Hypothesis Testing 1

What other types of Questions can Statistics answer?

Grinnell College

Review – Intervals

We have been covering Confidence Intervals for a bit now.

What was the purpose of confidence intervals?

- estimating a parameter

Methods

- Normal methods (p, $p_1 - p_2$)
- t-distribution (μ, $\mu_1 - \mu_2$)
- bootstrap distribution (simulating new statistics over and over)

Answering Questions with CIs

Sometimes when we had a confidence interval we would check whether a certain value was within that interval.

Example: We want to find out if a coin is fair. We flip a coin a whole bunch and from our data we construct some confidence intervals.

90% CI $\rightarrow$ (.47, .49)

- According to the CI, is the coin fair?

95% CI $\rightarrow$ (.45, .51)

- According to the CI, is the coin fair?

Answering Questions with CIs

Sometimes when we had a confidence interval we would check whether a certain value was within that interval.

Are there issues with this method? Yes!

- Our answer depends on the confidence level of our CI

Court Case

Suppose you are selected to be on a jury to determine if someone is a murderer.

What assumption do we make before the trial?

What are the two decisions we can make?

What do we use to make that decision?

How much evidence do we need to make a conviction?

If we find someone "not guilty" does that really mean they are "not guilty?"

If we find someone "guilty" does that really mean they are "guilty?"

Hypothesis Testing

Hypothesis Testing is the term we are going to give for figuring out how to answer binary questions using data

Examples

- did someone commit a crime? (guilty / not guilty)
- is a coin fair? (yes / no)
- is a new drug better than existing drugs? (yes / no)

Note: this process does not work for prediction

- ex) will it rain tomorrow?

Parameters and Statistics (Review)

Parameters are numerical summaries of the population
Statistics are numerical summaries of the sample

Typically we will use special notation to differentiate *population parameters* (things we wish to know) from *statistics* computed from our sample:

	Population Parameter	Sample Statistic
Mean	μ	$\overline{x}$
Standard Deviation	σ	s
Proportion	p	$\hat{p}$
Correlation	ρ	r
Regression	β	b's or $\hat{\beta}$'s

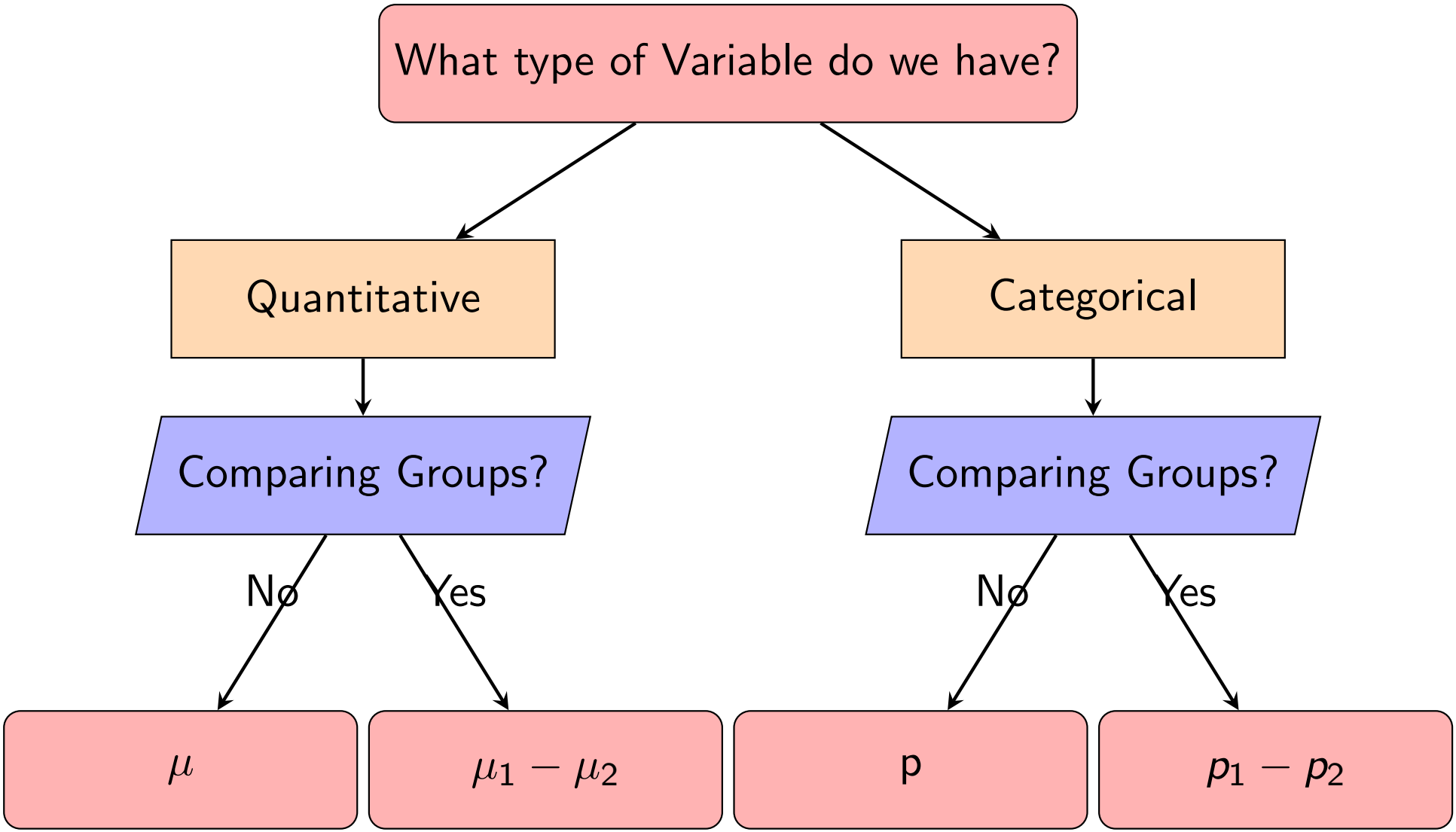
- It is EXTREMELY important that you can define the parameter and statistic in context from here on out
- failure to do so means we can't even start hypothesis testing

