HUDM4122
Probability and Statistical Inference

April 1, 2015

First Announcement

- HW8 will be due on April 15, rather than April 13
- I don’t expect us to get through the entire lecture today, so I decided to delay the homework rather than splitting it

HW7

Q5

- Take a variable with mean = 12 and SE = 6.
- What is the variable's lower bound for its 90% Confidence Interval?
(Give two digits after the decimal place)

- Answers were all over the place, so let's go over this together

Q6

- You are testing a new brand of Moose Chow.
  You feed it to 25 meese.
- The meese eat an average of 10 pounds of Moose Chow.
  The standard deviation for how much they eat is 1 pound.

  What is the upper bound of the 95% confidence interval for the average amount
  of Moose Chow a moose eats?

- A lot of people got incorrect answer of 11.96, which comes from confusing standard deviation with standard error…
- Let’s take a look

Q10

- Your favorite sports team is already 25 games into their season, and has a win-loss record of 15-10 (0.6). What is the lower bound on the 95% confidence interval for what percentage of games they will win by the end of the season?
  (Give two digits after the decimal)

- No common wrong answers, so let’s go over this together

Other questions/comments on the hw?

Statistical Significance Testing

- The core of the traditional "frequentist" paradigm of statistics

Statistical Significance Testing

- The core of the traditional “frequentist” paradigm of statistics

- Determining what is “probably not not true”
  - Not the same as determining what is true!

Let’s unpack this

In statistical significance testing

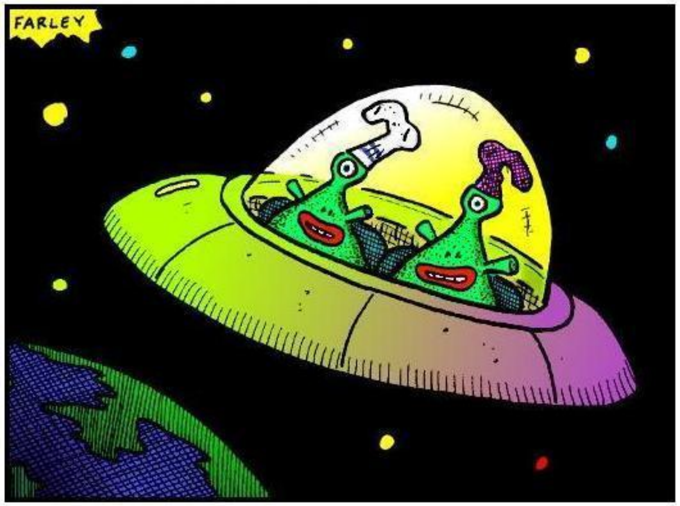
- We start with a hypothesis
  - Curriculum A is better than curriculum B

In statistical significance testing

- We start with a hypothesis
  - All swans are white

In statistical significance testing

- We start with a hypothesis
  - My missing socks are due to aliens



WHAT REALLY HAPPENS
TO YOUR LOST SOCKS

We don’t try to prove that our hypothesis is true

- It’s very difficult to prove something is true

We don’t try to prove that our hypothesis is true

- I looked at 30 swans. They were all white. Therefore, all swans are white.

- Insufficient evidence!

We don’t try to prove that our hypothesis is true

- I looked at 30 swans. They were all white. Therefore, all swans are white.

- Insufficient evidence! (Convenience sample?)

Instead, we try to look for evidence that our hypothesis is false

Instead, we try to look for evidence that our hypothesis is false

- We create what is called a *null hypothesis*

Instead, we try to look for evidence that our hypothesis is false

- We create what is called a *null hypothesis*

- Which basically means that we say “nothing is going on here”

Instead, we try to look for evidence that our hypothesis is false

- We create what is called a *null hypothesis*

- Some swans are not white
- My missing socks are due to some factor other than aliens
- Curriculum A is not better than Curriculum B

And we refer to our original hypothesis as the alternative hypothesis

Example

- Null Hypothesis: Some swans are not white
- Alternative Hypothesis: All swans are white

You Try It

- Null Hypothesis: My missing socks are due to some factor other than aliens
- Alternative Hypothesis:

You Try It

- Null Hypothesis: My missing socks are due to some factor other than aliens
- Alternative Hypothesis: Aliens stole my socks

You Try It

- Null Hypothesis: Curriculum A is not better than Curriculum B
- Alternative Hypothesis:

Usually It’s Thought of as

- Null Hypothesis: Curriculum A is not better than Curriculum B
- Alternative Hypothesis: There is a difference between Curriculum A and Curriculum B

- And we’ll get into why a little later

The Goal

- Find evidence that will help you distinguish between the null hypothesis and the alternative hypothesis

So why…

- Do we turn it around this way?

