Math 17, Section 2 Name _____SOLUTIONS_______
Spring 2011

Lab 6

Goal: To gain experience with confidence intervals for a proportion

Part 1 – 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: Answers will vary.

95% CI: Answers will vary.

99% CI: Answers will vary.

Report your computed confidence intervals to me. We'll enter and plot the intervals.

Q2] Are the intervals the same? The intervals will probably differ, but won't if you get two samples with the same number of red balls.

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

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?
We expect 90% of the class' 90% confidence intervals to contain 0.33. The answer is similar for 95% and 99% CIs.

Part 2 – 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?
The $\hat{p}$ is the middle of the interval, or 0.27.

Q6] What was the population proportion of patients who suffered minor side effects on the drug? We don't know! That's what we're trying to find out!

Q7] (T/F)

P(population proportion of patients who suffered minor side effects is in (.22,.32))=.99
False: We don't know the population proportion. The probability that it's inside the interval is either 0 or 1.

Q8] (T/F)

P(sample proportion of patients who suffered minor side effects is in (.22,.32))=.99
False: The interval is calculated using the sample proportion, so the probability that it is inside the interval is actually 100%.

Q9] What does the 99 percent confidence level mean?
It means that if we were to repeatedly sample and compute the confidence interval each time, then we'd expect 99% of the intervals to include the true mean.

Q10] Could you find the sample size this interval was based on?
Yes we could. The margin of error in this CI is 0.05. We can find

$$ME = Z^*\sqrt{\frac{\hat{p}\hat{q}}{n}} \Rightarrow 0.05 = 2.576\sqrt{\frac{(0.27)(0.73)}{n}} \Rightarrow n = \frac{2.576^2(0.27)(0.73)}{0.05^2} = 523.1$$

A sample of about 524 was used.

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?
I disagree. Based on the CI, the percentage of patients suffering minor side effects could be as low as 22%

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.

Yes, I would encourage adopting this new drug. We are 99% confident that the true proportion of patients suffering side effects is between 22% and 32%, well below 50%.

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?

Here, we couldn't really say they are different. The two intervals overlap, so there is not really evidence of a difference.

Part 3 – 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.

There are four conditions we need to check:

- The sample trees may not be independent. Beetle infestation likely affects a cluster of trees. This assumption may be questionable.
- We have a random sample.
- There are probably more than 5,000 trees in the forest, so we haven't sampled more than 10% of the population.
- We have $\hat{p} = \frac{357}{500} = 0.714$. We have $n\hat{p} = 500(0.714) = 357 \geq 10$ and $n\hat{p} = 500(0.286) = 143 \geq 10$. Our sample is large enough.

I'd say that three of the four conditions are satisfied. Proceeding with the confidence interval may be incorrect here.

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

The sample proportion of infected ash trees is $\hat{p} = \frac{357}{500} = 0.714$.

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).

Assuming $p = 0.70$, the sampling distribution of $\hat{p}$ is normal with a mean of 0.70 and a standard deviation of $\sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.70)(0.30)}{500}} = 0.0205$. Using the Z-score, $Z = \frac{0.714-0.70}{0.0205} = 0.683$.

Looking at the Z-table, $P(Z \geq 0.683) = 1 - 0.7517 = 0.2483$

There is a 24.83% probability of observing a sample proportion of 71.4%.

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?

$$\hat{p} \pm Z^{*}\sqrt{\frac{\hat{p}\hat{q}}{n}} = 0.714 \pm 1.96\sqrt{\frac{(0.714)(0.286)}{500}} = 0.714 \pm 0.0202 = (0.694\,, 0.734)$$

We are 95% confident that the true proportion of beetle infested trees is between 69.4 and 73.4 percent. A 99% confidence interval would be wider than this one, and a 90% confidence interval would be narrower than this one.

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?

It would not be reasonable to conclude that 80% of trees were infected. Our entire confidence interval lies below that.

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?

It would be reasonable to say that 70% of trees are infected, because 70 lies within our confidence interval, the range of plausible values.

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.)

We cannot conclude that the state have different proportions of infected ash trees. The confidence intervals overlap.