Hypothesis Testing, Part 2

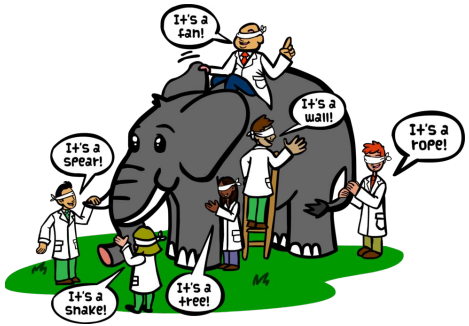
Lecture 8

Emma Ning, M.A.

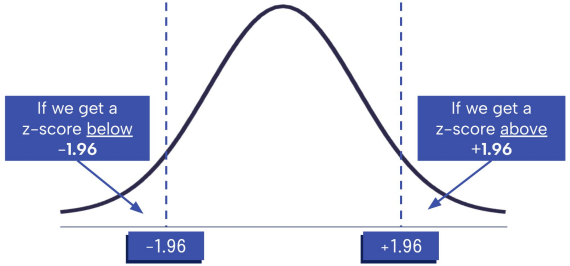
From our last lecture...

- Z-Score for Sample Mean $$Z = \frac{M - \mu}{\sigma_M}$$

- Null Hypothesis Significance Testing (NHST)/Hypothesis Testing

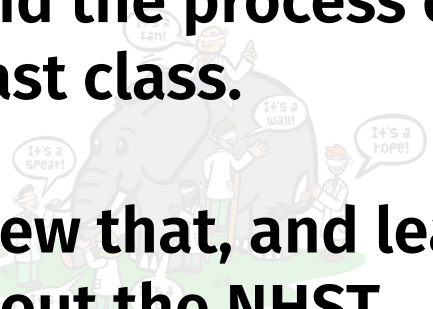


- Steps of NHST & z-Test



From our last lecture...

- Z-Score for Sample Mean $Z = \frac{M - \mu}{\sigma_M}$
- Null Hypothesis Significance Testing (NHST)/Hypothesis Testing
- Steps of NHST & z-Test



You got a general idea and the process of NHST from the last class.

Today we are going to review that, and learn a few more details about the NHST.

TODAY'S PLAN

01

Null & Alternative Hypotheses

02

Alpha Level & the *p*-Value

03

Z-Test Worked Example

04

Exam 1 Information

Learning objectives

- **Articulate** the null and alternative hypotheses for a given research design
- **Differentiate** between a one- and two-tailed hypothesis/test
- Define an alpha level and **explain** the factors that can influence the size and location of the critical region
- **Conduct a z-test** following the NHST steps and **interpret** the results in context of a specific research question

Null & Alternative Hypotheses

Recap: Why NHST, and why 2 hypotheses?

We have important research questions to answer, like whether a new drug improves symptoms.

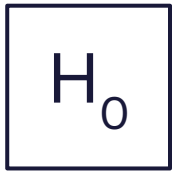
We can't afford to study the entire population. So we collect a **sample**, hoping it gives us a good estimate of what's happening in the population.

We can't just say "our drugs work". Regulators will ask: "Compared to what?", "**How do we know this isn't just random** — like someone feeling better because they got more sleep?"

Therefore, we need to test whether our drug works using **statistics** (based on probability). Just like how we use control groups to ask "**compared to what**?", we use two hypotheses so we can c**ompare our results to what we'd expect by chance alone**.

The Alternative and Null Hypothesis

I called this H_a last class, but they are the same!



Alternative Hypothesis

The **alternative** hypothesis (H_1) states an effect, difference, or relationship exists.

(usually what we *expect* to happen in our study)

Null Hypothesis

The **null** hypothesis (H_0) states that there is no relationship or difference.

This is our hypothesis of "**no effect**"

The Null Hypothesis

Typically, when we state the **null hypothesis** in words, we say something along the lines of...

1. There is **no difference** between...
2. There was **no effect** of ....
3. There was **no change** in ...

The Alternative Hypothesis

H_1 sometimes also called H_a

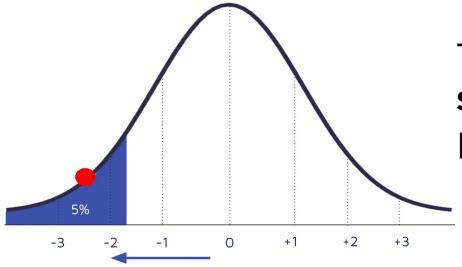
Directional (one-tailed):
We expect the sample mean to be **greater** or **smaller** than the population mean.

Nondirectional (two-tailed):
We expect there to be a difference between the sample and population mean.
(more common)

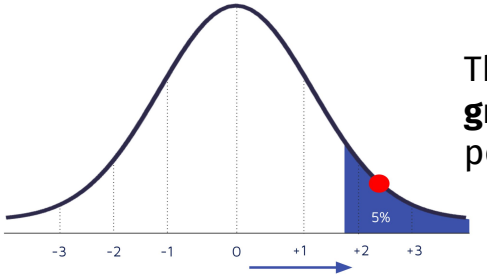
The Alternative Hypothesis

Directional (one-tailed):

We expect the sample mean to be **greater** or **smaller** than the population mean.



