Math 130 – Jeff Stratton Name _____

The Sampling Distribution of the Mean
Confidence Intervals for a Proportion

Goal: To gain experience with the sampling distribution of the mean, and with confidence intervals for a proportion.

Part 1 – The Sampling Distribution of the Mean

How does $\bar{y}$ behave? The surprising thing is that it behaves normally with the same center as your population, but a smaller variance.

For sample means, we will learn about the sampling distribution via an applet (link online). Steer your (Java-enabled) browsers to
http://onlinestatbook.com/stat_sim/sampling_dist/index.html

In this applet, when you first hit Begin, a histogram of a normal distribution is displayed at the top of the screen. This is the parent population from which samples are taken (think of it as the bin of balls) except it's showing the distribution. The mean of that distribution is indicated by a small blue line and the median is indicated by a small purple line. Since the mean and median are the same for a normal distribution, the two lines overlap. The red line extends from the mean one standard deviation in each direction.

The second histogram displays the sample data. This histogram is initially blank. The third and fourth histograms show the distribution of statistics computed from the sample data. The option N in those histograms is the sample size you are drawing from the population. We will be exploring the distribution of the sample mean by drawing many samples from the parent distribution and examining the distribution of the sample means we get.

Step 1. Describe the parent population. What distribution is it and what is its mean and standard deviation?

Step 2. You can see the third histogram is already set to "Mean", with a sample size of N = 5. Click Animated sample once. The animation shows five observations being drawn from the parent distribution. Their mean is computed and dropped down onto the third histogram. For your sample, what was the sample mean?

Step 3. Click Animated sample again. A new set of five observations are drawn, their mean is computed and dropped as the second sample mean onto the third histogram. What did the mean of the sample means (yes, we are interested in the mean of sample means as part of the sampling distribution) change to?

Step 4. Click Animated sample one more time. What did the mean of the sample means update to now?

Step 5. Click 10,000. This takes 10,000 samples at once (no more animation) and will place those 10,000 sample means on the third histogram and update the mean and standard deviation of the sample means. Record the mean and standard deviation of the sample means. What shape does this third histogram have? How do these findings compare to the parent distribution?

Step 6. Hit Clear Lower 3 in the upper right corner. Change N = 5 to N = 25 for the third histogram. Do animated sample at least once (convince yourself it is actually samples of 25 now). Then take 10,000 at once. Record the mean and standard deviation of the sample means. What shape does the third histogram have? How do these findings compare to the parent distribution?

Step 7. Compare the different standard deviations from Steps 5 and 6. What effect does sample size appear to have on standard deviation of the sample means?

Step 8. Hit Clear Lower 3. Change the parent distribution to Skewed. What are the new mean and standard deviation of the parent distribution? Which direction is this distribution skewed?

Step 9. Set N = 5 back for the third histogram. Set "Mean" and N = 25 for the fourth histogram. Hit 10,000 at once. (This will take 10,000 samples of size 5, compute the sample means and put those means in the third histogram, as well as take 10,000 samples of size 25, compute the sample means and put those means in the fourth histogram). What do the distributions look like for the third and fourth histograms? Are they skewed like the parent population? What are the means and standard deviations for each histogram?

Step 10. Hit Clear Lower 3. Change the parent distribution to Custom. Draw in a custom distribution (left click and drag the mouse over the top histogram). Sketch your custom distribution below. What are its mean and standard deviation?

Step 11. Hit 10,000 at once (leave the settings on the third and fourth histograms alone). (You could take animated once to convince yourself it was really drawing from your new distribution). What do the third and fourth histograms look like? Anything like the parent distribution? What are their means and standard deviations?

The Sampling Distribution for the sample mean, $\bar{y}$ can be described as having a mean $\mu_{\bar{y}} = \mu$, the same as that of the population mean. The standard deviation is $\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$.
The distribution is exactly normal if the parent population is normal. Finally, the Central Limit Theorem tells us the distribution will be approximately normal with the mean and standard deviation stated above if n is sufficiently large even if the population distribution is not normal.

