1. Confidence Intervals: Large Sample Size.

Suppose we poll Brown students and ask whether or not they think all undergrads should be required to take statistics. We ask 1000 students. 400 are STEM concentrators and of those, 300 say Yes. 600 are non-STEM concentrators and of those 300 say Yes.

Give a 95% confidence interval for the difference in proportions between STEM and non-STEM concentrators.

Because the sample size is large, we can well approximate the estimator with the Normal Curve. We can also use our estimators to approximate the standard error.

2. Confidence Intervals: Small Sample Size.

Suppose you want to estimate the average score on the midterm exam. Suppose 10 other people from class live in your dorm. You ask those 10 people their scores yielding \{37, 48, 42, 29, 35, 36, 43, 43, 39, 49\}.

In this example, we assume that the true midterm scores are Normally distributed. But, the sample size is small and the variance is unknown. We will use the T distribution instead of the Normal:

The density of the T distribution is:

\[
f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi\Gamma(\nu/2)}}(1 + \frac{t^2}{\nu})^{-\frac{\nu+1}{2}}
\]

The parameter \(\nu\) is known as the degrees of freedom (and is equal to (Sample size - 1) in sampling contexts). How does it arise?

The random variable

\[T = \frac{(\bar{X} - \mu)}{(S/\sqrt{n})}\]

has a T-distribution with \(n - 1\) degrees of freedom, where \(S\) is the sample standard deviation.

(a) Assuming a T-distribution for the estimator, give a 95% confidence interval for \(\bar{X}\) in terms of \(S\) and \(n\).

(b) Compute \(S\) for this data and give a numerical confidence interval.
3. A pollster would like to get an early start on the upcoming presidential elections. The pollster wishes to estimate the proportion of the public who favor Clinton over Bush.

Suppose the pollster wants an error of estimation less than .04 with probability .9. How many people do they need to poll? How is this influenced by the true value of $p$?

4. Suppose we come up with a different scheme to estimate the proportion of people who favor Clinton over Bush in the upcoming presidential election. We stop people walking by on the street and ask their preference (not asking the same person more than once), stopping the first time someone prefers Clinton.

Suppose the 7th person is the first person we ask who prefers Clinton. Given our observation, find the MLE for $p$, the proportion favoring Clinton. Namely, find the value of $p$ that maximizes the probability of seeing our observation.

5. Using real sources (newspaper, online sites) look at polls on the upcoming election. In particular, find the statistical reporting for the polls. What kind of sample sizes are used? What kind of margins of error are there? You do not need to write-up or turn anything in for this problem. Just look around and consider what you see.