The sample mean is **smaller** than the population mean.

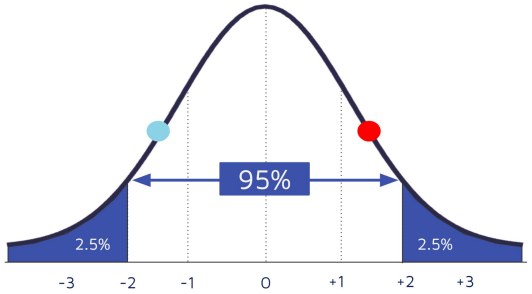


The sample mean is **greater** than the population mean.

Nondirectional (two-tailed):

We expect there to be a difference between the sample and population mean.

(more common)



We are not sure which direction the difference is – red dot or blue dot.

DISCUSS WITH TABLE

At your table, discuss the two studies and **articulate your alternative (research) and null hypotheses**. Is it **one or two-tailed**?

STUDY 1

Clinical psychologists are examining whether a 6-week cognitive-behavioral therapy (CBT) program impacts self-reported anxiety symptoms in adults diagnosed with generalized anxiety disorder.

STUDY 2

A psychologist is interested in whether college students who spend more than 3 hours per day on social media have higher anxiety levels than the general college population.

Alpha Level & the p-Value

The Basic Steps of NHST

✅ **01** Restate your research question as **hypotheses**

02 Decide what cutoff score is "**extreme**" or "**significant**"

03 Calculate some **statistics** (e.g., z-score)

04 **Make a final decision** about the null hypothesis

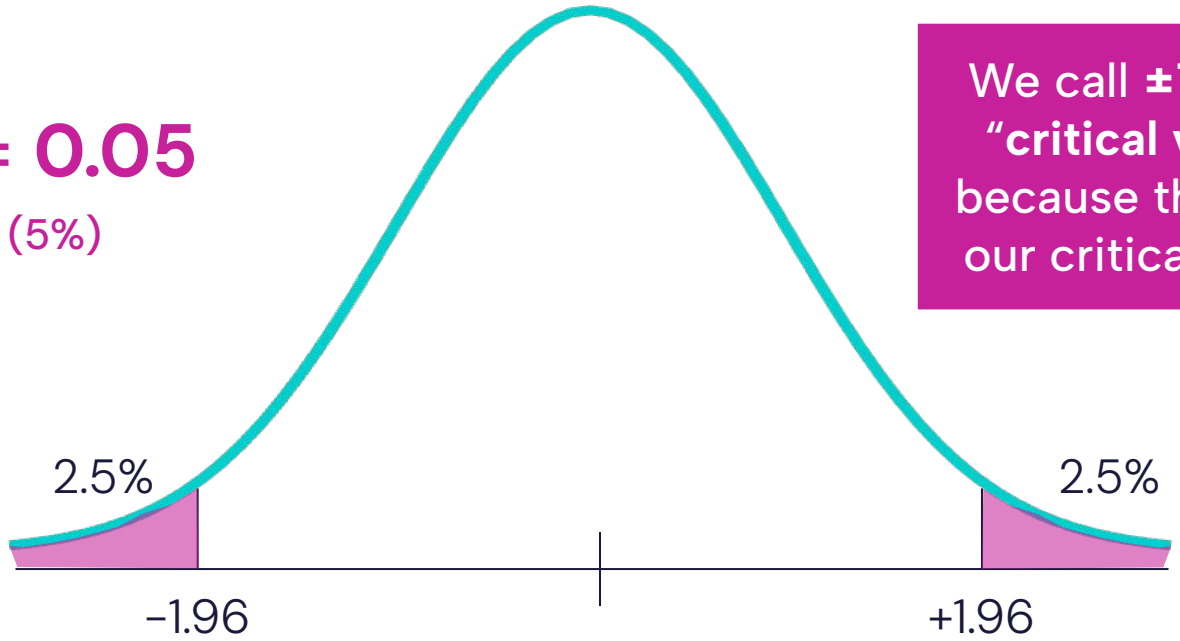
Alpha Level (α)

"Significance level"

The alpha level (α) determines our **"critical region"** for what we expect to be an **extreme** (unlikely due to chance) score.

