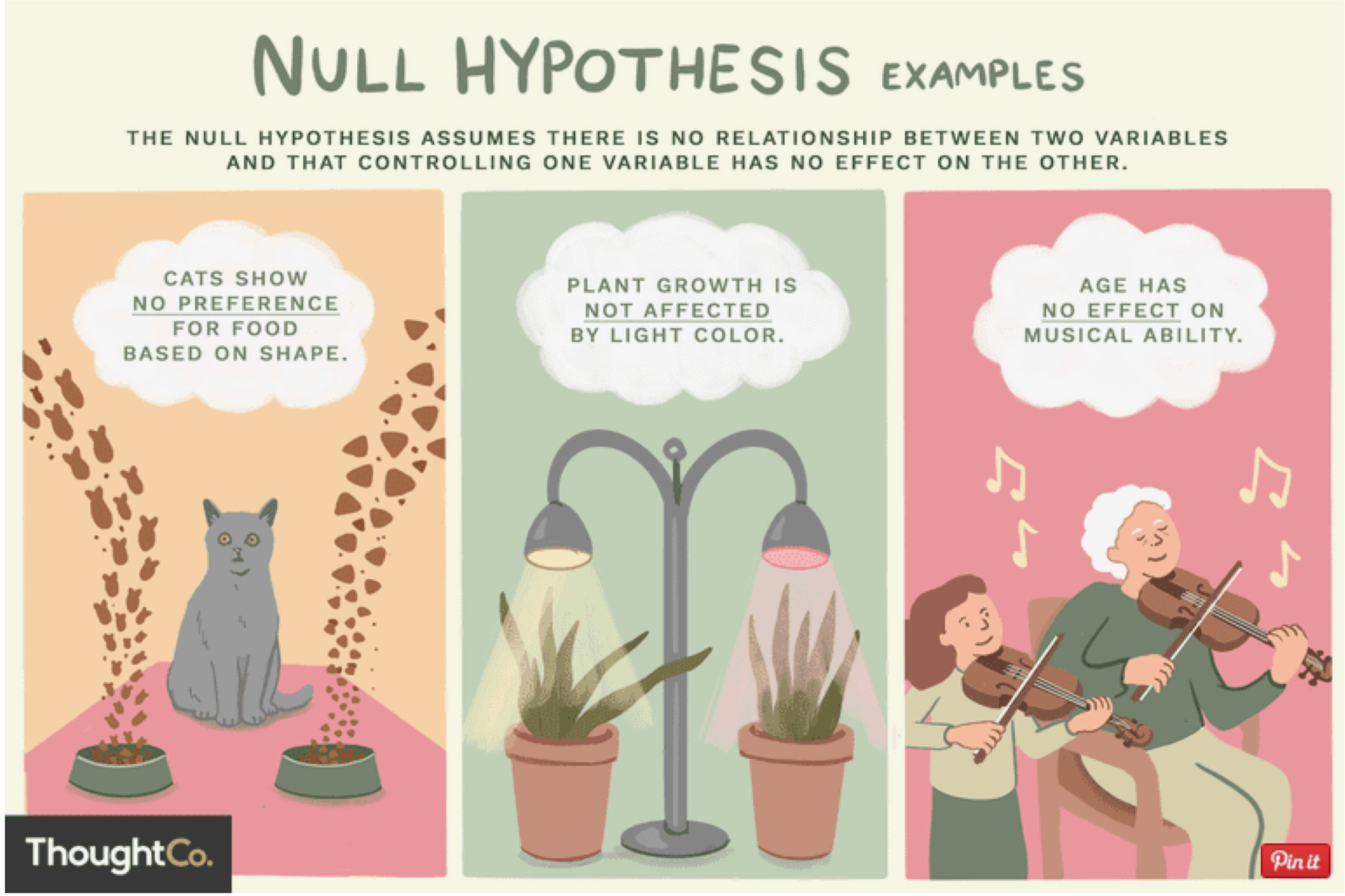
CSCI 3022: Intro to Data Science

Lecture 16: Introduction to Hypothesis Testing

Rachel Cox

Department of Computer Science



ThoughtCo / Hilary Allison

Announcements & Reminders

- Homework 3 Due Mar. 30 at 11:59pm - hard deadline, no submissions after 3/30
- Homework 4 - posted - due April 3
- office Hours: mostly Piazza, Zoom - sign up for 10 minute slots
- Zoom lectures will be Recorded - posted to Canvas

What will we learn today?