Part 2 – CIs for Proportions – The Ball Bin

For the purposes of this example, the bin filled with balls represents the population of all possible birds that could be captured as part of an upcoming study looking for a genetic trait which is known to be harmful to carriers and sometimes fatal to those which exhibit the trait (like sickle cell anemia idea but for birds). Let white balls denote birds that do not have the trait and are also not carriers. Let red balls denote birds that are carriers but do not themselves exhibit the trait, and let green balls denote birds that do exhibit the trait (also then carriers).

Looking at the bin, what are your initial guesses as to the composition of this population? What percentage of birds are carriers, but do not exhibit the trait (red).

Q1] Individually, I'd like each of you to take a sample of size 25 and size 50, and record the sample proportion of "red" birds. For these two samples, compute the 90%, 95%, and 99% confidence intervals for *p*, the true proportion of disease-carrying birds who don't exhibit the trait.

Recall the confidence interval formula: $\hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$

90% CI:

95% CI:

99% CI:

Report your two $\hat{p}$ values to me. We'll enter and plot the intervals.

Q2] Are the intervals the same?

Q3] We are lucky to know the true population proportion in this case, $p = 0.33$. How did your intervals perform?

Q4] We consider an interval to be "good" if it contains the true population proportion. How many of the 90% confidence intervals made by the class would you expect to be good? What about the 95% and 99% confidence intervals?

Part 3 – Checking Understanding

A report gives a 99 percent confidence interval for the proportion of patients who suffer minor side effects from a certain drug as (.22,.32). You may assume the CI was based on a random sample of patients.

Q5] What was the sample proportion of patients who suffered minor side effects on the drug?

Q6] What was the population proportion of patients who suffered minor side effects on the drug?

Q7] (T/F)

P(population proportion of patients who suffered minor side effects is in (.22,.32))=.99

Q8] (T/F)

P(sample proportion of patients who suffered minor side effects is in (.22,.32))=.99

Q9] What does the 99 percent confidence level mean?

Q10] Could you find the sample size this interval was based on?

Q11] The report states that more than a quarter of patients should expect to suffer minor side effects from the drug based on the CI. Do you agree or disagree? Why?

Q12] The standard drug for treating this condition has a minor side effect proportion of .5 (roughly 50 percent of patients suffered minor side effects). Would you encourage adopting this new drug over the standard drug if they only differed in minor side effect proportions? Explain.

Q13] Another drug for treating the same condition has a similar confidence interval of (.27,.37). Would you be able to conclude that you should use the first drug because it has a lower rate of minor side effects assuming the drugs are equal in all other respects?

Part 4 – Beetles

In Michigan and surrounding states, there has been substantial attention in recent years on the emerald ash borer, an invasive beetle that destroys ash trees. In some states, like Wisconsin and Michigan, ash trees make up a sizable percentage (20%) of "urban" forest, so the death of these trees and the spread of the beetle must be dealt with. The beetle has now been found in numerous other states and a quarantine region for moving ash wood has expanded to three states (it was just southeastern Michigan at one point). You have been asked by one of the afflicted states to determine the proportion of infected ash trees (assume you can identify the infected trees). A random sample of 500 ash trees in the state leads you to conclude that 357 are infected.

Q14] Argue whether the conditions necessary to do a confidence interval are satisfied here.

Q15] What is the sample proportion of infected ash trees?

Q16] If the population proportion was really 70 percent, what is the probability of observing the sample proportion you observed or something higher? (Hint, I'm thinking of normal theory here).

Q17] Provide the state with a 95 percent confidence interval for the proportion of infected ash trees. How would a 99% CI compare to your 95% CI? How about a 90% CI?

Q18] Based on your 95% CI, if the state asked you if it was reasonable to conclude that 80 percent of ash trees were infected, what would you say?

Q19] Based on your 95% CI, if the state asked you if it was reasonable to conclude that 70 percent of ash trees were infected, what would you say?

Q20] Another state got a similar 95% CI from .66-.72 for their proportion of infected ash trees. What was the value of their sample proportion that led to that CI? (Think about the formula.) Think about and explain (but don't solve) how you would work backwards to find the sample size they used to estimate *p*.

Can you conclude that the two states have different proportions of infected ash trees based on the 95% CIs? Why or why not? (We will learn how to deal with two proportion problems directly in about a week.)