Again…

- It’s hard to prove something is true
- It’s not as hard to find evidence that there must be something going on

Again…

- It’s hard to prove something is true
- It’s not as hard to find evidence that there must be something going on

- Determining what is “probably not” “not true”

Questions? Comments?

The conceptual structure of a statistical test

- I assume that H_0 is true
- What is the probability that I see the data I see, if H_0 is true?

Not the same

- What is the probability that I see the data I see, if H_0 is true?

- What is the probability that H_0 is true, if I see the data I see?

Example

- If I want to study the difference between two curricula
- I ask the question
- What is the probability that I see the data I see, if there is no difference between curricula?

You try it

- If you want to study whether Japanese high school students are off-task less than American high school students

- What question do you ask?

You try it

- If you want to study whether students who take your curriculum have an average learning gain greater than zero

- What question do you ask?

Questions? Comments?

A statistical test of a hypothesis requires

- A null hypothesis, H_0
- A alternative hypothesis, H_a
- An α value and tailedness

- You then look at the data to compute
  - A p-value

We’ve already discussed the null and alternative hypotheses

- The third part of the test is the alpha and tailedness, which come together to identify the *rejection region*

You may remember α from last class

- α was the parameter we used to define the area outside the confidence interval

- If α = 0.05, 95% CI region is [0.025, 0.975]
- If α = 0.01, 99% CI region is [0.005, 0.995]
- If α = 0.10, 9o% CI region is [0.05, 0.95]

When we are doing a statistical test

- We are looking to see whether our probability is in the α range
- Or in other words, whether p is less than α
- Or in other other words, α is the probability that we will reject the null hypothesis, even when it is true

Remember from Confidence Intervals

- A 95% Confidence Interval means
- That given our data, the true value can be expected to be inside this range 95% of the time
- And outside the range 5% of the time

Analogy

- A 95% Confidence Interval means
- That given our data, the true value can be expected to be inside this range 95% of the time
- And outside the range 5% of the time

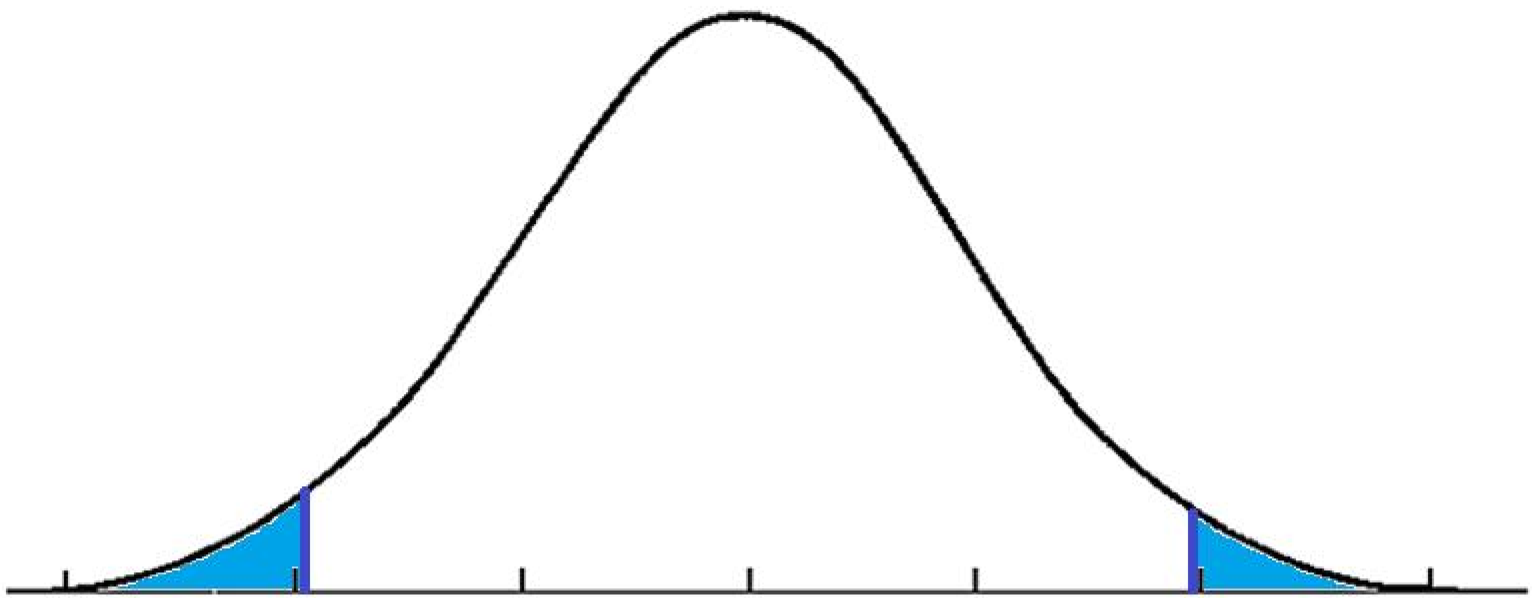
- Similarly, with a statistical test and $\alpha = 0.05$
- We can trust that the null hypothesis is false 95% of the time
- But 5% of the time we may be rejecting the null hypothesis even though it is true