α = 0.05

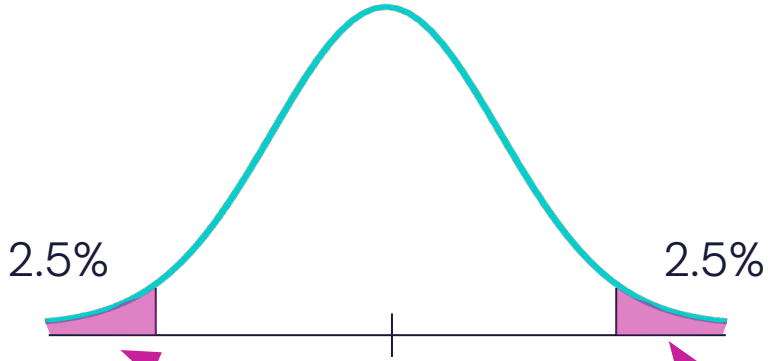
(5%)



We call **±1.96** our **"critical values"** because they start our critical region

Where do the critical values come from?

z	Body (B)	Tail (C)
1.94	.9738	.0262
1.95	.9744	.0256
1.96	.9750	.0250
1.97	.9756	.0244
1.98	.9761	.0239

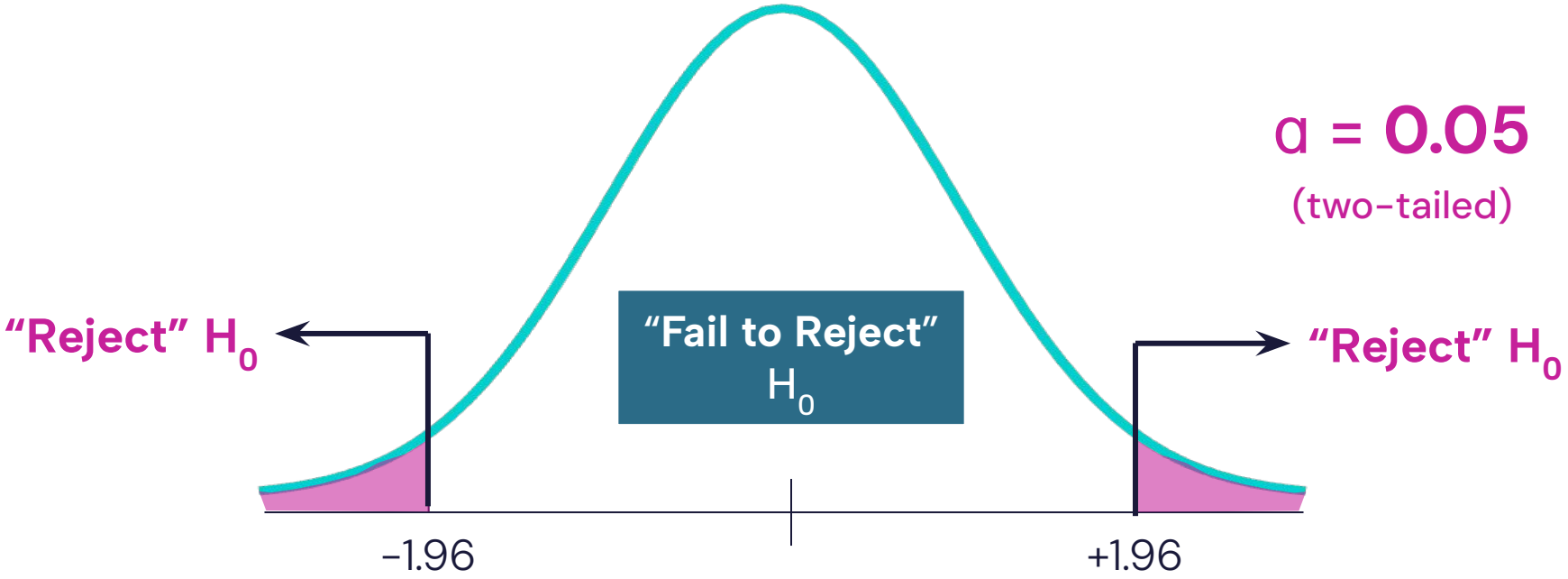


This is why we use **±1.96** as our cutoff score for a <u>two</u>-tailed test with $\alpha = 0.05$.

z	Body (B)	Tail (C)
1.94	.9738	.0262
1.95	.9744	.0256
1.96	.9750	.0250
1.97	.9756	.0244
1.98	.9761	.0239

Region of "Rejection"

