Math 244 Lecture Notes

Chapter 22 Day One: Comparing Two Proportions Using a Confidence Interval

Overview: Today, we will compare the success rates/proportions for two groups using a confidence interval. This will use many of the techniques from the previous chapters.

We found our confidence interval for one-proportion using the formula

$$\hat{p} \pm Z^* \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

or rather (Center)±(Dist)*SE. Confidence intervals are based off _____

We used the model in step two of a hypothesis test of

$$\hat{p} = N\left(p, \sqrt{\frac{pq}{n}}\right).$$

Hypothesis tests are based off _____.

Both our CI and our H-Test were created from the same model. So, if we have a model for comparing two proportions, we'll be able to test claims and determine the actual difference in two groups.

Example 1. Suppose that $P_{Seattle} - P_{Portland} > 0$ for two winning rates for local roller derby teams. Which city is better?

Example 2. Suppose that $P_{Seattle} - P_{Portland} < 0$ for two winning rates for local roller derby teams. Which city is better?

Example 3. Suppose that $P_{Seattle} - P_{Portland} = 0$ for two winning rates for local roller derby teams. Which city is better?

NOTE:

We need a model for $\hat{p}_1 - \hat{p}_2$. We start with some quick observations...

(a) The model for a single proportion is

$$\hat{p}_1 = N\left(p_1, \sqrt{\frac{p_1 q_1}{n_1}}\right).$$

(b) If you add/subtract two normal models, you will get another normal model.

(c) If X and Y are independent variables, then

$$Var[X - Y] = Var[X] + Var[Y].$$

Our model for $\hat{p}_2 - \hat{p}_1$ is normal based off the first two facts. Let's determine which normal.

Example 4. CENTER: $E[\hat{p}_1] = p_1$ and $E[\hat{p}_2] = p_2$ are the centers for each of our individual models. What is the center for $E[\hat{p}_1 - \hat{p}_2]$?

Example 5. VARIANCE: $Var[\hat{p}_1] = \frac{p_1 q_1}{n_1}$ and $Var[\hat{p}_2] = \frac{p_2 q_2}{n_2}$ are the variances for each of our individual models. What is the variance for the difference, $Var[\hat{p}_1 - \hat{p}_2]$?

Example 6. Based off the variance, what is the standard deviation, $SD[\hat{p}_1 - \hat{p}_2]$?

Our model is

Example 7. Using the model

$$\hat{p}_1 - \hat{p}_2 = N\left(p_1 - p_2, \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}\right)$$

create the confidence interval for $p_1 - p_2$.
HINT: Our interval will be based off samples and look something like (Center)$\pm Z^*$ SE.

Our assumptions are...

(a)

(b)

(c)

(d)

Example 8. Create a 90% CI for $p_1 - p_2$ if $x_1 = 18$, $n_1 = 25$, $x_2 = 22$, and $n_2 = 50$.

Which group appears to have a larger success rate? Why?

Example 9. Mr. Wherry performs a random study of PCC students and finds that out of 200 students, under the age of 20, 120 have played Pokémon Go! In contrast, he finds that out of 100 students, age 20 or older, 40 have played Pokémon Go! Create a 95% for the difference in proportions for Pokémon Go! players for the two age groups. Assume that people were sampled individually to avoid bias.

Example 10. Suppose you got a 95% CI for $P_{Seattle} - P_{Portland}$ of $(-0.04, -0.03)$. What does this tell us?

Example 11. Suppose you got a 95% CI for $P_{Seattle} - P_{Portland}$ of $(-0.04, 0.10)$. What does this tell us?

Example 12. Suppose you got a 95% CI for $P_{Seattle} - P_{Portland}$ of $(-0.04, -0.03)$. What would the 95% for $P_{Portland} - P_{Seattle}$ look like?

Batman Preference: Create a 95% CI to examine the difference in proportions for people that like Batman.

Example 13. What was the best point estimate in the previous problem?

Example 14. What was the MOE in the previous problem?

Checking work with Calculator: In our calculator, we use "2-propZInt" to check our interval. This is found in either [Stat]→[Tests] on the TI-83/84 OR [Stat/List]→[F7:Ints] on the TI-89. Type in your relevant information and you are good to go!

Calc Interval:

Example 15. Mike and his friend El are playing a game of Dungeons and Dragons. Mike thinks El might be cheating by moving the die. In 50 rolls, Mike gets 4 "critical rolls". El gets 14 "critical rolls" in 60 rolls. Determine if El is getting more "critical hits" using a 95% confidence interval.

Example 16. Jenny and Dory are competing to see who can remember more street names. Dory managed to remember 40% of 150 tested street names. Jenny remembered 60% of 150 tested street names. Determine if Dory remembers less street names using a 99% confidence interval.

Example 17. Jack and Diane, two American kids growing up in the heartland, are competing to see how many Slurpee flavors they can correctly identify. Jack identified 10 correct out of 30. Diane identified 13 correct out of 30. Determine if the difference is statistically significant using a 90% confidence interval.

Example 18. Mr. Wherry has two cats: Victor and Fiona. In the 200 selected afternoons, Fiona is at the front door waiting 180 times. In contrast, Victor is at the front door waiting 40 times in a different 50 randomly selected days. Determine if Fiona is more frequently at the front door waiting using a 99% confidence interval.