MA117 - Worksheet 5
Sampling Distributions
Week 2, Monday

Problem 1. A poll conducted in 2013 found that 52% of U.S. adult Twitter users get at least some news on Twitter. The standard error for this estimate was 2.4%, and a normal distribution may be used to model the sample proportion.

(a) Explain what the standard error means in this context.

(b) Explain how this standard error is calculated.

(c) Construct a 97% confidence interval for the fraction of U.S. adult Twitter users who get some news on Twitter.

(d) Interpret the confidence interval in context.

Problem 2. Of all freshman at a large college, 16% made the dean's list in the current year. As part of a class project, students randomly sample 40 students and check if those students made the list. They repeat this 1000 times and build a distribution of sample proportions. What is the variance of this distribution?

Problem 3. The General Social Survey surveyed a simple random sample of 1151 Americans to ask, "For how many days during the past 30 days was your mental health (which includes stress, depression, and problems with emotions) not good?" Based on responses from 1151 US residents, the survey reported a standard error of 0.21 and a 95% confidence interval for the mean of 3.40 to 4.24 days in 2010.

(a) Explain clearly what the standard error means.

(b) Interpret this confidence interval in context.

Problem 4. The `csv` file

https://sagrawalx.github.io/teaching/data/atlantacrimes.csv

contains data about a simple random sample of 1000 crimes that occurred in the city of Atlanta between 2009 and 2017. Each row contains information about a single crime from this simple random sample. There are three columns: `crime` classifies the crime into one of several possible types, `date` indicates the date the crime occurred, and `neighborhood` indicates the neighborhood in which the crime occurred. Import this data into R using `read_csv` like in the labs, and then use the data to construct a 98% confidence interval for the proportion of crimes that occurred in Atlanta between 2009 and 2017 that were classified as `AUTO THEFT`.

Problem 5. How big a sample would you need in the previous problem if you wanted your 98% confidence interval to be half as wide?