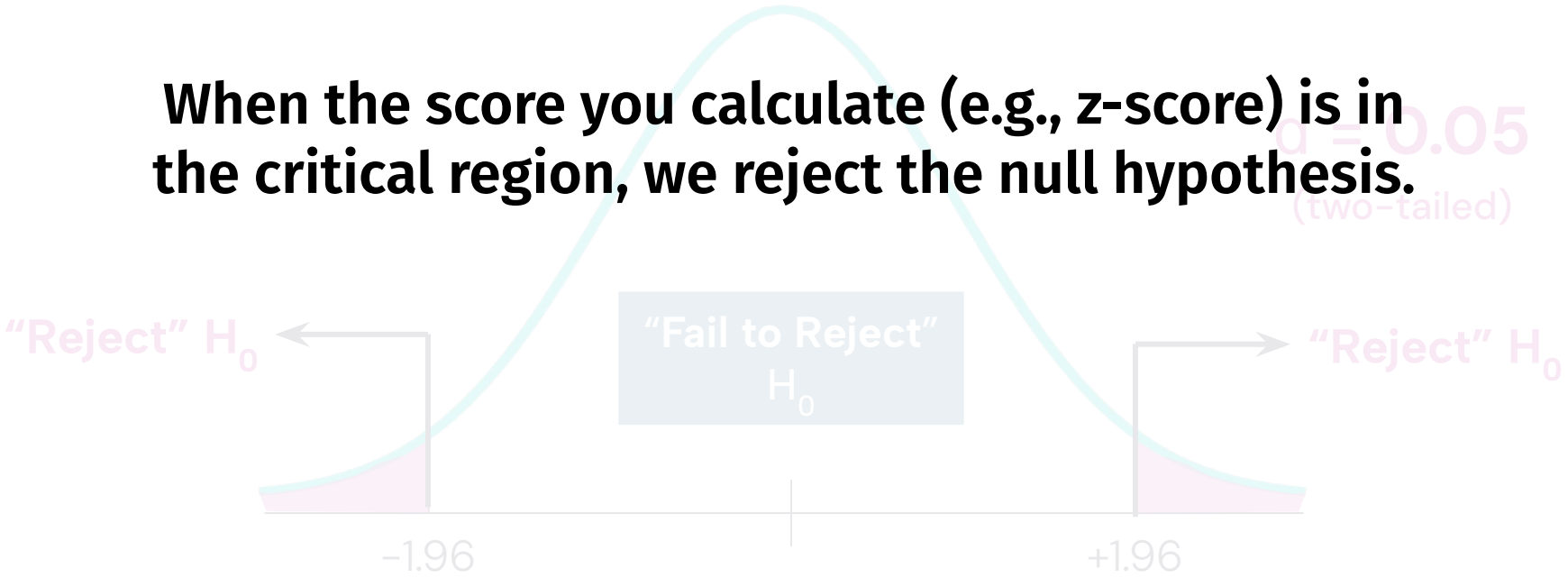
Another name for the critical region is the **region of "rejection"** because we reject the null hypothesis if our test statistic falls within it.



Region of "Rejection"

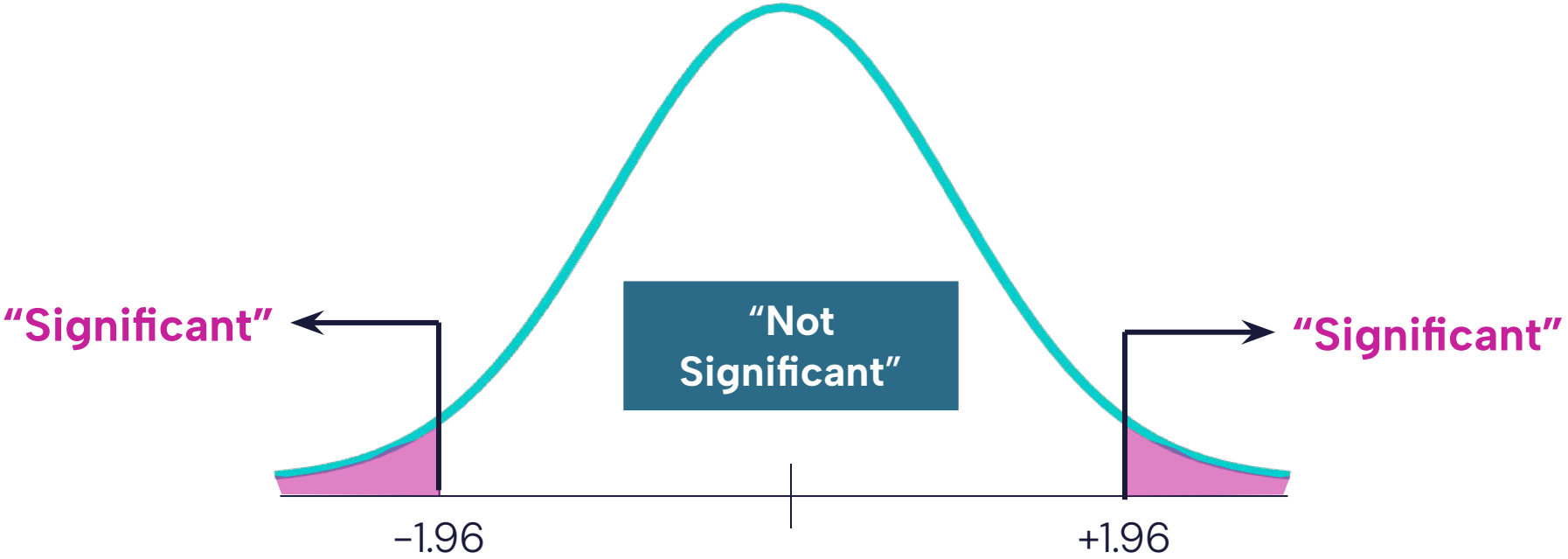
Another name for the critical region is the **region of "rejection"** because we reject the null hypothesis if our test statistic falls within it.

