Introductory Statistics – Day 4

Sampling Distributions

Activity 1: Variability in coin flips

- Preview: how many heads do you expect if you flip a coin 20 times.
- Let's build a fair coin simulation in TinkerPlots to model flipping a coin 20 times and repeating that experiment many many times.
- Be sure to click along and create the same simulation on your own computer.

Activity 1, continued:

Once your simulator is built and run, you will have a graph displaying the sampling distribution. Use your sampling distribution to answer the following:

A. How likely is it to get exactly 10 heads (out of 20)?

B. How likely is it to get 8 heads or less?

C. How likely is it to get 5 heads or less?

D. How likely is it to get between 40% and 60% (inclusive) heads?

E. Fill in the blank: The middle 95% of experiments had between _____ and _____ heads.

F. If you flip a coin 20 times and get 18 heads, would you think the coin was unfair/weighted? Why or why not?

Activity 2:

A few college students were discussing the prevalence of underage drinking on campus. Based on anecdotes from their peers, the students believe that roughly 50% of underage college students drink. They later find out that a survey was given to 76 randomly selected underage college students on their campus. Only 32 of the 76 reported regularly participating in underage drinking.

A. Identify the research question.

B. Is this a controlled experiment? If so, what are the treatment(s)?

C. Null Hypothesis:
Alternative Hypothesis:

D. We can use a simulator (TinkerPlots!) to determine the chance of getting a random sample as low as 32 out of 76 (42.1%). If the null hypothesis is true, the probability of getting data this unusual or more unusual is called the **p-value.**

E. Estimate the p-value. What conclusion can we make?

F. Can we draw a cause / effect relationship out of this study?

Definition

A statistical *simulation* is a way to model a statistical situation, generally using a computer to generate repeated samples of a statistical situation. TinkerPlots is a tool for creating statistical simulations.

- A sampling distribution shows us how spread out we expect randomly selected samples to be.
- A sampling distribution gives us the power to recognize whether one sample (of size 76 in this case) is weird or typical.

If a sample is particularly weird, we should suspect that

- It didn't come from the population we thought it did ... *(interesting)* ... or
- Our null hypothesis might be wrong... *(interesting)* ... or
- Our participants are lying ... *(interesting)* ... or
- The sample was unusual because variability is part of statistics
  This is possible, but it has a low probability of being the explanation

If our sample is not very weird, its variability could be just from random chance.

Our job is to quantify **weird** in a mathematical and/or statistical way.

Activity 3:

At a larger state school, a study from 2005 indicated that 65% of underage students drank alcohol. The school engaged is a sustained campaign to lower the rates of drinking across campus. This year, a survey was given to 150 randomly selected underage students at the school. Only 83 of the students reported drinking (55.3%).

A. What is the null hypothesis and how do you use it in your simulator?

B. What is the alternative hypothesis and why does it not appear in your simulator?

C. If the drinking rates have not changed since 2005, whats the chance of getting a sample as low (or lower) than the one in this study? i.e. What portion of your randomly generate samples were this low?

D. What is the p-value and what does it mean?

E. Was this a controlled experiment?

F. Can we draw a cause / effect relationship out of this study?

Activity 4:

At a typical 4 year college, about 25% of students are in each class. Build a sampler that has 25% Freshmen, 25% Sophomores, 25% Juniors, and 25% Seniors. The number of students who access health services on campus should be the same across the 4 classes, but there is a concern that freshmen are less likely to seek help. Last week, out of 120 students who visited health services, only 20% were freshmen. Is this 20% in the range of normal variability, or does it represent a statistically extreme situation?

A. What is the null hypothesis and how do you use it in your simulator?

B. What is the alternative hypothesis?

C. Build a simulation to learn how random samples of size 120 behave. One might anticipate that there would be 30 students from each class, but the simulation will show what level of variability is typical.

D. Create a sampling distribution for the proportion of freshmen and determine where the middle 95% of samples are.

E. How unusual is a sample with 20% or fewer freshmen? Find the p-value and state your results as a complete sentence related to the context.