Parameters and Statistics

For the rest of hypothesis testing (until told otherwise), we are going to focus on these four specific scenarios and the following parameters

- population mean (μ)
- difference in population means ($\mu_1 - \mu_2$)
- population proportion (p)
- difference in population means ($p_1 - p_2$)

Which parameters to use?



Hypothesis Statements

There are two possible outcomes for testing a research question.

- The data supports the research question
- The data *does not* support the research question

A **Hypothesis Statement** is a statement about a parameter based on the research question.

The Null Hypothesis

Null Hypothesis – hypothesis statement that represents an assumption of no effect or no relationship or no difference between variables (status-quo)

- uses the most basic assumption we can make about a parameter
- sometimes based on previous information
- denoted H_0 (H-"naught" or H-"oh" or H-zero)

Form will always be...

- H_0: parameter = 'hypothesized value'
- "our null hypothesis is that the parameter equals the hypothesized value"
- 'hypothesized value' is often written as the parameter with a 0 subscript
  - ex) μ_0, p_0

Null Hypothesis Examples

Common examples include:

- Testing if a pop. mean is equal to zero:

$$H_0 : \mu = 0 \qquad (\mu_0 = 0)$$

- Testing if difference of proportions between groups is zero

$$H_0 : p_A - p_B = 0 \qquad (p_0 = 0)$$

- Testing if odds ratio is equal to one (won't spend time on this):

$$H_0 : \theta = 1 \qquad (\theta_0 = 0)$$

The Alternative Hypothesis

Alternate Hypothesis – a hypothesis statement that represents what we want to show with evidence, based on the research question

- claim we want to find evidence for
- denoted H_A

Will look similar to H_0 but with a change.

- H_A: parameter $<$ 'hypothesized value' (left-tailed test) **OR**
- H_A: parameter $>$ 'hypothesized value' (right-tailed test) **OR**
- H_A: parameter $\neq$ 'hypothesized value' (two-tailed test)

The research question will determine which of these we actually use

- always same hypothesized value as H_0

Example: Writing Hypotheses

In 1998, as an advertising campaign, the Nabisco Company announced a "1000 Chips Challenge" claiming that every 18-ounce bag of their Chips Ahoy cookies contained at least 1000 chocolate chips.