- ❑ Hypothesis testing
- ❑ Significance level – how much evidence do you need in order to reject the null hypothesis
- ❑ Rejection regions – if your test statistic falls in here, you have evidence to reject the null hypothesis
- ❑ Type I and Type II errors – false positive and false negatives, respectively

- ❑ *A Modern Introduction to Probability and Statistics, Chapter 25 & 26*

Review from Last Time

Proposition: If X is a normally distributed random variable with mean μ and standard deviation σ, then Z follows a standard normal distribution if we define:

$$Z = \frac{X-\mu}{\sigma} \quad \text{and} \quad X = \sigma Z + \mu$$

A $100 \cdot (1-\alpha)\%$ confidence interval for the mean μ when the value of σ is known is given by:

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

A Thought Experiment

Example: After the introduction of the Euro, Polish mathematicians claimed that the new Belgian 1 Euro coin is not a fair coin. Suppose that I hand you a Belgian 1 Euro coin. How could you decide whether or not it is fair?

Flip coin 100 times, Record # of heads

→ If # heads ≈ 50, you probably would conclude it a fair coin.

Otherwise not.

Statistical Hypotheses

A **statistical hypothesis** is a claim about the value of a parameter of a population characteristic.

Examples:
Suppose the recovery time of a person suffering from a disease is normally distributed with mean μ_1 and standard deviation σ_1.
Hypothesis: $\mu_1 > 10$ days

Suppose that μ_2 is the recovery time of a person suffering from the same disease, but also given some kind of new treatment.
Hypothesis: $\mu_2 < \mu_1$

Suppose that μ_1 is the mean internet speed for Comcast and μ_2 is the mean internet speed for Century Link.
Hypothesis: $\mu_1 \neq \mu_2$

Null vs. Alternative Hypotheses

In any hypothesis-testing problem, there are always two competing hypotheses under consideration:

1) H_0: null hypothesis

2) H_1 or H_A: alternative hypothesis

- ❖ The objective of hypothesis testing is to choose, based on sampled data, between two competing hypotheses about the value of a population parameter.

Classic Jury Analogy

Consider a jury in a criminal trial. When a defendant is accused of a crime, the jury presumes that they are not guilty.

- Null hypothesis: $\underbrace{H_0 = \text{ not guilty}}$

The jury is then presented with evidence. If the evidence seems implausible under the assumption of non-guilt, we might reject the null hypothesis of non-guilt, and claim that the defendant is (likely) guilty.

- Evidence supported the Alternative hypothesis: $\underbrace{H_1 = \text{guilty}}$

Null vs. Alternative Hypotheses

H_0 : status quo

Is there strong evidence for the alternative hypothesis?

- The burden of proof is placed on those that believe the alternative claim.

- The initially-favored claim (H_0) will not be rejected in favor of the alternative claim (H_1) unless the sample evidence provides enough support for the alternative.

Two Possible Conclusions:

1) Reject H_0 (in favor of H_1)
2) Fail to reject the null hypothesis H_0

Null vs. Alternative Hypotheses

Why assume the Null Hypothesis?

Sometimes we don't want to accept a particular assertion unless data can be shown to strongly support it.

Reluctance (cost, time, effort) to change

Example: Suppose a company is considering hiring a new outside advertising company to help generate traffic to their website. Under their current advertising, they get 200,000 hits/day on average. With μ denoting the true average number of hits/day they'd get using the new company's advertising, they would not want to switch companies (because it would be costly) unless evidence strongly suggested that μ exceeds 200,000.

H_0: $\mu = 200{,}000$ hits per day under the new ad company

H_1: $\mu > 200{,}000$

Null vs. Alternative Hypotheses

The alternative to the null hypothesis H_0: $\theta = \theta_0$ will look like one of the following assertions (and variations of these):

1) $\theta > \theta_0$
2) $\theta < \theta_0$
3) $\theta \neq \theta_0$ coin flip

- The equals sign is always in the null hypothesis.
- The alternative hypothesis is the one for which we are seeking statistical evidence.

Test Statistics and Evidence

A **test statistic** is a quantity derived from the sample data and calculated assuming that the null hypothesis is true. It is used in the decision about whether or not to reject the null hypothesis.