Terminology

- If a statistical test is such that $p < \alpha$
- Then we say the result is statistically significant

Questions? Comments?

Now, for 95% CI, we used α symmetrically



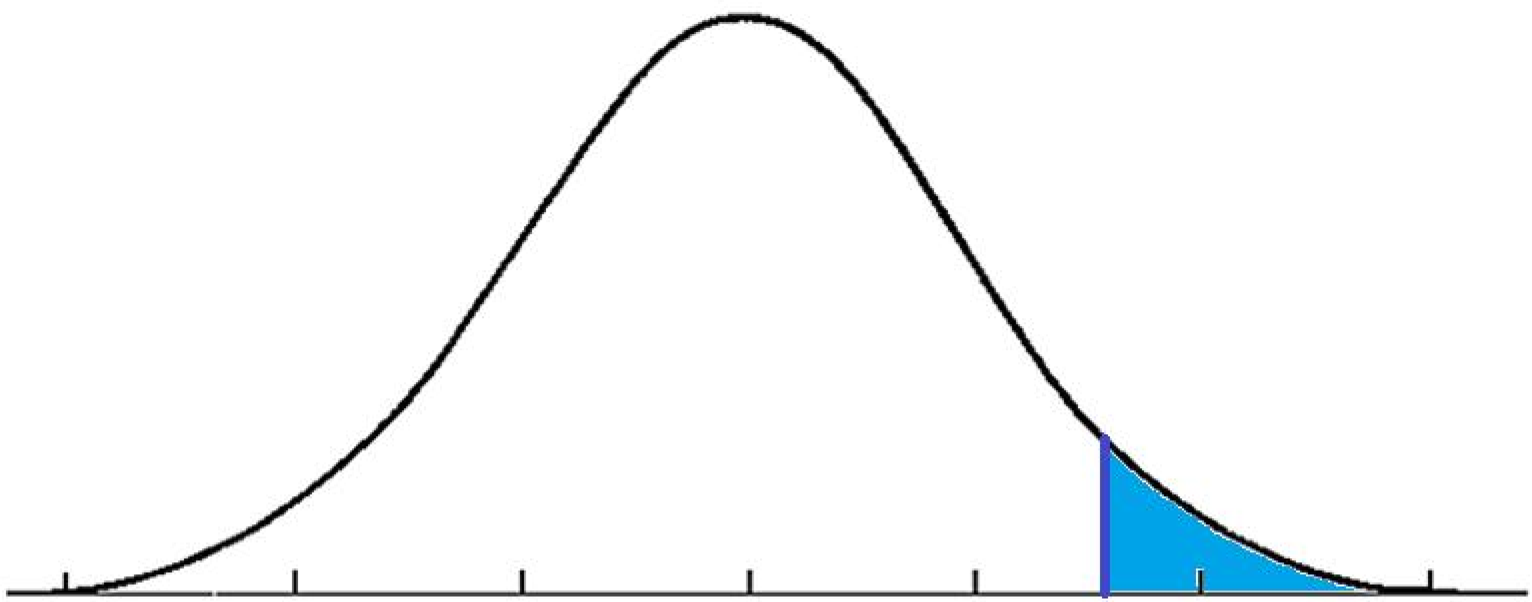
There is another alternative

There is another alternative

