Hypothesis Testing 2

More on Null Distributions and P-values

Grinnell College

Review

Hypothesis Testing: formal technique for answering a question with two competing possibilities

Null Hypothesis: represents a skeptical view or a perspective of no difference

- H_0: 'parameter' = (some value)

Alternate Hypothesis: what the researchers actually want to show with the study

- H_A: 'parameter' [$<$ / $>$ / $\neq$] (some value)
- choose the sign to match the research question
- both hypotheses use the same value

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

Coin Flip Example

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

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

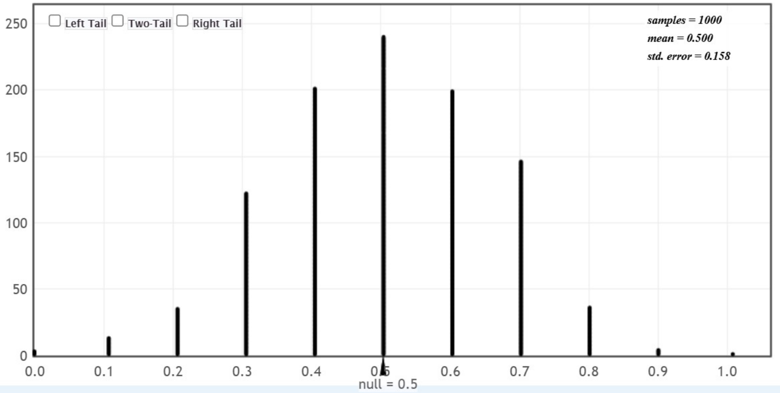
- We expect to get a proportion of heads around 0.5 if the coin is fair (hypothesized value)
- Will we get exactly $\widehat{p} = .5$ every time even if the coin is fair?
- What '# heads'/10 would make you think the coin is unfair?

Coin Flip Simulation

We are going to simulate a bunch of coin flips.

- 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 $p > 0.5$

Let's say I flipped the coin in question 10 times and got 8 heads

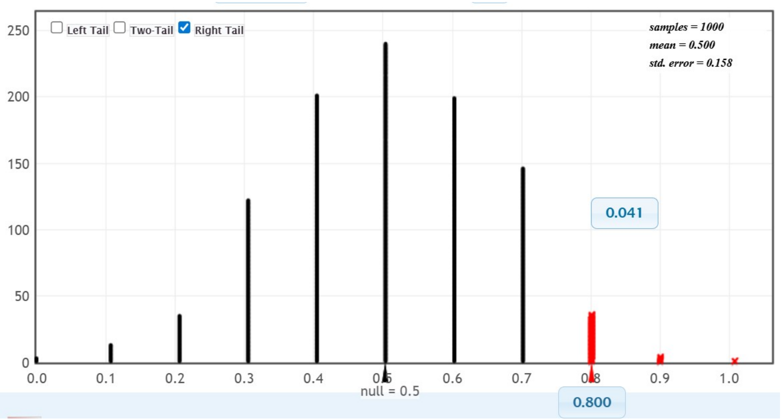
- this means $\widehat{p} = .8$
- where is 0.8 on our randomization distribution? Is this rare?

What is the probability of getting this result or a result more extreme? ($\widehat{p} \geq 0.8$)

- 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 Flipping: P-value

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



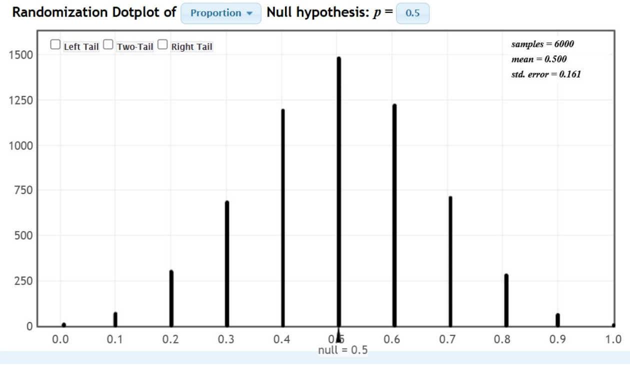
This is the prob. of 8 or more H in 10 flips of a fair coin. Does 'fair' seem reasonable?