Intuition:

We can think of the test statistics as our evidence about the competing hypotheses.

We consider the test statistic under the assumption that the null hypothesis is true by asking questions like:

- How likely would we be to obtain this evidence if the null hypothesis were true?

Test Statistics and Evidence

Example: To determine if the Belgian 1 Euro coin is fair, you flip it 100 times and record the number of heads. What is the test statistic? What are the null and alternative hypotheses?

Test Statistic: the <u>proportion</u> of heads.

let x = # flips that come up heads

$$\hat{p} = \frac{x}{100}$$

What would it take to convince you that the coin is not fair?

$\hat{p} > 60/100$ or $\hat{p} < 40/100$

subjective

Test Statistics and Evidence

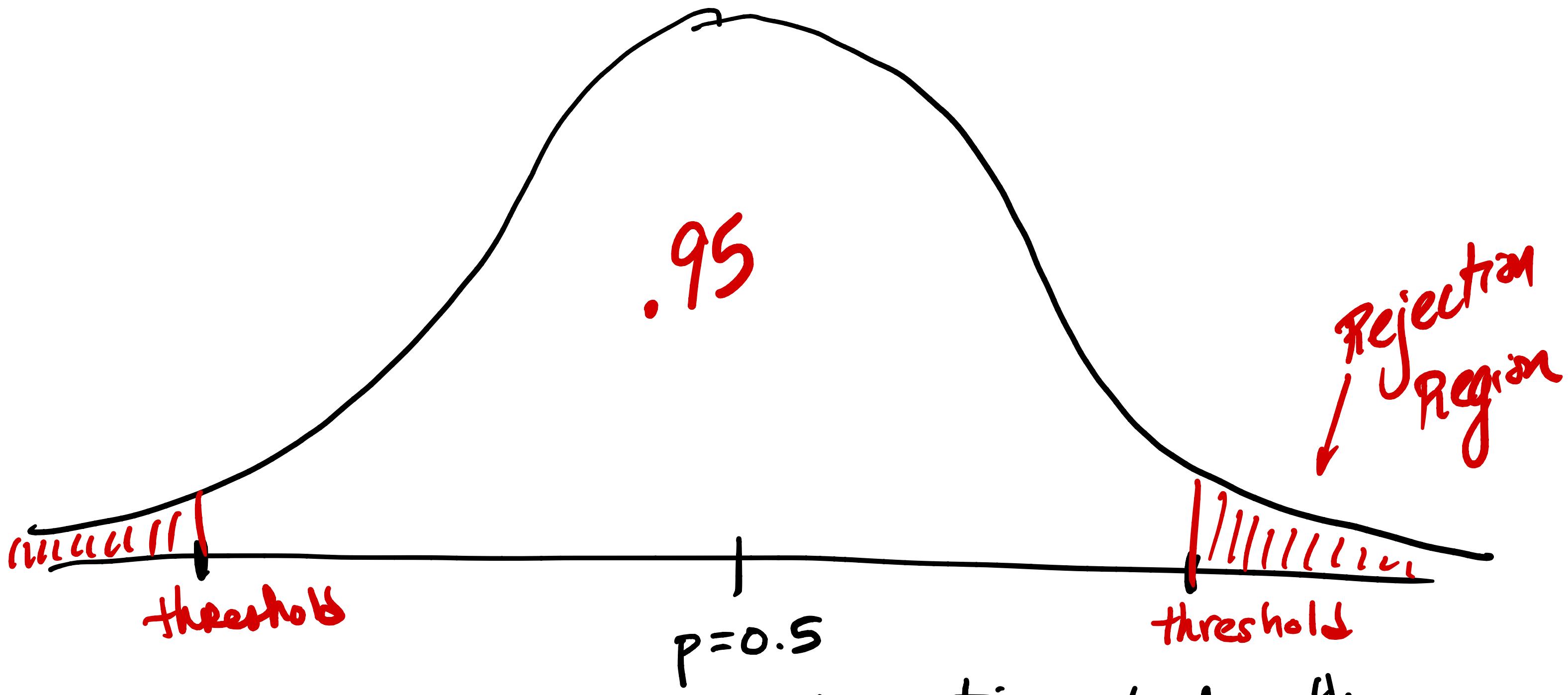
Example: To determine if the Belgian 1 Euro coin is fair, you flip it 100 times and record the number of heads. What is the test statistic? What are the null and alternative hypotheses?

