

Introductory Statistics – Day 11

Confidence Intervals with Simulations

Today's activities are to be completed in a Word Document.

The activity details can be found on these slides, which are available on Moodle.

We will work through parts of the activities together, and you will have some time to finish them.

Submit your work to Moodle at the end of class today.

Activity 1: What happens when we reject H_0 ?

Create a simulation to answer this question: A new cardiac drug is given to patients after open heart surgery. The pharmaceutical company claims that 20% of people that take this drug suffer from headaches. A doctor has been using this drug on many patients and notices that more than 20% of his patients have complained of headaches. He has the hospital's statistician gather a random sample of 50 patients that have used this drug and finds that 40% complain of headaches while on the drug. Does the doctor have evidence that the 20% advertised side effect rate is incorrect?

- Null Hypothesis: $H_0 : p = 0.2$
(where p is the proportion of patients that have headaches)
- Alternative Hypothesis: $H_A : p > 0.2$
- p-Value =
- Decision =

Activity 1, continued. Since we are rejecting H_0 , we know that the proportion of patients with headaches is not 0.2. If the proportion isn't 0.2, then what is it? When we reject the null hypothesis, it will be our responsibility to report a better estimate for the population parameter.

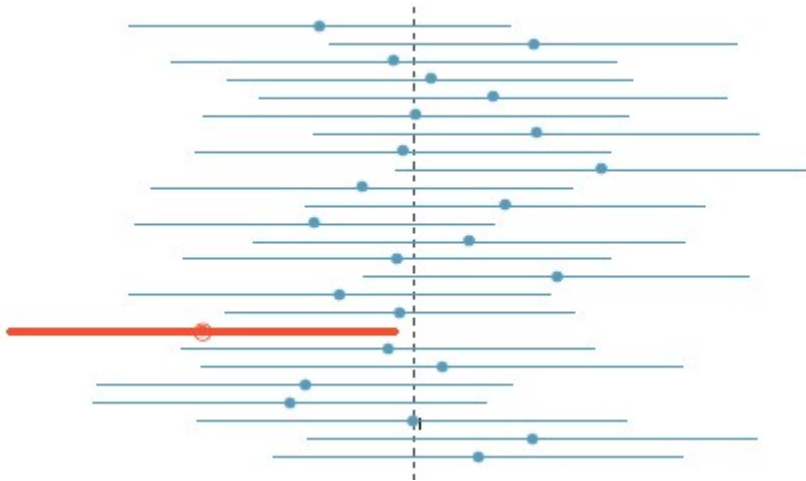
The best estimate we have is our data point ($\hat{p} = 0.4$), so will use this statistic along with a simulator to get a sense for the level of variance in the scenario. This will allow us to create a confidence interval for the population parameter of the proportion of patients with headaches.

- Build a simulation centered around your data point $\hat{p} = 0.4$.
- Use the divider to gather 95% of the simulated results
- Make a statement about where you think the true population proportion could be.

We are 95% confident that the true proportion of people that get headaches on this drug is between _____ and _____.

- Notice that your interval does not contain the pharmaceutical company's claimed 20%. Does that seem appropriate?

Confidence intervals usually capture the population parameter you are estimating (e.g. p or μ). A 95% confidence interval has a 95% chance of capturing the population parameter. As the image shows, that means about 1 in 20 confidence intervals does not capture the population parameter.



Activity 2: What if we simply had no prior knowledge? Or no prior claims to evaluate?

While hiking the Pacific Crest Trail (from Mexico to Canada along the Sierra Nevada and Cascade mountains) Jerry notices that many fellow hikers have adopted a trail name. Many do this to maintain anonymity on the trail and some do it just to embrace the full experience of the trail. Jerry polls 40 randomly selected PCT hikers and finds that 65% of them have adopted trail names (26 out of 40). He knows that it is unlikely that exactly 65% of all Pacific Crest hikers have adopted trail names, but he has no other information.

- 1** Build a simulation to estimate a 95% confidence interval for the proportion of PCT hikers that have adopted a trail name.
- 2** Using the same simulated data, create a 90% confidence interval by capturing the middle 90% of data points.
- 3** Using the same simulated data, create a 80% confidence interval by capturing the middle 80% of data points.
- 4** As the confidence level goes lower, what happens to the width of the interval? Will this always happen when you reduce your confidence level?

Activity 3.

A physical therapy researcher does a baseline study to understand if a new test for an ACL tear is working. She gathers a random sample of 100 patients that she knows have an ACL tear (usually indicated by an MRI scan). Of those 100, the new test gives a positive result on 82 patients. Estimate a 95% confidence interval for the likelihood that this test will give a true positive result for an ACL tear.

Activity 4.

Open the *videoGamesXBoxPS3* data set from Moodle.

This data set is limited to games on the XBox and PS3 platforms.

- 1 Find a 95% confidence interval for the proportion of games rated E for everyone. State your conclusions in a complete sentence related to the context of video games.
- 2 Find a 95% confidence interval for the proportion of games rated M for mature. State your conclusions in a complete sentence related to the context of video games.