- Which I totally, totally, totally don’t recommend

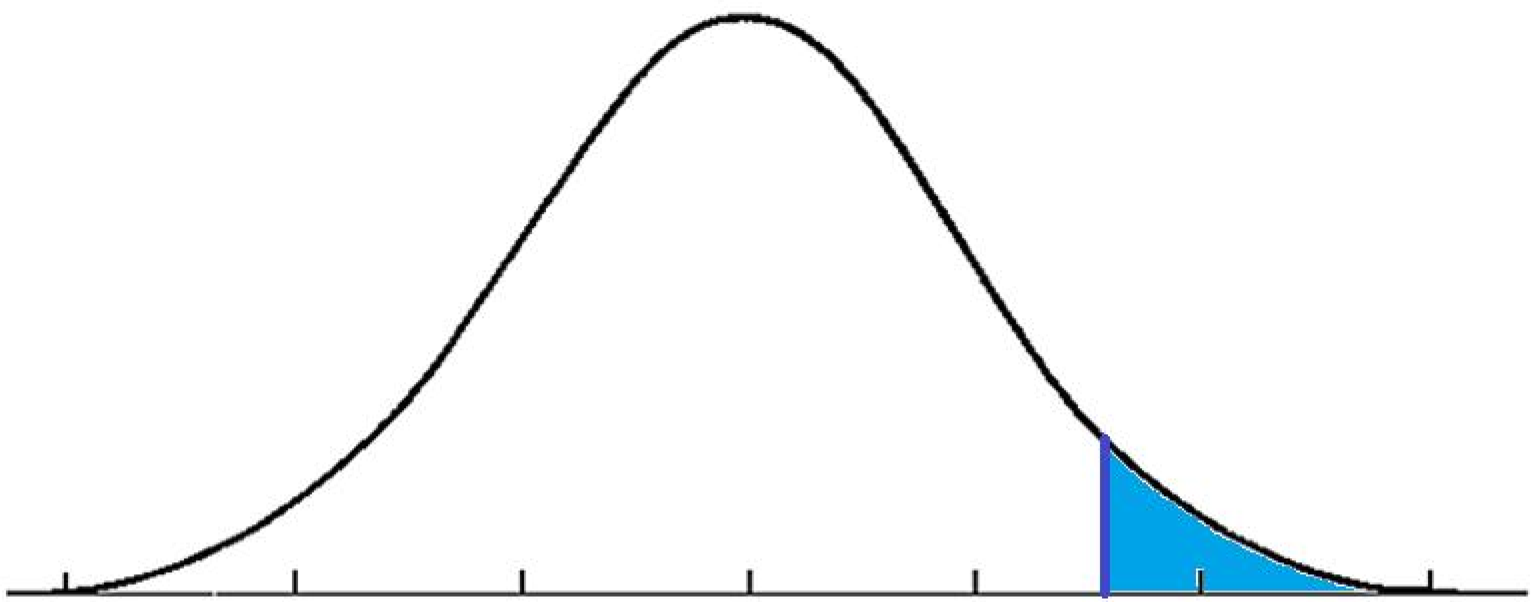
There is another alternative

- Which I totally, totally, totally don’t recommend

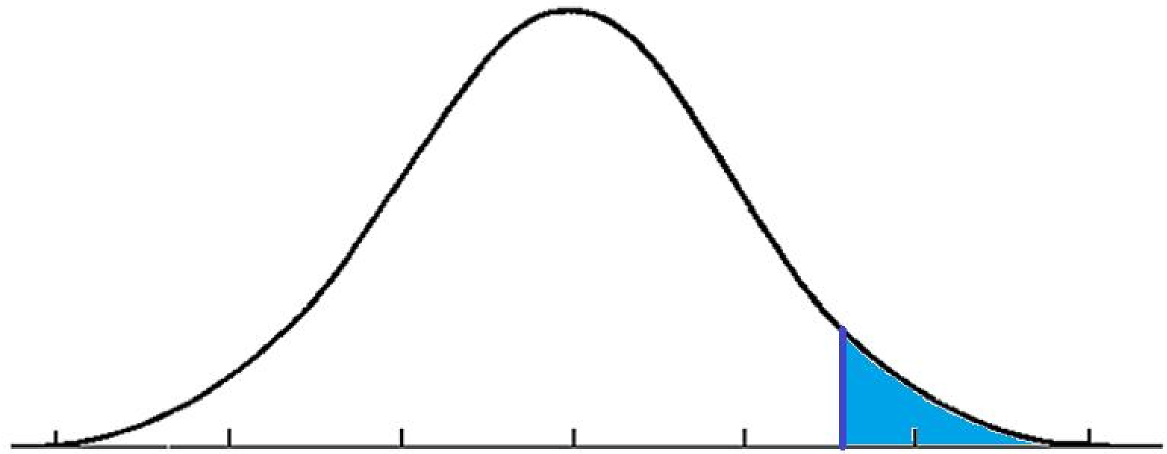


One-tailed test

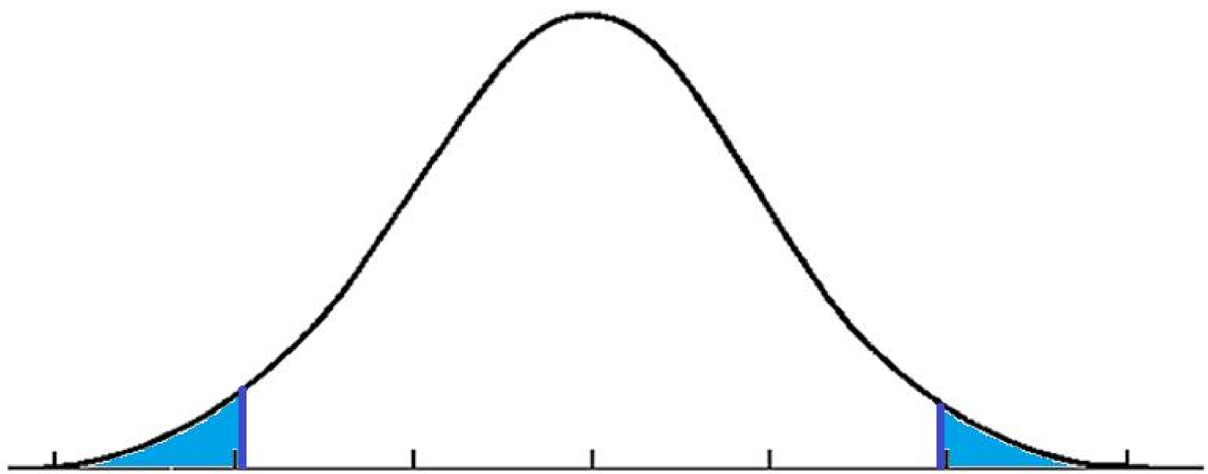
- Which I totally, totally, totally don’t recommend