Let's say we want a 95% CI.

Test statistic : $\hat{p} = \frac{x}{100}$

$H_0 : p = 0.5$

$H_1 : p \neq 0.5$



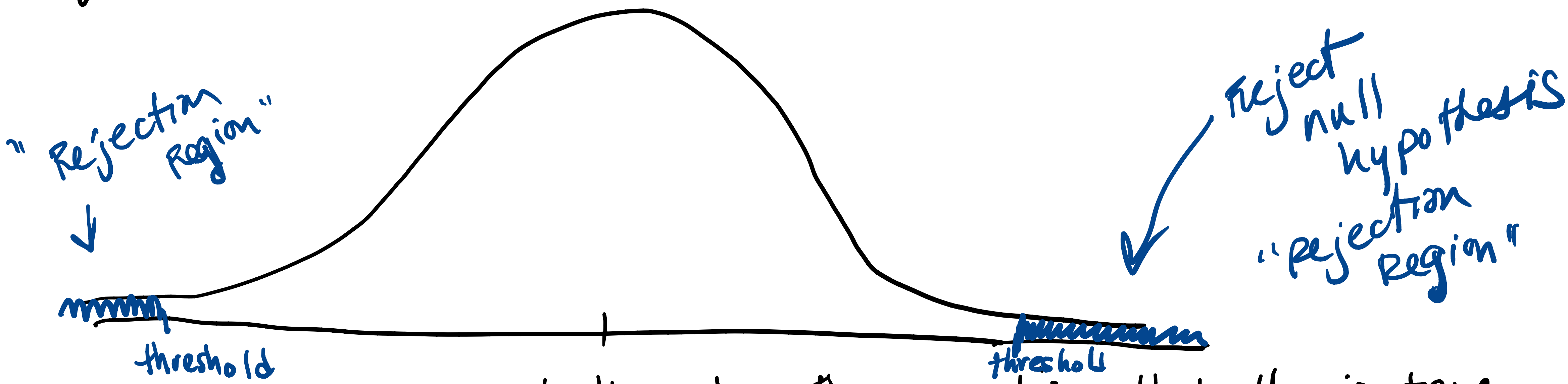
Using the assumption that the Null Hypothesis is True.

Rejection Regions and Significance Level

The **rejection region** is a range of values of the test statistic that would lead you to reject the null hypothesis.

The **significance level** α indicates the largest probability of the test statistic occurring under the null hypothesis that would lead you to reject the null hypothesis.

significance level: set this ahead of time



Curve is built under the assumption that H_0 is true.

Rejection Regions and Significance Level

Example: To determine if the Belgian 1 Euro coin is fair, you flip it 250 times and find that it comes up heads 139 times. Do you reject the null at the 0.1 significance level or not?

Standard procedure:

1. Write down H_0, H_1
2. Determine the test statistic and its "z score", z_{stat} or z_{TS}
3. Compute a threshold value, $z_{critical}$ (based on the significance level α)

1. $H_0: p = 0.5$
   $H_1: p \neq 0.5$

2. $\hat{p} = \frac{139}{250}$

CLT: $N\left(p = 0.5, \sigma^2 = \frac{p(1-p)}{250}\right)$

$$z_{stat} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

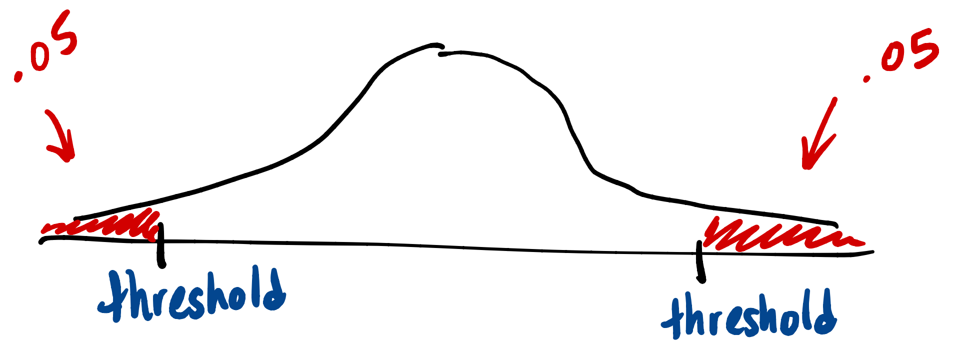
$$= \frac{\frac{139}{250} - 0.5}{\sqrt{\frac{0.5(1-0.5)}{250}}} = 1.771$$

