1 Populations, Parameters, Samples and Statistics

1. What is the difference between a population and a sample? Given an example of a population you work with and your resulting sample data.

2. What does it mean for a sample to be random? What does it mean for observations to be independent of one another?

3. What does a large sample size imply about the distribution of the parent population the data is generated from? What does a large sample size imply about the distribution of a statistic involving a sum (ie a mean or proportion)?

4. What does it mean for an estimator to be unbiased? What does it mean for an estimator to be consistent?

2 Inference for a Proportion

Let X be a Binomial random variable with probability of success p and n trials. Consider the below hypothesis test.

$$\begin{aligned} H_0 &: p = p_0 \\ H_1 &: p > p_0 \end{aligned}$$

where p_0 is the value of the hypothesized proportion. You observe x successes in n trials so that $\widehat{p} = \frac{x}{n}$.

1. What is the standard deviation of $\widehat{p}$ assuming H_0 is true? What is the test statistic?

2. What is the approximate distribution of the test statistic assuming the null hypothesis is true? What do you require for this approximation to hold?

3. How do you find the p-value for this test? How do you interpret the p-value in this situation?

4. How do you use the p-value to make a decision?

3 Inference for a Mean

Let X be a Normal random variable with mean μ and standard deviation σ. Consider the below hypothesis test.

$$\begin{aligned} H_0 &: \mu = \mu_0 \\ H_1 &: \mu \neq \mu_0 \end{aligned}$$

You observe $x_1, \ldots, x_n$ summarized with $\overline{x}$ and s.

1. What is the test statistic?
2. What is the approximate distribution of the test statistic assuming the null hypothesis is true? What do you require for this approximation to hold?
3. If the observations are not independent, what part of the test statistic is not correct?
4. If you don't reject the null hypothesis, does this prove the null hypothesis is true? If not, what does it prove or not prove?
5. Suppose X actually represents difference between two observations of the same experimental unit (for example, the difference before and after some intervention). Explain why you cannot use a two-sample T test in this situation.

4 Confidence Intervals

1. For a confidence interval, what does it mean to be 95% confident?
2. How can you use a confidence interval to make a conclusion about a hypothesis test?
3. What is the difference between a confidence interval and a prediction interval?