What type of parameter will we work with to test this?

Define the Null hypothesis for this research question

Define the Alternative hypothesis for this research question

Example: Writing Hypotheses

We have data on two random samples of diamonds: one with diamonds that weigh 0.99 carats and one with diamonds that weigh 1 carat. We want to find out if the price of these 2 different types of diamonds are the same, or if they differ.

What type of parameter will we work with to test this?

Define the Null hypothesis for this research question

Define the Alternative hypothesis for this research question

Coin Flip Example

Let's go back to testing the fairness of a coin. What is the best possible guess we could give for the 'true proportion of heads' a coin will land on if we haven't yet tested a coin?

Research Question: Is the coin biased in favor of heads?

What type of parameter will we work with to test this?

Define the Null hypothesis for this research question

Define the Alternative hypothesis for this research question

More on Null Hypotheses

Some students find defining the null hypothesis confusing. How do we know what to use for the null hypothesis?

It may be helpful to think of what the null hypothesis is being used for: It is a point of comparison (or argument) that we are going to compare our data against to see if they match up.

Another way to think about H0 is that it is a position that gets the 'benefit of doubt' or can be 'taken for granted', that we are OK allowing (or pretending) it is true without evidence.

More on Null Hypotheses

In our statistics class, we are primarily focused on testing specific parameter values. When framing a null hypothesis, sometimes we may run into an issue where people argue for a different default position.

Example: Energy Drinks
A company claims a new energy drink is safe for consumptions. Which should be the default?

- H_0: the energy drink is safe (until 'proven' otherwise)
- H_0: the energy drink is harmful (until 'proven' otherwise)

Defining our null hypothesis actually requires us to think about who the 'burden of proof' falls on. Now, consider which situation we will most likely find ourselves in when performing a hypothesis test.

Coin Flip Example

Goal: Finding out if $p > 0.5$

Method: We are going to conduct a study, gather data by flipping the coin, and see how our data compares to H_0: $p = 0.5$

- essentially seeing does data match up with the coin being fair
- defining H_0: $p = 0.5$ gives us the point of comparison

Coin Flip Example

Research Question: Is the coin biased in favor of heads? $\rightarrow$
H_0: true proportion of heads = p = 0.5

How would we go about testing this question? Let's say we flip the coin 10 times.

- We expect to get a proportion of heads around 0.5 if the coin is fair (hypothesized value)
- Will we get exactly $\widehat{p} = .5$ every time even if the coin is fair?
- What '# heads'/10 would make you think the coin is unfair?