Rejection Regions and Significance Level

Example: To determine if the Belgian 1 Euro coin is fair, you flip it 250 times and find that it comes up heads 139 times. Do you reject the null at the 0.1 significance level or not?

total area in Rejection Region = 0.1

3.



$z_{critical}$ = stats.norm.ppf(.95)
= 1.645

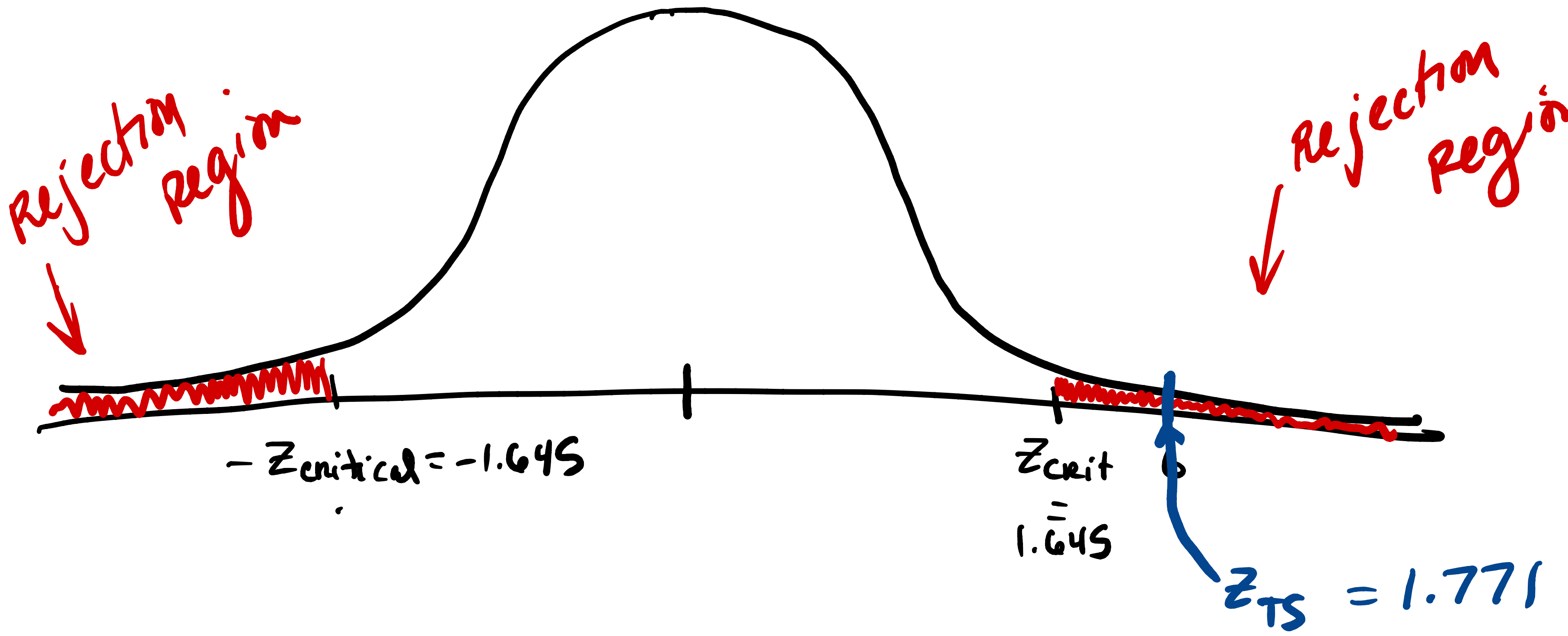
conclusion

Rejection Regions and Significance Level

Example: To determine if the Belgian 1 Euro coin is fair, you flip it 250 times and find that it comes up heads 139 times. Do you reject the null at the 0.1 significance level or not?

What is our conclusion?