When the score you calculate (e.g., z-score) is in the critical region, we reject the null hypothesis.



Region of "Significance"

Our α is often called our **significance level**. If our test statistic falls within our critical region, we conclude that there is a **significant effect/difference/relationship**.



Region of “Significance”

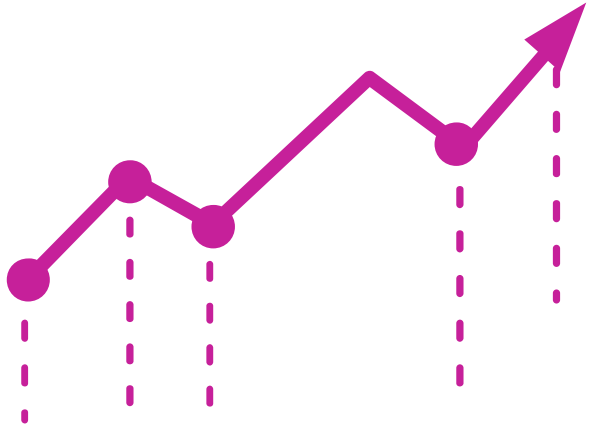
Our α is often called our **significance level**. If our test statistic falls within our critical region, we conclude that there is a **significant effect/difference/relationship**.

In psychology, α = 0.05 is the most commonly used threshold for significance.
It’s a widely accepted standard, so unless otherwise stated, we usually assume this is the cutoff being used.

A Note on “Significance”

When we say our sample mean is **statistically significant**, we are saying that the result is so *unusual* that it is only found in the outermost 5% of the sample means*.

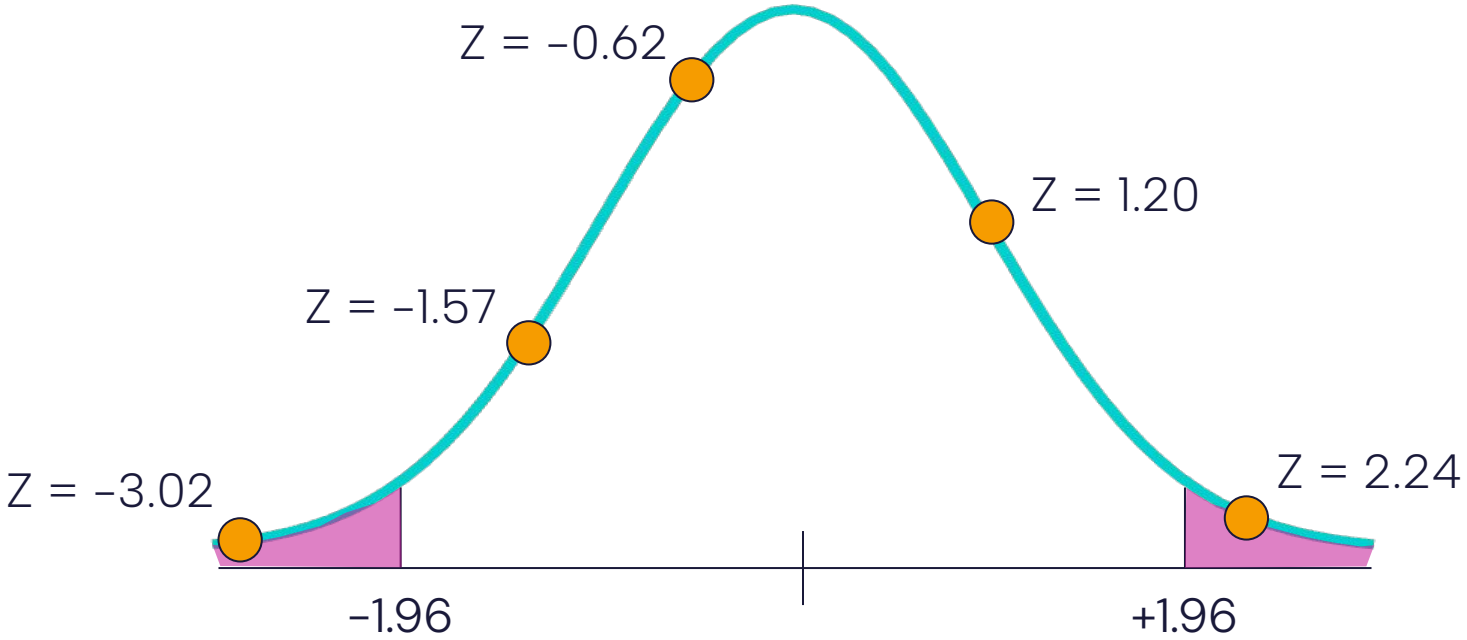
It is not a statement of importance or how “big” a difference is. It is simply stating that the **observed result is unlikely due to chance**.