Null Distribution

The distribution of the statistics if the null hypothesis is true

- simulates what the null hypothesis looks like
- use this to compute p-values

We looked at the coin-flip scenario. Null distribution of fair coin, 10 flips



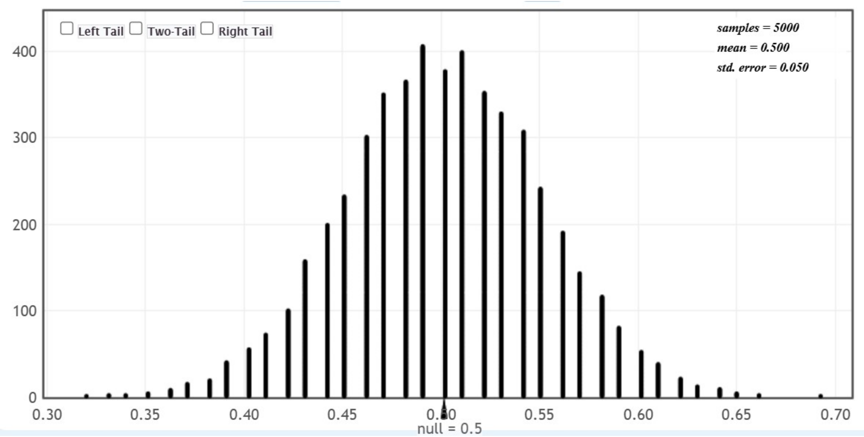
Null Distribution

What if we flipped 100 coins instead of 10

- Would getting $\hat{p} = .8$ be more common or less common?

Null Distribution

This represents 'H_0: p = 0.5', when the sample size (# of flips) is 100



- What do we see?
- Are large or small values of $\hat{p}$ more/less common?

Null Distribution

The **null distribution** will look very similar to the sampling distribution stuff we saw before

- this time it is simulating the null hypothesis
- for means and proportions this looks like a Normal curve

These distributions looks Normal when certain conditions are met. Very similar to what we had when we were using confidence intervals.

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*

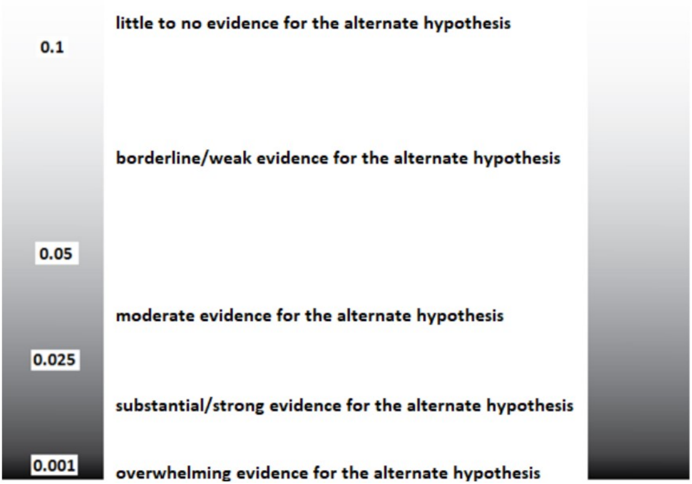
- this is a conditional probability

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 $= 0.041$

Coin Flip Example

With $\widehat{p} = .8$, the p-value of 0.041 indicates moderate evidence against the Null hypothesis

- Null hypothesis: coin is fair
- $\widehat{p} = .8$ is not particularly rare in a trial of 10 flips $\rightarrow$ moderate evidence to say the coin is biased in favor of heads

Conclusion: there is moderate evidence that the coin is biased in favor of heads

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

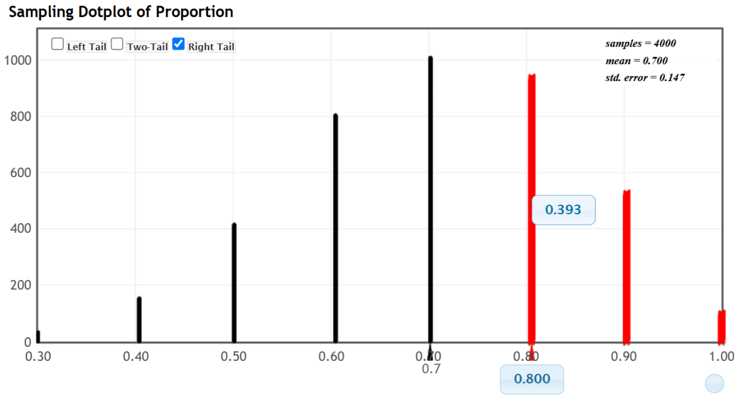
Conclusion:

- 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 hypothesized value $\rightarrow$ practical takeaway

P-value relies on Null Distribution

P-value is a conditional probability, which depends on the null hypothesis and thus the null distribution. If you use a different null hypothesis you get a different p-value.



- $\hat{p}$ of 0.8 or more is less rare when the true proportion of heads is 0.7 than when it was 0.5