$Z_{TS} = 1.771$



Because $Z_{stat} > Z_{crit}$ we reject the null hypothesis

Rejection Regions and Significance Level

Example: To determine if the Belgian 1 Euro coin is fair, you flip it 250 times and find that it comes up heads 139 times. Do you reject the null at the 0.05 significance level or not?

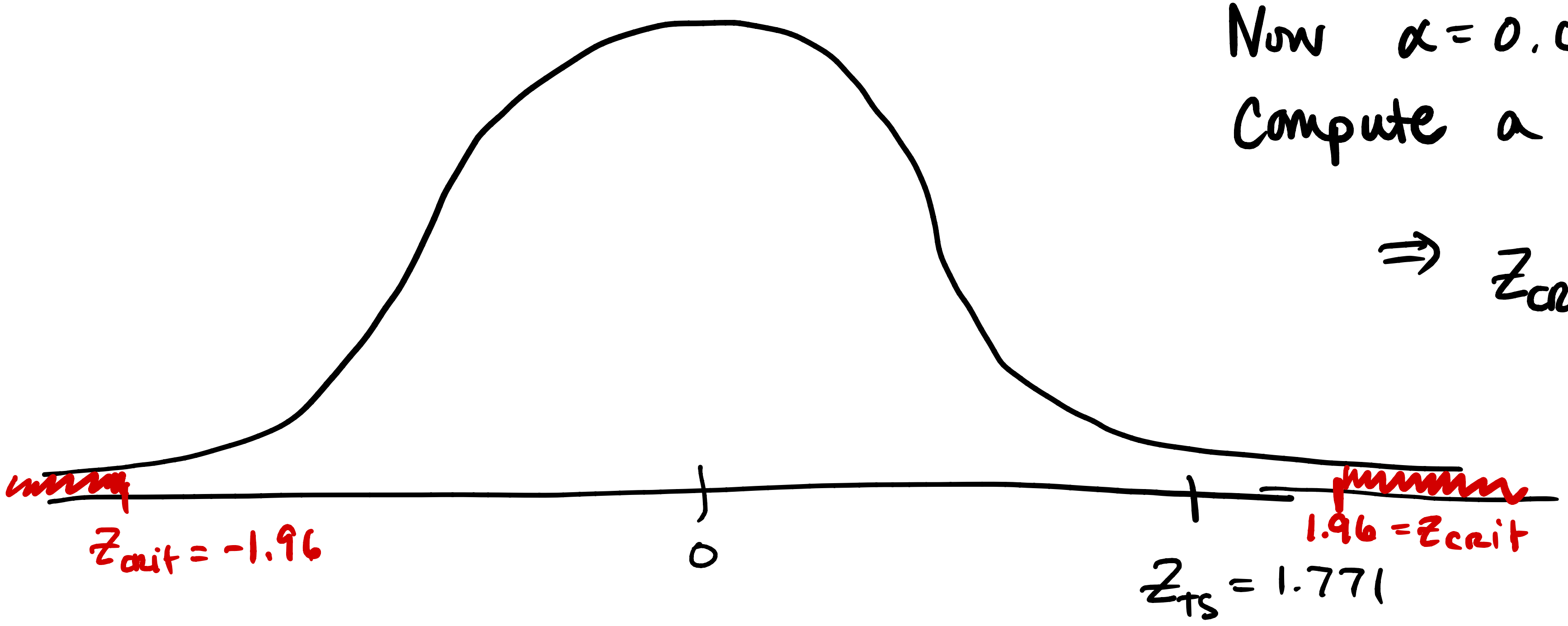
Now $\alpha = 0.05$

Compute a new $z_{critical}$

$\Rightarrow z_{critical} = $ stats.norm.ppf(.975)

