the sample proportion, to estimate p, the population proportion. What is the approximate probability that $\hat{p}$ will be greater than .5, causing the polling organization to incorrectly predict the result of the upcoming election?

7. A certain chromosome defect occurs in only one out of 200 Caucasian adult males. A random sample of n = 100 adult Caucasian males is obtained.

(a) What are the mean and standard deviation of the sampling distribution of $\hat{p}$?

(b) Does the sampling distribution of $\hat{p}$ have an approximately normal distribution? Explain.

(c) What is the smallest value of n for which the sampling distribution of $\hat{p}$ is approximately normal?

8. A manufacturer of computer printers purchases plastic ink cartridges from a vendor. When a large shipment is received, a random sample of 200 cartridges is selected, and each is inspected. If the sample proportion of defectives is more than .02, the entire shipment will be returned to the vendor.

(a) What is the approximate probability that the shipment will be returned if the true proportion of defectives in the shipment is .05?

(b) What is the approximate probability that the shipment will *not* be returned if the true proportion of defectives in the shipment is .10?