Coin Flip Simulation

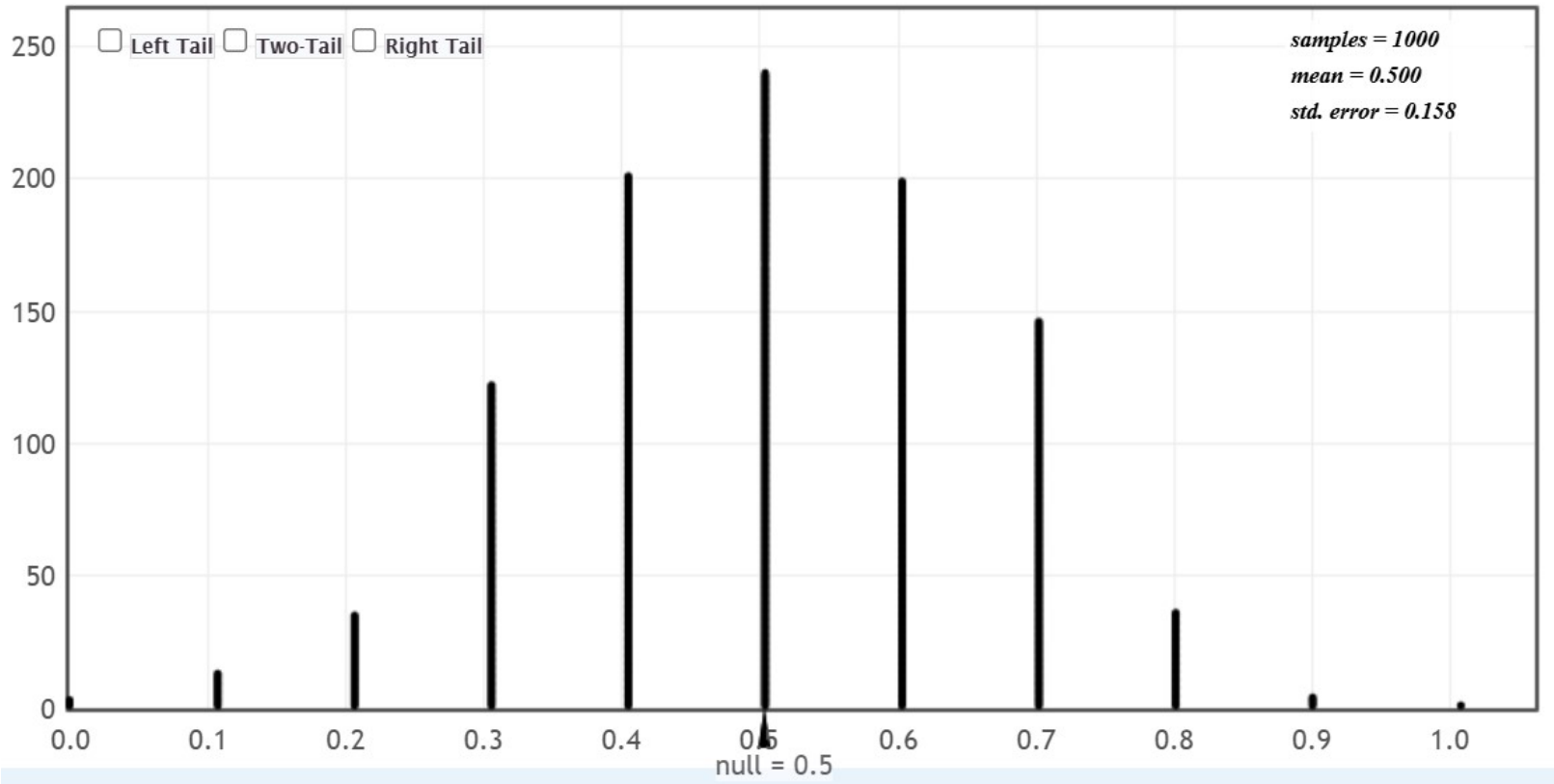
How do we go about this?

If we want to compare our results to what we get with a fair coin... let's see what simulating a fair coin actually looks like.

We are going to simulate a bunch of coin flips.

- flip **fair** coin 10x, compute $\hat{p}$
- repeat 1000x
- make distribution of 1000 $\hat{p}$'s

Coin Flip Example



This resulting distribution is called a "Null Distribution".

- it simulates what results would look like if H_0 is really true.

What shape do we see? What is the center of the distribution?

Coin Flip Example

Goal: Finding out if $p > 0.5$

Now we need to test the actual coin. Let's say I flipped the coin in question 10 times and got 8 heads

- this means $\widehat{p} = .8$
- where is 0.8 on our randomization distribution? Is this rare?
  - let's calculate a probability associated with this to quantify 'rareness'

Coin Flip Example

Goal: Find out if $p > 0.5$

What we have:

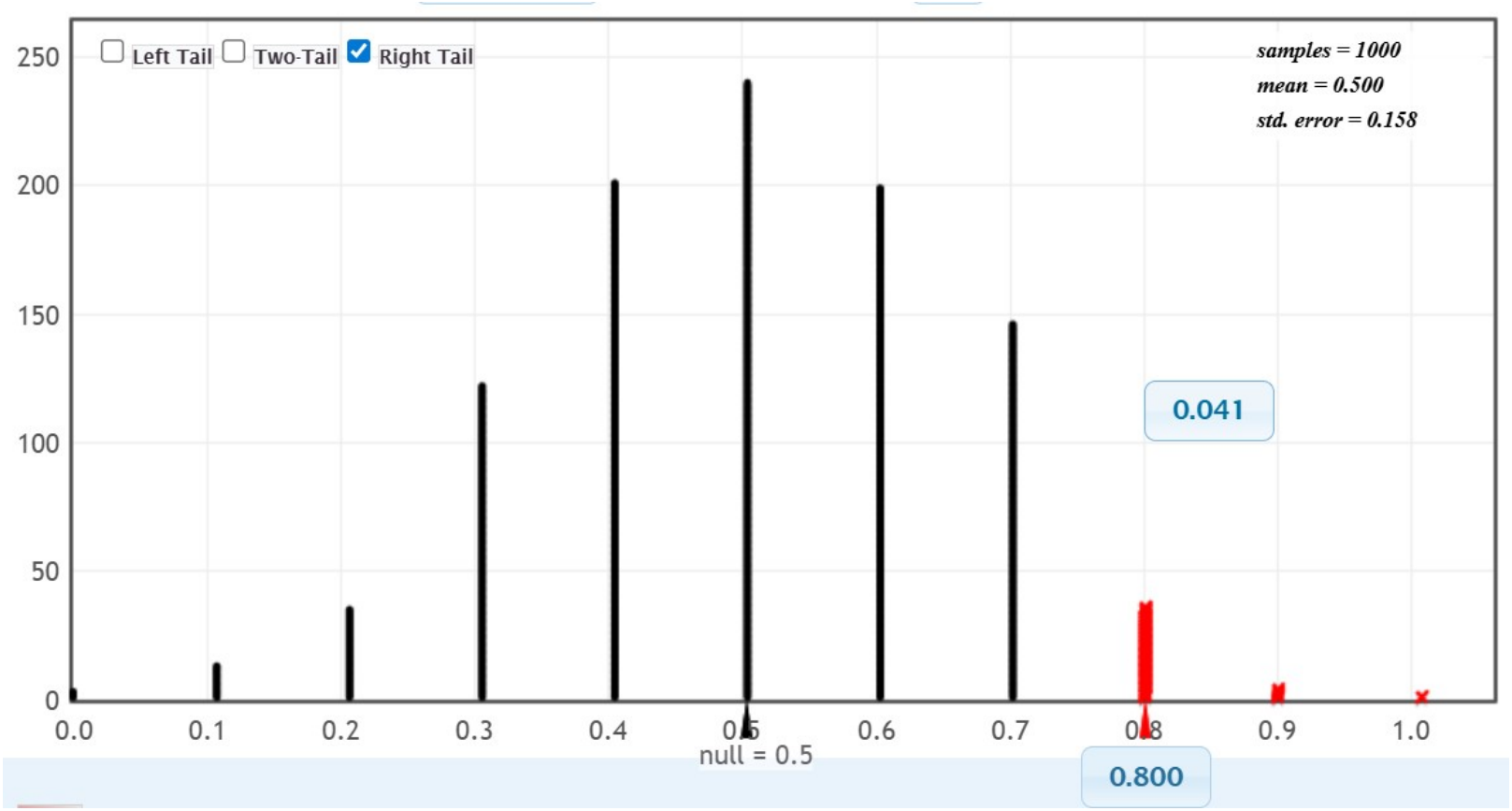
- Null distribution (simulates results of fair coin flipped 10x)
- results of us flipping the real coin: $\hat{p} = 0.8$

Quantifying how rare this result is:
What is the probability of getting this result or a result more extreme?

- compare our result to the null distribution
- give this a special name: **p-value**

Coin Flipping: P-value

If we have $\widehat{p} = 0.8$, we get a p-value of:



This is the prob. of 8 or more H in 10 flips of a fair coin. Does 'fair' seem reasonable?

P-values

We will talk more about these next week too.

P-value Definition:
The probability of getting a statistic equally or more extreme than what we did UNDER THE ASSUMPTION THAT H_0 IS TRUE

- P(getting statistic value more extreme than we got $\|$ H_0 is true)
- looking at how unlikely our data/result is
- NOT looking at how unlikely H_0 is

Wrapping up

What kinds of questions can we answer with CIs?

What kinds of questions can we answer with hypothesis tests?

What does a null distribution show us?

Broadly, what do p-values tell us?