$(1 - \alpha/2)$

$= 1.96$



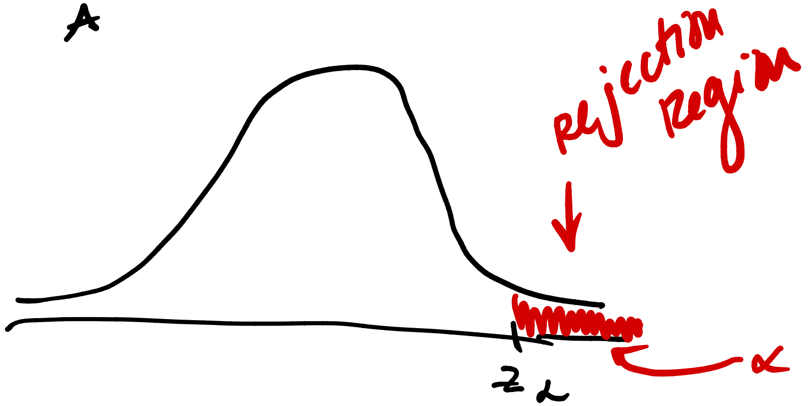
Reject $z_{stat} \geq z_{crit}$

Different Tests for Different Hypotheses

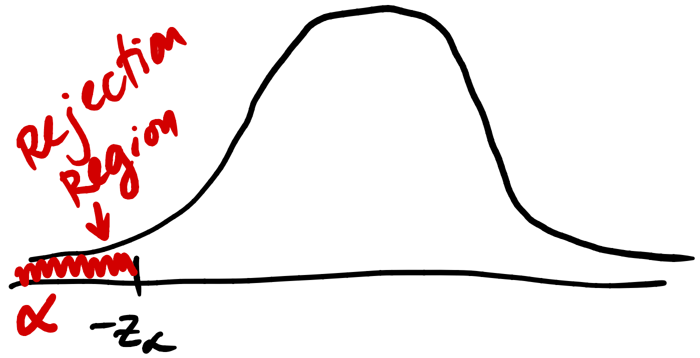
The coin example was an example of a **two-tailed hypothesis test**, because we would have rejected the null hypothesis if the coin had been biased towards heads or towards tails.

	Alternative Hypothesis	Rejection Region for Level α test	
A	$H_1\!: \ \theta > \theta_0$	$z \geq z_\alpha$	one-tailed
B	$H_1\!: \ \theta < \theta_0$	$z \leq -z_\alpha$	one-tailed
C	$H_1\!: \ \theta \neq \theta_0$	$(z \geq z_{\alpha/2})$ or $(z \leq -z_{\alpha/2})$	two-tailed

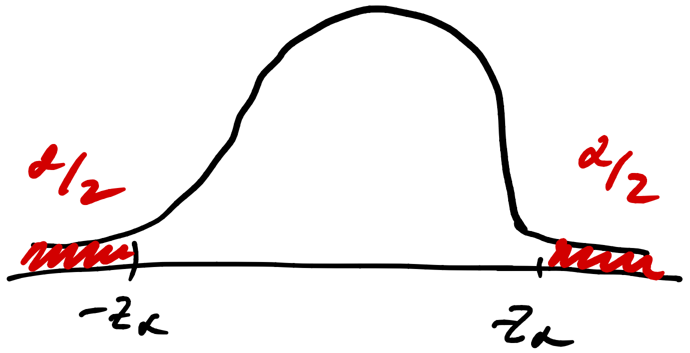
A



B



C



Rejection Regions and Significance Level

Example: Suppose a company is considering hiring a new outside advertising company to help generate traffic to their website. They currently get 200 thousand hits/day, with a standard deviation of 50 thousand hits per day. Suppose they hire the new ad agency for a 30-day trial. During those 30 days, their website gets 210 thousand hits/day. Perform a statistical hypothesis test to determine if the new ad campaign outperforms the old one at the 0.05 significance level.

- $H_0: \mu = 200$
  $H_1: \mu > 200$

This is a one-tailed test.

- Now, look at our test statistic

$$Z_{TS} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} = \frac{210 - 200}{50/\sqrt{30}}$$

$$= 1.095$$

- Z_{crit} = stats.norm.ppf$(1-\alpha)$
  = stats.norm.ppf$(.95)$
  $= 1.645$

Since $Z_{TS} < Z_{crit}$

$1.095 < 1.645$

$\boxed{\Rightarrow \text{Fail to reject the null hypothesis}}$

Rejection Regions and Significance Level

What assumptions did we make in the previous example?

1) Assumed that the CLT would hold -- n = 30 samples

2) Assumed that we can represent the distributions involved as Normal.

Type I & Type II Errors

- Type I Error – Rejecting a null hypothesis when it is actually true.
- Type II Error – Accepting a null hypothesis when it is actually false.

α - significance

α = P(making a type I error)

Next Time:

- P-values