- One-tailed

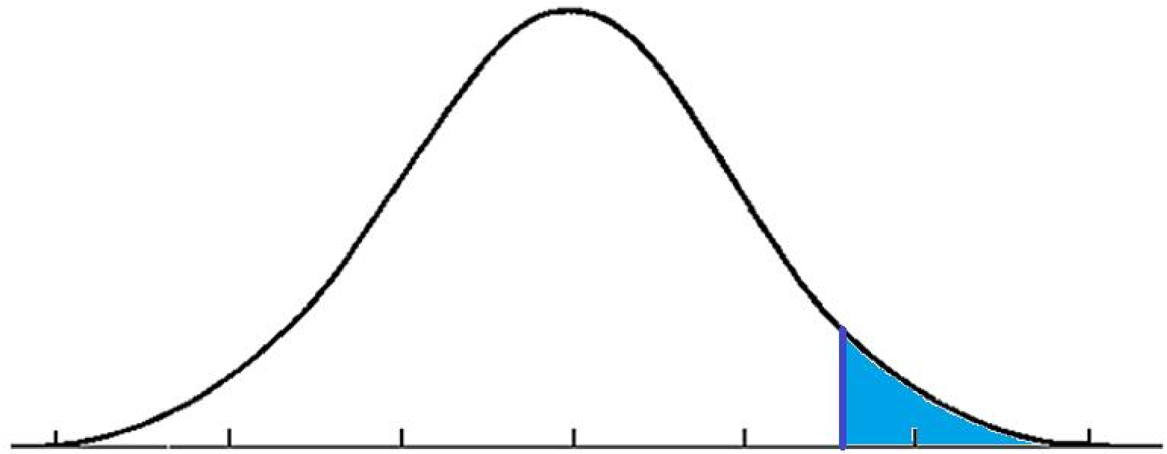


- Two-tailed

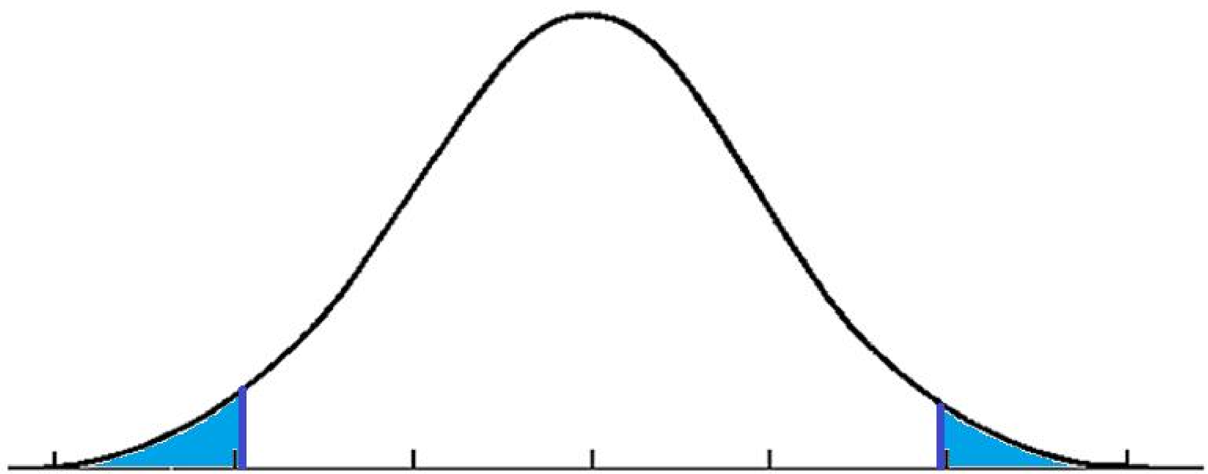


The area in blue is called the "Rejection region"

- One-tailed



- Two-tailed



Rejection region

- If our probability is in the rejection region
- Then the null hypothesis appears to be false
