MA206, Lesson 4 - Strength of Evidence

What is Statistical Significance?

Statistical Significance indicates the strength of evidence that the observed result was unlikely to have occurred by random chance

What is the 3S Strategy?

Statistic, Simulate, Strength of Evidence

What is the difference between a **parameter** and a **statistic**?

A parameter is the long-run behavior of a random process and it is not observable. A statistic is the summary of observed results and is used to infer about the parameter.

Define:

H_0 :
Null Hypothesis, the "Random Chance" model for the long-run behavior of the parameter.

H_a :
Alternate Hypothesis, the "Something More" model for the long-run behavior of the parameter.

π :
The proportion parameter, the long-run behavior of a random process. It is unobservable.

$\hat{p}$:
The observed proportion statistic from data, used to argue or infer about the parameter

n :
The sample size; The number of observations within each sample to construct the statistic

p-value :
The probability of observing an observation at least as extreme as $\hat{p}$ assuming the null hypothesis is true.

Classify the strength of evidence for each range of p-values.

0.1	$< p$		Weak Evidence against the null
0.05	$< p \leq$	0.1	Moderate Evidence against the null
0.01	$< p \leq$	0.05	Strong Evidence against the null
	$< p \leq$	0.01	Very Strong Evidence against the null

1) CPT Rocha dreams of running West Point's best cookie bakery. To test his chances of success, he bakes chocolate chip cookies and has his cadets taste them in a blind comparison to two unnamed competitors. If we assume there is no difference in taste between the three cookies, we would expect an even distribution of preferences as cadet choice would be left to randomly guessing. This method was conducted among the class.

For your class, what is the ratio of cadets that chose CPT Rocha's cookie as their favorite?

What is the ratio of cadets that chose mystery cookie K as their least favorite?

a) What is the research question?

Do cadets prefer CPT Rocha's cookies to the competitors, in general?

b) What are the observational units in this study?

The individual cadets in the classroom.

c) What are the variable(s)? Classify each as categorical or quantitative.

The variable is which cookie was chosen as the tastiest, and it is categorical (with three levels).

d) Describe the parameter of interest. What symbol is used to represent it?

The parameter of interest is the proportion of people who choose CPT Rocha's cookies over the competitors, represented by π

e) For your class, list the appropriate values for the variables below and describe them in words.

n: Will vary by class, generally 18;
The number of observations in the sample, aka the number of individuals given cookies and asked to choose one to eat. Here, class size.

$\hat{p}$: Will vary based on class; The observed proportion of students in the class who choose CPT Rocha's cookies as their favorite to eat. $\frac{Count\ who\ chose\ CPT\ Rocha's}{Total\ cadets}$

H_0: $\pi = \frac{1}{3}$. The null hypothesis. The true long run proportion of those who prefer CPT Rocha's cookie is equal to 1/3 (the class will choose CPT Rocha's cookie randomly.)

H_a: $\pi > \frac{1}{3}$. The alternate hypothesis, the true long run proportion of cadets who will choose CPT Rocha's cookie is greater than $\frac{1}{3}$

f) Is it possible for the sample to match the null hypothesis if the null hypothesis is false?

Yes, it is possible.

g) Is it possible to differ significantly from the null hypothesis if the null is true?

Yes, it is possible.

h) How might we use dice to randomly simulate and test our null hypothesis?

We might roll a 6-sided dice and select a 5 or 6 to represent a "success," or that someone chose CPT Rocha's cookie. We could repeat this n times for a sample of size n, or alternatively roll n dice and count the 5s and 6s.

i) Using the applets, run a simulation using your observed statistic, your hypotheses, and at least 1,000 simulations. Describe the null distribution. Describe where the observation falls (near the center or in the tails).

Answers may vary based on class output. There should be a sample size n matching the class size, simulations of 1,000 or more, and probability of success of $\frac{1}{3}$. The distribution should be centered on 1/3, roughly normal looking, with variability from 0.08 to 0.67 (ish).

j) List your p-value from the simulation and describe the strength of evidence in words. Have we proven the alternate hypothesis?

Answers will vary. Strength of evidence should match the scale given in class. No, we have not proven the alternate hypothesis - it is still possible to get this result from chance, even if unlikely.

0.1	$< p$		Weak Evidence against the null
0.05	$< p \leq$	0.1	Moderate Evidence against the null
0.01	$< p \leq$	0.05	Strong Evidence against the null
	$< p \leq$	0.01	Very Strong Evidence against the null

h) Be able to repeat the above work for a different research question: Is the Keebler Chips Deluxe the least favorite cookie, in general? Repeat the analysis, provide a p-value, and write your conclusion.