*assuming $\alpha = 0.05$

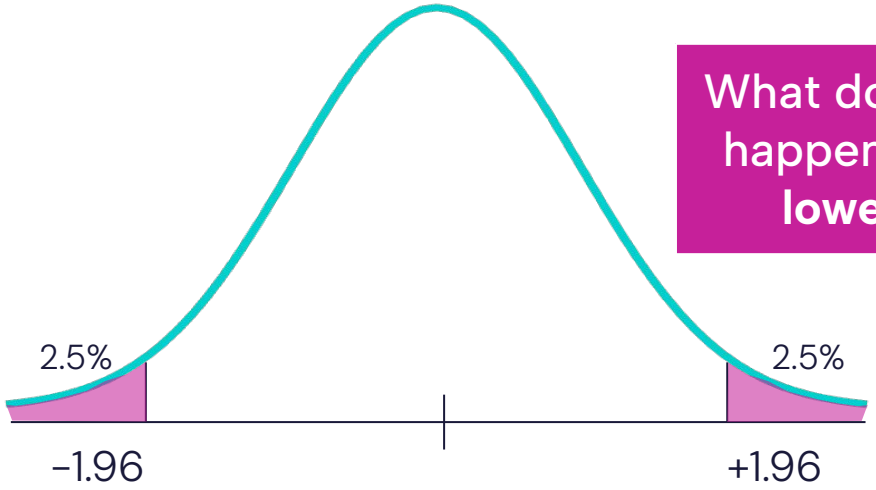
Practice

What would we **conclude** for each z-score below?



Alpha Level (α) Changes Our Critical Region

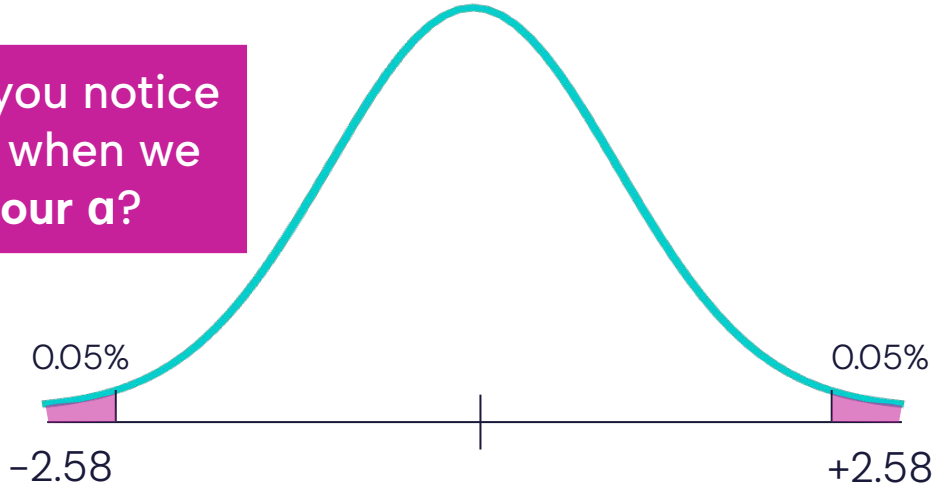
In psychology, we often use an alpha level of **5% (0.05)**, but sometimes we want to use a lower alpha, such as 1% (0.01).



α = **0.05**

(5%)

