

# Lab: Statistical Inference of Proportions

*Mary Ryan*

*due 11/13 at end of discussion*

## As A Class

- 1) Go to: <http://www.rossmanchance.com/applets/OneProp/OneProp.htm?candy=1>. Set “proportion of orange” to 0.4, and “number of candies” to 50. Select “proportion of orange” and “summary statistics”.
- 2) Draw 1 sample. What sample proportion of orange did you get?

Text:

- 3) Uncheck “Animate”. Draw 50 more samples. What is the mean of the sample proportions? What is the standard deviation?

Text:

- 4) By central limit theorem, what should the mean of the sample proportions be? What should the standard error be?

Code (for use as calculator):

Text:

- 5) Draw 1 more sample. Conduct a hypothesis test to determine if this sample proportion is greater than 0.4.

Code:

Text:

## On Your Own, In Groups

- 1) According to [a survey](#) of 1,364 4-year colleges and universities conducted by the Pew Research Center, 53.3% of colleges admitted at least two-thirds of their applicants in 2017.
  - a) Calculate an 85% confidence interval for the proportion of colleges/universities that admitted at least two-thirds of their applicants in 2017. Interpret it.

Code:

Text:

2) In the fall of 2018, Girls Who Code (GWC) surveyed women affiliated with the organization about their experiences applying for internships in the tech industry (read the full report [here](#) ). Out of 1,017 women, 29% of survey respondents reported having negative experiences during an internship application process.

a) Calculate a 95% confidence interval for the proportion of women affiliated with GWC who had negative internship applications experiences. Interpret it.

Code:

Text:

b) Conduct a hypothesis test to evaluate whether the proportion is larger than 20%.

Code:

Text:

c) Do your answers for (a) and (b) agree with each other?

Text:

## Optional Exercises

- 1) Below is a table illustrating the prevalence of two different biomarkers of Alzheimer's Disease, drawn from a [recent paper](#). Having at least 1 APOE e4 allele is thought to increase your risk for Alzheimer's; having elevated levels of AB protein in the brain is also seen as an increased risk.

##	AB Elevated	AB Non-Elevated
## 0 APOE Alleles	53	182
## 1-2 APOE Alleles	47	53

- a) What is the estimated difference in proportions of elevated AB levels between those who have no APOE e4 alleles, and those who have at least 1 APOE e4 allele? What is the standard error?

Code:

Text:

- b) Conduct a hypothesis test for whether this difference in proportions is equal to 0.

Code:

Text:

- 2) Below are results from a FiveThirtyEight survey about people's opinions of the Oxford (or serial) comma. Answers on the left correspond to the question "How much, if at all, do you care about the debate over the use of the word 'data' as a singular or plural noun?". Answers on the top correspond to the question "How much, if at all, do you care about the use (or lack thereof) of the serial (or Oxford) comma in grammar?"

##		Not at all	Not much	Some	A lot
##	Not at all	77	53	46	27
##	Not much	32	143	152	76
##	Some	8	57	178	109
##	A lot	7	11	38	77

- a) Calculate a 70% confidence interval for the difference in proportions of those who only care "Some" about the Oxford comma debate and "A lot" about the data debate, and those who only care "Some" about the Oxford comma debate and "Not Much" about the data debate. What does this interval tell you?

Code:

Text:

- b) Now conduct a hypothesis test for whether the difference in these two proportions is equal to 0. What can we conclude?

Code:

Text: