Hypothesis Testing

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 (everything else)

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

- is someone a murderer? (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?

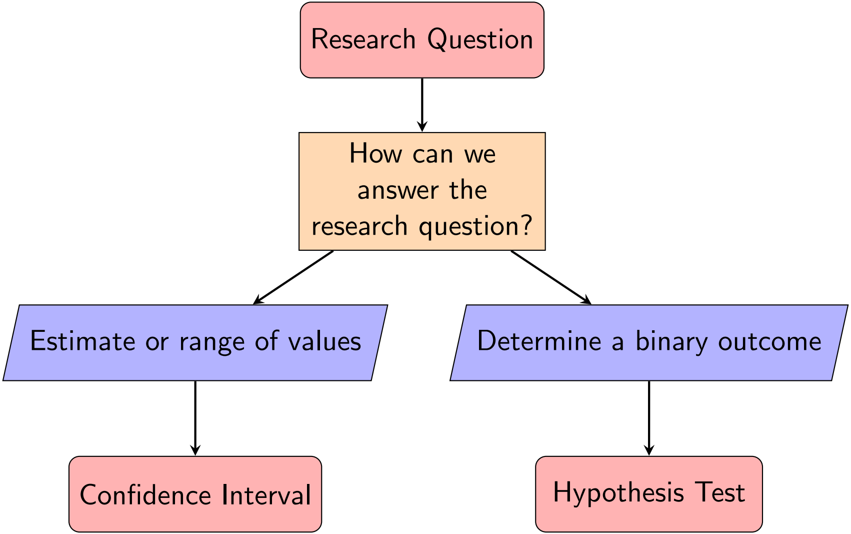
Types of Questions in Statistics

So far we have broadly encountered two scenarios in stats.

- we want to estimate something (a parameter)
- we want to answer a binary question about a population

For the remainder of the semester, I want you think about Research Questions that we formulate with this idea in mind...

Which method to use?



Parameters and Statistics

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

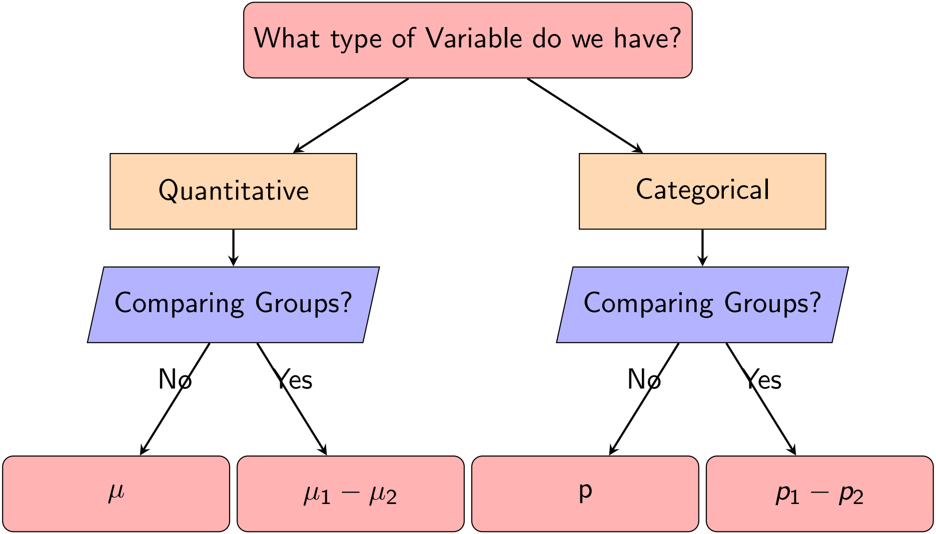
- It is EXTREMELY important that you can define the parameter and statistic in context from here on out
- failure to do so means we can't even start hypothesis testing

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:

$$H_0 : \mu = \mu_0 = 0$$

- Testing if difference of proportions between groups is zero

$$H_0 : p_A - p_B = p_0 = 0$$

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

$$H_0 : \theta = \theta_0 = 1$$

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) **<u>OR</u>**
- H_A: parameter $>$ 'hypothesized value' (right-tailed test) **<u>OR</u>**
- 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

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 fair?

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

Research Question: Is the coin fair?

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

We are going to simulate a bunch of coin flips.

Go to "https://flipsimu.com/"

- at the bottom, adjust the # of coin flips to 10
- flip the coins and compute the sample proportion of heads $\widehat{p}$
- do it again
- mark results on the board

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 close to?

- What name did we give to this value?

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? **Normal**

What is the center of the distribution close to? **0.5**

- What name did we give to this value? **hypothesized value**

Coin Flip Example

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?
- If the coin was not-fair, couldn't we also expect to get low values of $\widehat{p}$ if the coin was biased in the other direction?
  - both small and large values for $\widehat{p}$ provide evidence the coin isn't fair

Let's find the proportion of times we got a result of as extreme as the data we collected from our results ($\widehat{p} \geq 0.8$ and $\widehat{p} \leq 0.2$)

- these are the values that provide just as much, if not more, evidence against the coin being fair
- this has a special name: **p-value** (short for 'probability-value')

Coin Flip Example

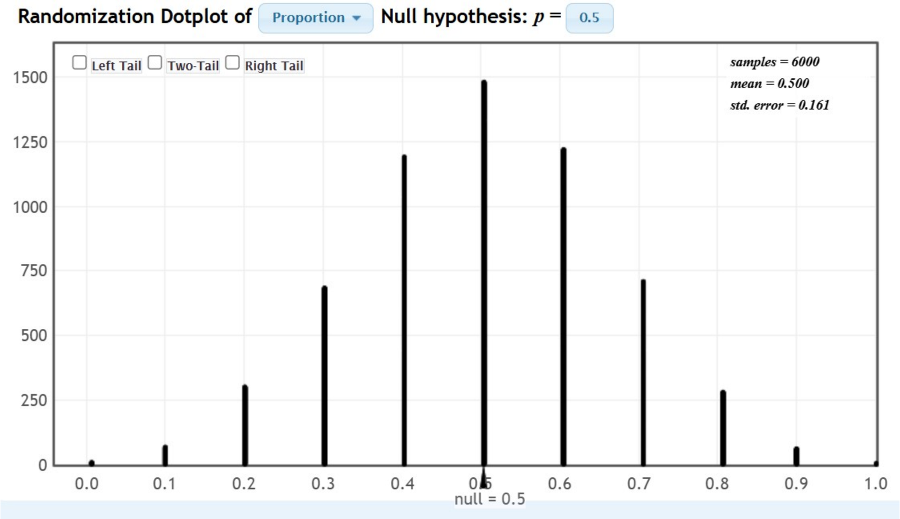
If we have $\widehat{p} = 0.8$, we get a p-value of:

What if we had $\widehat{p} = 0.9$?

- p-value $=$
- is this stronger or weaker evidence that the coin is not fair?

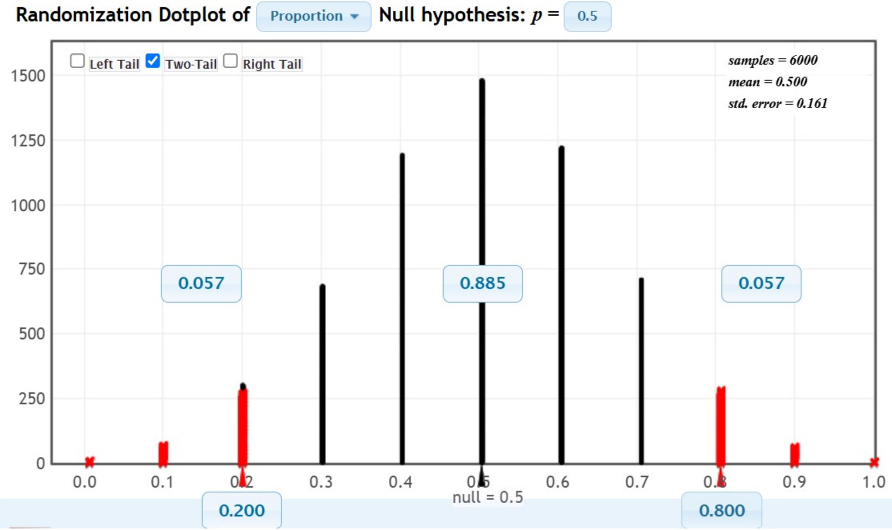
Coin Flip Example

What if we had even more trials of the 10 coin flips?



Coin Flip Example

With $\widehat{p} = .8$, what is the p-value? $0.057 + 0.057 = 0.114$



P-values

P-values are a way of quantifying how strong the evidence is *against* the Null Hypothesis

- equivalently: how strong the evidence is *in support* of the Alternative Hypothesis

Formal definition:
The probability of getting an observed statistic equal to or more extreme than what we got *IF THE NULL HYPOTHESIS IS TRUE*

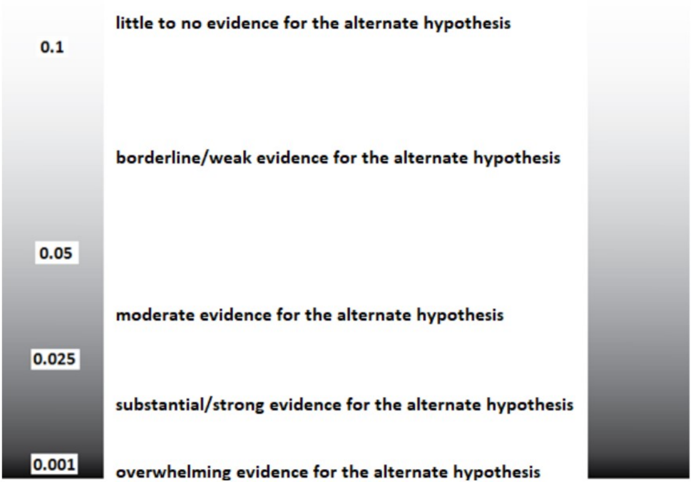
- this is a conditional probability

Interpretation:

- 'More extreme' = contrary to the Null hypothesis
- 'More extreme' is defined by the alternative hypothesis
- smaller p-value → more evidence against H_0
- smaller p-value → more evidence in favor of H_A

How do we quantify the evidence strength?

To do this we will look at the p-value. "Where does it fall on this chart?"



$\widehat{p} = 0.8 \rightarrow$ p-value $= 0.114$

Coin Flip Example

With $\widehat{p} = .8$, the p-value of 0.114 indicates there is little to no evidence against the Null hypothesis

- Null hypothesis: coin is fair
- $\widehat{p} = .8$ is not particularly rare in a trial of 10 flips $\rightarrow$ little to no evidence to say the coin is biased

Conclusion: there is little to no evidence that the coin is biased

Null Distribution

When we calculate p-values we are always going to do so using a distribution that simulates what the null hypothesis would look like

- compare our statistic to the null distribution
- compute p-value using the sign in the alternate hypothesis

Null Distribution:
The distribution of the statistics if the null hypothesis is true

- simulates what the null hypothesis looks like

Finding P-values

When calculating a p-value you must ALWAYS look at the sign in the alternative hypothesis. This determines the direction in which you find the area. Then look at the null distribution and find where the statistic value is

H_A: parameter $>$ 'hypothesized value'

- Find area 'greater than' the statistic on the null distribution
- "right-tailed test"

H_A: parameter $<$ 'hypothesized value'

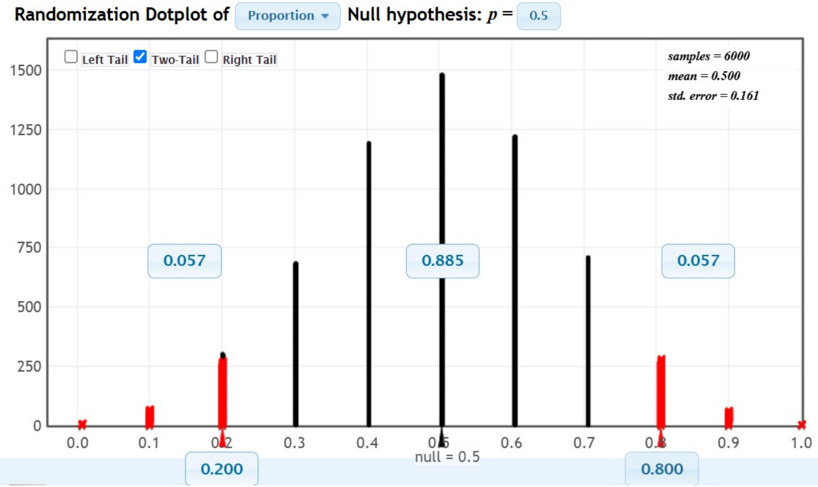
- Find area 'less than' the statistic on the null distribution
- "left-tailed test"

H_A: parameter $\neq$ 'hypothesized value'

- Find area 'greater than' the statistic (if statistic is positive) and area 'less than' -statistic
- "two-tailed test"

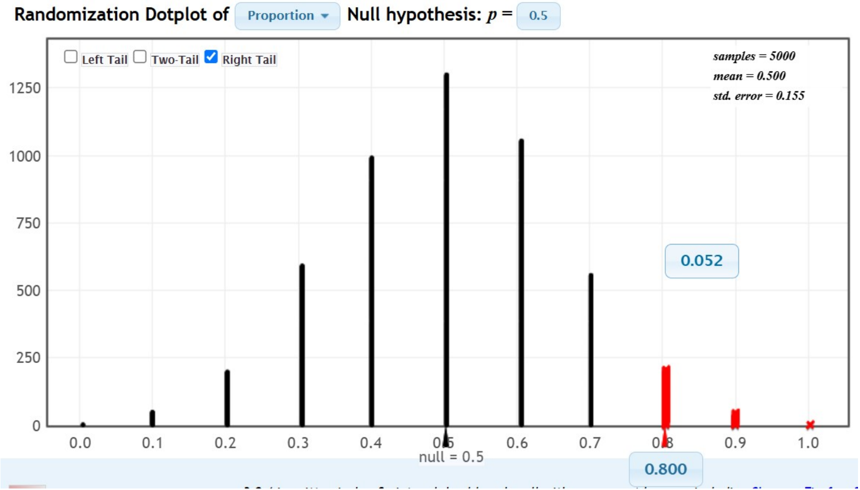
Coin P-value Example 1

H_A: $p \neq 0.5$, with $\widehat{p} = .8$, the p-value is $0.057 + .057 = 0.114$



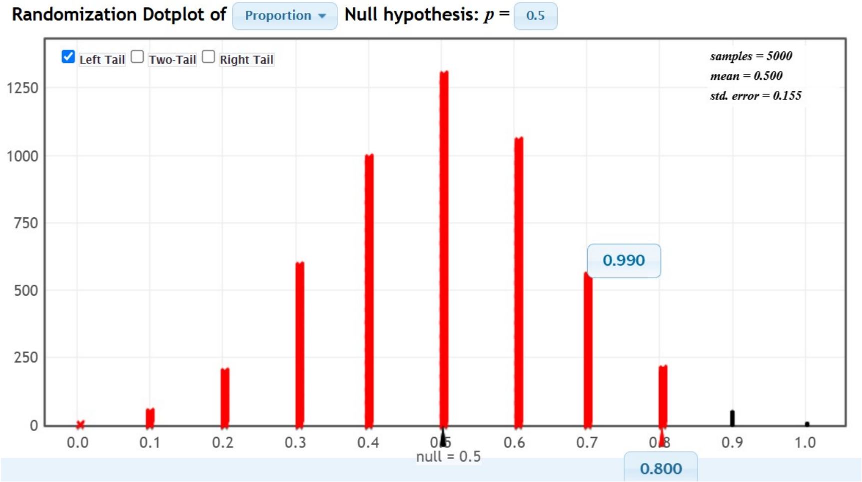
Coin P-value Example 2

H$_A$: p $>$ 0.5, with $\widehat{p} = .8$, the p-value is 0.052



Coin P-value Example 3

H_A: $p < 0.5$, with $\widehat{p} = .8$, the p-value is 0.99



P-value Interpretation

Interpretation:
(value of the p-value) is the probability of getting a (statistic) of (statistic value) or more if (null hypothesis) is true

- replace placeholders with our values and context
- basically just saying the definition of a p-value with values and context

Conclusion:

- mention strength of evidence, parameter, and context
- try to answer the research question with our evidence

There is (strength) evidence that the (parameter) is (pick $>$, $<$, $\neq$) the hypotheized value

Test-Statistic

For null distributions, we will use the Normal and t-distribution stuff like we did with CI's, but with a small twist...

To make things, easier, we are going to calculate something called a "test-statistic" that will help us in calculating p-values.

- will always be labeled with a Z or a T

The test-statistic is a value we calculate and are going to use with the pnorm() or pt() functions in R to get p-values

It will always have the form:

$$\text{test-statistic} = \frac{\text{statistic - hypothesized value}}{\text{standard error}}$$

Central Limit Theorem

When certain conditions are met (normal pop. or large sample size), we saw that the sampling distribution of means and proportions is a Normal distribution.

We will use this as the starting point for constructing null distributions.

Hypothesis Test – Single Proportion

H_0: $p = p_0$

Under the null hypothesis we have:

$$Z := \frac{\widehat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathbf{N(0,1)}$$

- use pnorm() with value of Z

Conditions:

- Random Sample
- $n \times p_0 \geq 10$
- $n \times (1 - p_0) \geq 10$

Wrapping up

What kinds of questions can we answer with CIs?

What kinds of questions can we answer with hypothesis tests?

Why do we calculate p-value (think of strength of evidence)?

What is a test-statistic and what form do they have in general?