What do you notice happens when we **lower our α?**

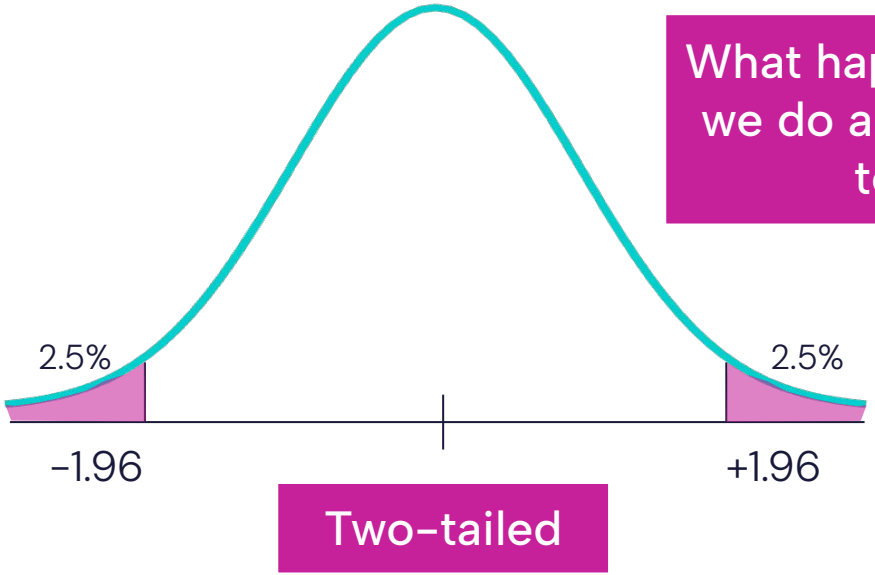


α = **0.01**

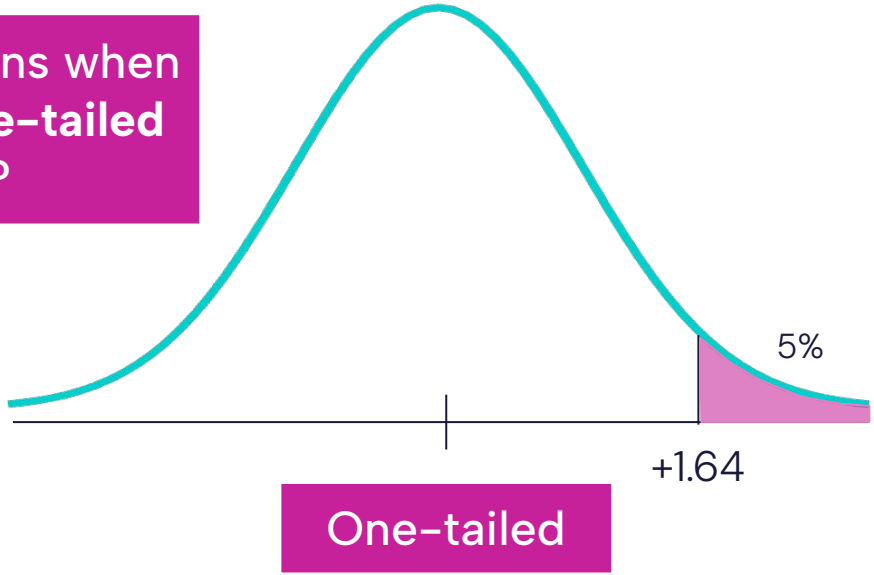
(1%)

How do one vs. two-tailed hypotheses impact our critical region?

If we use a one-tailed test, our critical region (5%) will only be on one side of the distribution. This changes our critical value.



What happens when we do a **one-tailed** test?



The Basic Steps of NHST

✅ **01** Restate your research question as **hypotheses**

✅ **02** Decide what cutoff score is "**extreme**" or "**significant**"

03 Calculate some **statistics** (e.g., z-score)

Let's say we calculated our z-value, and checked to see if it's more extreme than t-crit already

04 **Make a final decision** about the null hypothesis

The Basic Steps of NHST

✅ **01** Restate your research question as **hypotheses**

✅ **02** Decide what cutoff score is "**extreme**" or "**significant**"

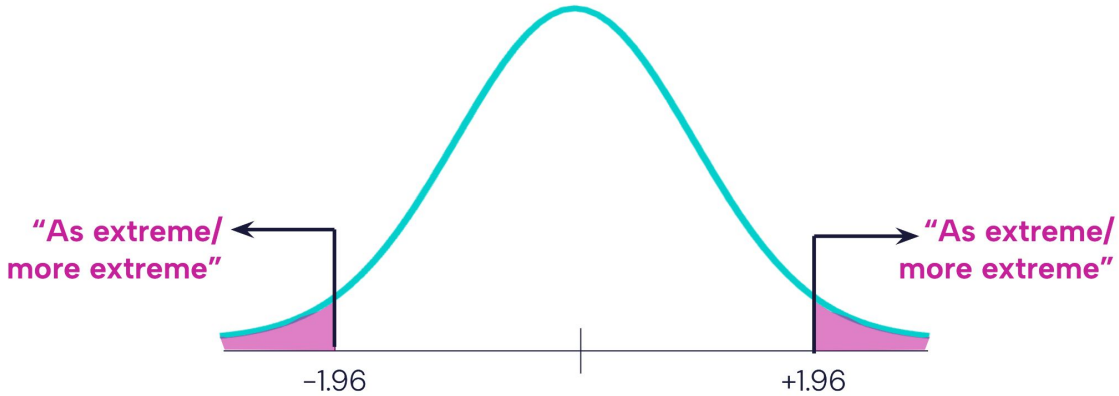
✅ **03** Calculate some **statistics** (e.g., z-score)

04 **Make a final decision** about the null hypothesis

The "p-value"

A p-value measures the probability of obtaining a result as extreme as, or more extreme than, our observed result, assuming the null hypothesis is true.

In other words, if our sample p-value is below α, then we question the null hypothesis.



Decisions regarding our hypotheses

$p < \alpha$

Reject the null hypothesis
(we found support for our hypothesis)

$p \geq \alpha$

Fail to reject the null hypothesis
(we did not find support for our hypothesis)

Notice how we didn't say "we prove the null hypothesis"?

Well, since null represents "no effect", we can never prove that.

Imagine you are searching for a spider in your room. You shine your flashlight around for 10 minutes and don't see anything.

You didn't see it, but you can't prove it's not there.

Z-Test Worked Example

Worked Example

A clinical psychologist is investigating whether college students who use social media more than 3 hours per day have **different** anxiety levels than the general college student population. The population mean anxiety score, based on national data, is **40**, with a known standard deviation of **6**. The psychologist selects a sample of **36** college students who report using social media for more than 3 hours per day and finds that their mean anxiety score is **43**.

Conduct a z-test (use $\alpha = 0.05$).