- There is ***something*** going on

Comments? Questions?

You don’t actually have a choice

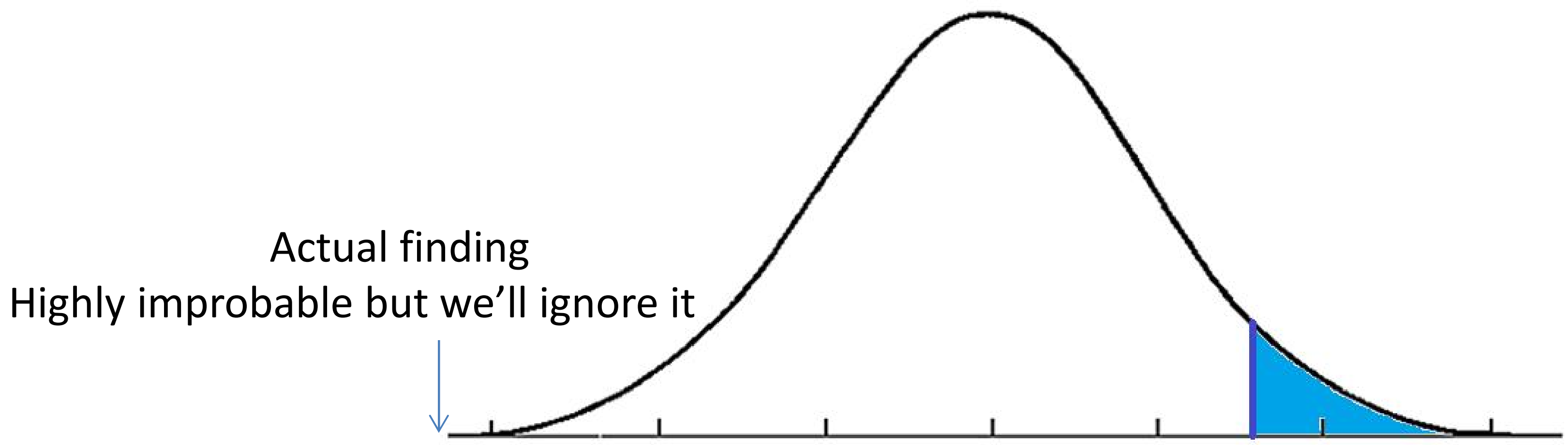
- Despite what textbooks will tell you
- Everyone uses $\alpha = 0.05$
  - Caveat: Sometimes people do refer to *marginal significance,* where they compare probabilities to $\alpha * 2 = 0.10$