2) Many types of studies have been done to see whether animals have a sense of number. In one study (Beran & Beran, 2004), researchers wanted to see whether chimpanzees could remember how many items were added to a container when they see them added one at a time. To do this they had two containers set up so the chimpanzees could not see inside them. They randomly added bananas to the containers one at a time and then had the chimpanzees choose the container they wanted. In one set of experiments, they added 3 bananas to one container and 4 to the other and had a chimpanzee named Mercury pick which container he wanted. In 20 trials, he picked the container containing 4 bananas 16 times.

a) What is the observed statistic? Include the appropriate symbol.
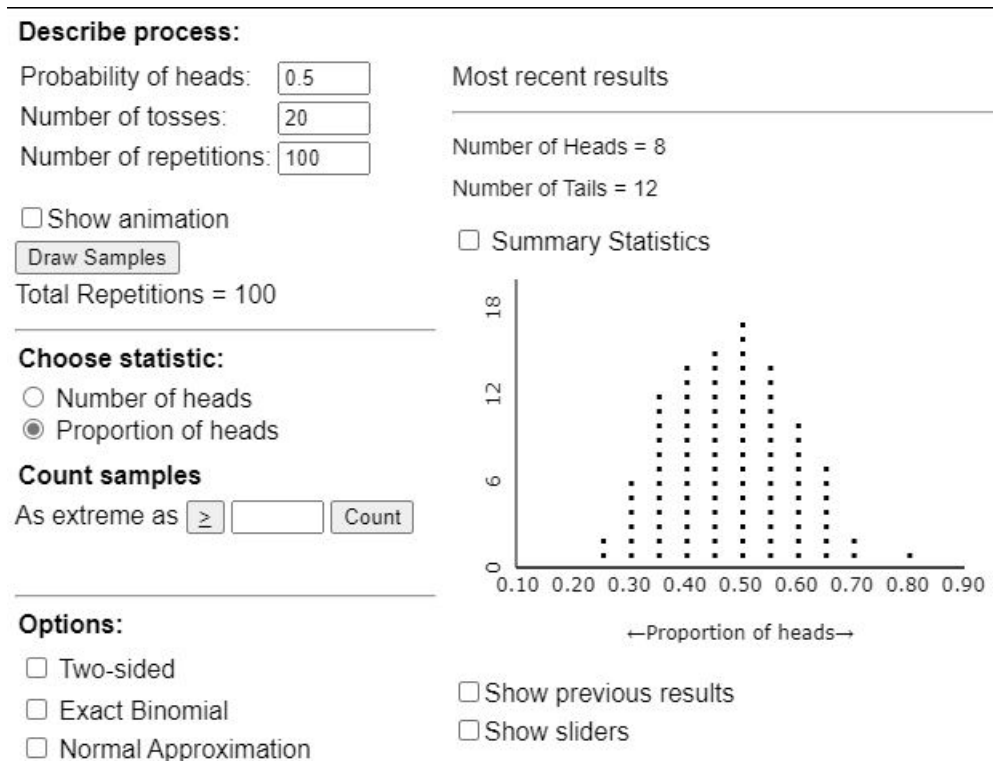
$\hat{p} = \frac{16}{20} = 0.8$

b) In words and symbols, define the null and alternate hypotheses for this study.

$H_0 : \pi = 0.5$. The long run proportion of times that Mercury picks the container with more bananas is 50%.
$H_a : \pi \geq 0.5$. The long run proportion of times that Mercury picks the container with more bananas is greater than 50%.

c) Below is a screenshot of a simulation done 100 times. What is the p-value, given our observed statistic

Describe process:

Probability of heads: 0.5
Number of tosses: 20
Number of repetitions: 100

☐ Show animation
Draw Samples
Total Repetitions = 100

Choose statistic:
○ Number of heads
◉ Proportion of heads

Count samples
As extreme as [≥] [   ] Count

Options:
☐ Two-sided
☐ Exact Binomial
☐ Normal Approximation

Most recent results
_____

Number of Heads = 8
Number of Tails = 12

☐ Summary Statistics

0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90
←Proportion of heads→

☐ Show previous results
☐ Show sliders

Our p-value is $\frac{1}{100} = 0.01$. This is very strong evidence against the null hypothesis that Mercury was choosing the boxes randomly.

We would expect it to be centered on 10. We have a $\pi = 0.5$, $n = 20$, and $0.5 \times 20 = 10$.

e) Given the above distribution, what is the fewest number of times that Mercury could pick the box with more bananas and we would still consider it strong evidence against randomly selecting boxes?

14 times. Above, a proportion 0f 0.70 has a p-value of 0.03, but a proportion of 0.65 has a p-value of 0.1. Strong evidence is ≤ 0.05, so we choose 0.70. With a sample size of 20, $0.7 \times 20 = 14$.

3) Your friend says he can shoot free throws as well as someone in the NBA and you don't think he is that good. You know that the NBA average for shooting free throws in the 2021-2022 season was 77.8%. You test your friend by asking him to make 40 shots, and he makes 24 of them. Make a conclusion about your friend's claim, including your p-value.

$\hat{p} = \frac{24}{40} = 0.6$
$H_0 : \pi = 0.778$

$H_a : \pi < 0.778$

p-value will change depending on simulation results, but should be about 0.008.

With a p-value of 0.008, we have very strong evidence against the null hypothesis that your friend is as good as an NBA player at making free throws.