Name: ____________________________________ Date: __________________

SAMPLE MEANS
COMMON CORE ALGEBRA II

The vast majority of the statistics that you've done so far has been **descriptive**. With descriptive statistics, we summarize how a data set "looks" with measures of central tendency, like the mean, and measures of dispersion, like the standard deviation. But, the more powerful branch of statistics is known as **inferential** where we try to **infer** properties about a population from samples that we take. We do this by using **probability** and **sampling variability** to estimate how likely the sample is given a certain population.

We begin our multi-lesson investigation into **inferential statistics** with the most basic question. How can we estimate the **population mean**, μ, if we know a sample mean, $\bar{x}$? Before answering this question, though, we need to investigate the distribution of sample means.

***Exercise* #1:** Say we are investigating the heights of 16 year old American males. Say we know that the population mean height is 65.3 inches with a standard deviation of 4.2 inches. Let's say we take a sample of 16 year old American males. The sample has a size of 30.

(a) Will the mean height of the sample always be 65.3? Why or why not? Could it be significantly different?

(b) Will the standard deviation of the sample be 4.2 inches? Would you expect more or less variation in a sample versus a population?

(c) Run the program NORMSAMP with a mean and standard deviation given above and a sample size of 30. Do at least 200 simulations (500 is preferable). State the minimum and maximum of the sample means, the mean of the sample means, and the standard deviation of the sample means.

min sample mean = _____________ Mean of means: Standard deviation of means:

max sample mean = _____________

(d) How does the **variability** of the sample means compare to the **variability** of the population? Is it more or less?

(e) State the mean of the sample standard deviations. How does it compare to the population standard deviation? Could you use the standard deviation of a sample to estimate the standard deviation of the population?

We can use our simulation to decide whether a sample could have come from a given population. We can even quantify how likely it would be to happen. This is known as establishing **confidence**.

***Exercise* #2:** Mr. Weiler took a sample of 30 16-year old males and found the mean height of the sample to be 66.4 inches. Do you believe this sample came from this population? Why or why not? Examine the results of your simulation. Quantify how likely this sample (or one greater) was to come from the population simulated.

Strangely enough, this process can be used in order to give a **confidence interval** for the population mean if we know the sample mean. This is important because in reality **the population mean is almost never known and is what we want to infer from the sample** mean. The next set of exercises will illustrate how this is done using simulation.

***Exercise* #3:** A sample of 50 ripe oranges were taken from a large orchard in order to estimate the mean weight of a ripe orange. The sample mean was 212 grams and the sample standard deviation was 34 grams.

(a) Why does it seem reasonable to use the sample standard deviation as an estimate of the population standard deviation? See Exercise #1(e).

(b) Run a simulation using the 212 as the population mean (even though it is the sample mean) and use 34 as the population standard deviation. State the 5^{th} percentile sample mean and the 95^{th} percentile sample mean.

5^{th} Percentile Sample Mean = _____________

95^{th} Percentile Sample Mean = ____________

(c) Now, try your simulation again, but use the 5^{th} percentile sample mean as the population mean. Where does the 212 lie on the distribution in terms of percentile? Notice how close this is to the 95^{th} percentile.

(d) Now, try your simulation again, but use the 95^{th} percentile sample mean as the population mean. Where does the 212 lie on the distribution in terms of percentile? Notice how close this is to the 5^{th} percentile.

(e) What both (c) and (d) tell us is that by using the 5^{th} and 95^{th} percentile values based on our original sample mean, we have actually found the **lowest possible population mean** and **highest possible population mean** that could have resulted in that sample mean 90% of the time. Write the 90% confidence interval below for μ based on (b).

Name: ____________________________________ Date: __________________

SAMPLE MEANS
COMMON CORE ALGEBRA II HOMEWORK

FLUENCY

1. The mean of the sample means is

(1) Greater than the population mean.
(2) Less than the population mean.
(3) Equal to the population mean.
(4) Could be greater or less than the population mean. _________

2. The variation within the sample means is

(1) Less than the variation within the population.
(2) More than the variation within the population.
(3) Equal to the variation within the population. _________
(4) Could be more or less than the variation within the population.

APPLICATIONS

3. A factory has machines that fill 12 ounce soda bottles repeatedly with an average volume of 12.2 ounces and standard deviation of 0.9 ounces. A new machine was installed and 30 bottles were sampled. It was found that they had an average volume of 11.8 ounces. We want to investigate whether this mean is significantly lower than the original population mean.

(a) Run NORMSAMP with a mean of 12.2 and a standard deviation of 0.9. Run 100 simulations. What percentile rank would you give the 11.8 ounces (this will vary based on the simulation)?

(b) Based on your findings from (a), can you conclude that this sample mean likely came from the same population or a different population with a lower mean? Explain.

(c) Use the sample mean of 11.8 ounces and a standard deviation of 0.9 ounces to generate the 90% confidence interval for the population mean of the new soda filling machine by simulation. Use a sample size of 30 and at least 100 samples to generate your interval. Round you lower and upper estimate for μ to the nearest hundredth.

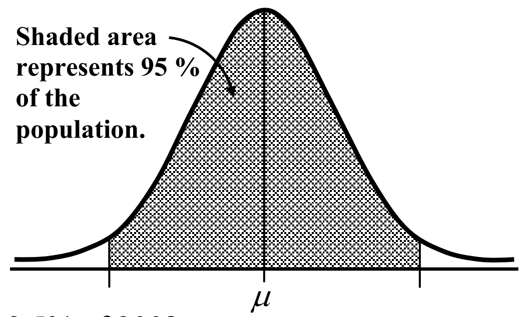
$\mu_L =$ ____________________

Interval: __________ $\le \mu \le$ __________

$\mu_H =$ ____________________

Let's work some now with the **95% confidence interval**. It will be easiest to use 200 simulations to generate this interval as we will soon see.

4. Consider a normal distribution with 95% of the probability (or distributions) **centered about the population mean** μ.



(a) What percent lies in half of the shaded area shown in the diagram?

(b) Explain why 2.5% of the population must lie in each of the un-shaded areas of the graph.

(c) What is 2.5% of 200?

5. Researchers have found that the average number of hours per week spent by adults watching television is 34.2 with a standard deviation of 6.8 hours. Researchers wanted to determine if there was an effect to sampling people with tablets. They found that a random sample of 40 people with tablets had a sample mean of 36.3 hours per week with a sample standard deviation of 5.4 hours.

(a) Run NORMSAMP with a population mean of 34.2 hours and a standard deviation of 6.8 hours. Do 200 simulations. This will take some time (approximately 8 minutes). Based on your results, what is the approximate percentile rank of 36.3 (remember it is out of 200 now)?

(b) Do you have significant evidence that the 36.3 comes from a population with a mean that is higher than 34.2? Explain your thinking.

(c) Now, let's attempt to construct the **95% confidence interval** for the sample whose mean was 36.3. Run a simulation with a mean of 36.3, a standard deviation of 5.4, a sample size of 40, and with 200 simulations. Find the 2.5th percentile as the lower limit and the 97.5th percentile as the upper limit.

$\mu_L =$ ____________________

Interval: __________ $\le \mu \le$ __________

$\mu_H =$ ____________________

(d) The **theoretical** (versus simulated) 95% confidence interval can be found using the formula below, where $\bar{x}$ is the observed sample mean and s is the sample standard deviation. Use this formula and compare to the interval from above.

$$\bar{x} - 2 \cdot \frac{s}{\sqrt{n}} \le \mu \le \bar{x} + 2 \cdot \frac{s}{\sqrt{n}}$$

Interval: __________ $\le \mu \le$ __________