- Everyone uses two-tailed tests

Why two-tailed tests?

- Because one-tailed tests have a weird implication

- It commits you to ignoring extreme findings in the unexpected direction



In practice

- Considering marginal significance, where you compare probabilities to $\alpha * 2 = 0.10$
- Is the same level of stringency as doing a one-tailed test where $\alpha = 0.05$

In practice

- Considering marginal significance, where you compare probabilities to α * 2 = 0.10
- Is the same level of stringency as doing a one-tailed test where α = 0.05

- Never ever ever say “a marginally significant one-tailed test”
- Your paper will be rejected

In practice

- Never use one-tailed tests

- Some reviewers are dogmatically opposed to them

Questions? Comments?

A statistical test of a hypothesis requires

- A null hypothesis, H_0
- A alternative hypothesis, H_a
- An α value and tailedness

- You then look at the data to compute
  - **Whether the result is statistically significant**
  - **A p-value**

One-sample Z-test

- A statistical test involving the Z distribution
- Which, yes, means that your sample should have N>30

The test

- *H_0:* The sample mean is no different than some known value
- *H_a:* The sample mean ***is*** different than that known value

- Calculate a Z value for the mean

Significance Criterion

- For a two-tailed test, where $\alpha = 0.05$

- We consider the test significant if

$$Z < -Z_{\propto/2}$$
$$Z > Z_{\propto/2}$$

Significance Criterion

- For a two-tailed test, where $\alpha = 0.05$
- We consider the test significant if

$$Z < -1.96$$
$$Z > 1.96$$

Abstract Example

- $\bar{x}$ is 6, SE is 3

- We want to know if M is greater than 0

Abstract Example

- $\bar{x}$ is 6, SE is 3
- We want to know if M is greater than 0
- Null hypothesis: M is not significantly different than 0

Abstract Example

- $\bar{x}$ is 6, SE is 3
- We want to know if M is greater than 0
- Null hypothesis: M is not significantly different than 0
- $Z = \frac{6}{3} = 2$ **Z > 1.96**
- **So it is significant!**

Concrete Example

- 36 students use a curriculum and take pre and post tests

- The students average a gain of 10 points
- The students get a standard deviation of 12

- Do the students learn from this curriculum?

Hypotheses

- Null hypothesis: The students’ learning gain is not significantly different from 0
- Alternative hypothesis: The students’ learning gain ***is*** significantly different from 0

$$Z = \frac{\bar{x}}{SE} = \frac{10}{\frac{12}{\sqrt{36}}} = \frac{10}{\frac{12}{6}} = \frac{10}{2} = 5$$

- 36 students use a curriculum and take pre and post tests

- The students average a gain of 10 points
- The students get a standard deviation of 12

- Do the students learn from this curriculum?

5> 1.96
It is statistically significant

- 36 students use a curriculum and take pre and post tests

- The students average a gain of 10 points
- The students get a standard deviation of 12

- Do the students learn from this curriculum?

Class Ends

- See next slide deck for continuation

Final questions or comments for the day?

Upcoming Classes

- 4/8 No class
- 4/13 Types of Errors
- 4/15 Statistical power
  - ***HW8 due***