Introductory Statistics – Day 18

Confidence Intervals with 1 proportion

To estimate a population mean or proportion based on a sample of data,

- a point estimate (sample mean $\bar{x}$ or sample proportion $\hat{p}$.)
- a likely interval for the population parameter.

> "A plausible range of values for the population parameter is called a **confidence interval.** Using only a point estimate is like fishing in a murky lake with a spear, and using a confidence interval is like fishing with a net. We can throw a spear where we saw a fish, but we will probably miss. On the other hand, if we toss a net in that area, we have a good chance of catching the fish." –OpenIntro Stats, Sect 2.8

To construct a 95% confidence interval

95% CI: point estimate $\pm 1.96 \times SE$

90% CI: point estimate $\pm\ 1.64 \times SE$

99% CI: point estimate $\pm\ 2.58 \times SE$

α-CI: point estimate $\pm z_\alpha \times SE$

- z_α is called the critical z-score for the given α-value.
- the second half of the above expression, $z_\alpha \times SE$, is called the **margin of error** or MOE.

Standard error measures the variability within data. Standard error is the standard deviation for sampling distributions. The formula for standard error when working with a single proportion is

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

where p is the population proportion and n is the sample size. However, when working with confidence intervals, we do not know what the population proportion p actually is. So we use $\hat{p}$ as an approximation, which alters the SE formula to

$$SE \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Warning: If you shift from calculating a p-value to computing a confidence interval, you will need to recalculate SE using $\hat{p}$. Why? Because if you reject the null hypothesis, that means you have rejected p and you should stop using it. Switch to $\hat{p}$.

Example: According to the ASPCA, approximately 44% of households in the US have dogs. A researcher wants to know if the rate of dog ownership is higher in rural areas. He takes a random samples of 80 rural households and finds that 51 had a dog.

A. What is are the null and alternative hypotheses?

B. What are the mean and standard error in the sampling distribution?

C. What is the p-value for this hypothesis test? What do we conclude?

If we rejected the null hypothesis that $p = 0.44$, then we should tell the reader what proportion of rural households do have dogs.

How do we create a 95% confidence interval? We can use what we know about sampling distributions for samples of size 80 with a center of $\hat{p} = 0.638$.

Activity 1: Creating confidence intervals from real world data.
Use the NCBabySmoke data from North Carolina (adapted from OpenIntro Stats) for the following problems.

A. Create a 95% confidence interval for the proportion of new moms who are married.

B. Create a 99% confidence interval for the proportion of newborns who are premies.

column name	description and units
fage	father's age
mage	mother's age
mature	under 35 vs. 35 or older
weeks	length of pregnancy
premie	premie or full term
visits	number of doctor visits
marital	married or not married
gained	weight gained by mom (lbs)
weight	weight of baby (lbs)
lowbirthweight	low is ≤ 5.5 lbs
gender	baby's gender
habit	smoking habit of mom
whitemom	white or not white

Activity 2: For each of the following, conduct a hypothesis test using a real world data set. Then follow-up with a confidence interval if appropriate. Using the NCBabySmoke data from North Carolina (adapted from OpenIntro Stats), conduct a hypothesis test for each of the following. Then follow-up with a 95% confidence interval if appropriate.

A. Are premies 50% girls and 50% boys, or are premie boys more common (in NC)? Note: For this question, you will have considerably less than 1000 babies. Use a pivot table to get a count of premies vs. full term babies, and to sort boys and girls.

B. According to www.childtrends.org, approximatley 8% of pregnant women in the US reported smoking in 2014. Is the rate of smoking higher than this for new moms in NC?