Step 0: Annotate Your Problem (recommended)

A clinical psychologist is investigating whether college students who use social media more than 3 hours per day have **different** anxiety levels than the general college student population. The population **mean** anxiety score, based on national data, is **40**, with a known **standard deviation** of **6**. The psychologist selects a **sample** of **36** college students who report using social media for more than 3 hours per day and finds that their **mean** anxiety score is **43**.

Conduct a **z-test** (use $\alpha = 0.05$).

$\mu = 40$

$\sigma = 6$

$M = 43$

$n = 36$

Step 1: State Hypotheses (Null & Alternative)

State both of your hypotheses in words. You can also write it mathematically.

- **Null Hypothesis (H_0)**: the average anxiety of college students using social media for more than 3 hours per day is **equal to** the population mean.

$$\mu = 40$$

Step 1: State Hypotheses (Null & Alternative)

State both of your hypotheses in words. You can also write it mathematically.

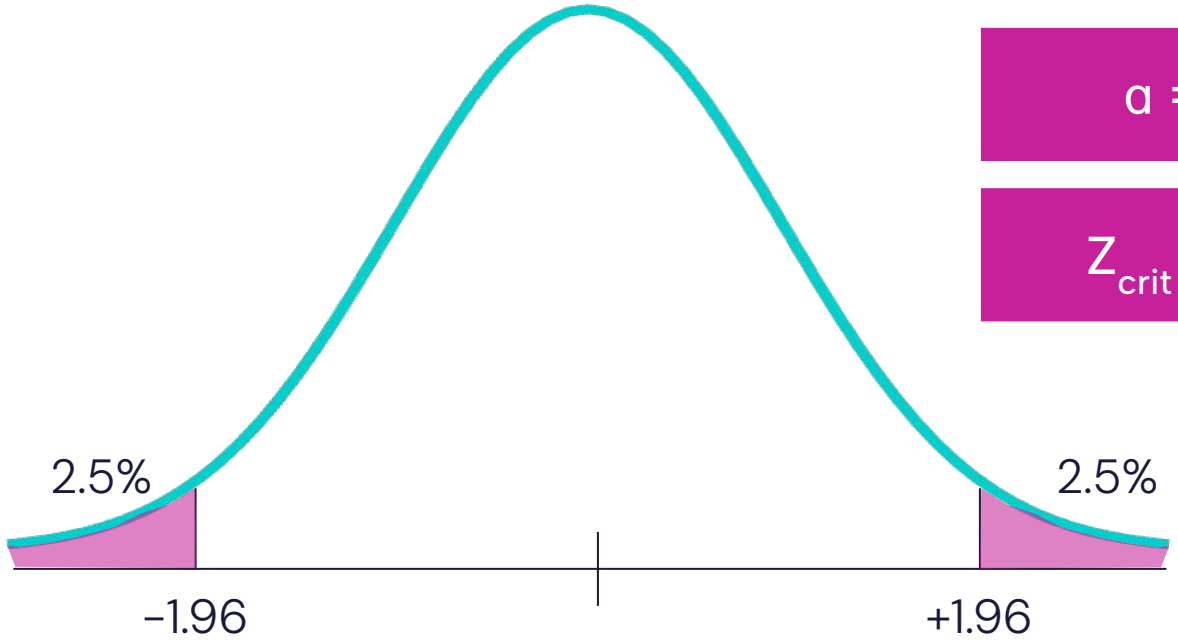
- **Null Hypothesis (H_0)**: the average anxiety of college students using social media for more than 3 hours per day is **equal to** the population mean.

$$\mu = 40$$

- **Alternative Hypothesis (H_1)**: the average anxiety of college students using social media for more than 3 hours per day is **different than** the population mean

$$\mu \neq 40$$

Step 2: Determine Cutoff Values (draw critical region)



$\alpha = 0.05$

$Z_{crit} = \pm 1.96$

$\mu = 40$ $\sigma = 6$

$M = 43$ $n = 36$

Step 3: Calculate your test statistic

We will assume the following for your group:

- First, calculate the **standard error** (σ_M) using the formula.

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{36}} = 1.00$$

$\mu = 40$ | $\sigma = 6$ | $M = 43$ | $n = 36$

Step 3: Calculate your test statistic

We will assume the following for your group:

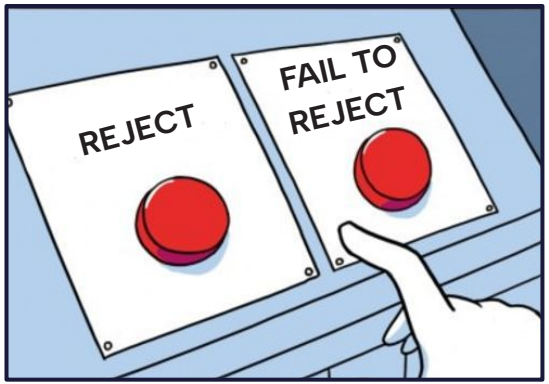
- First, calculate the **standard error** (σ_M) using the formula.

$$\sigma_M = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{36}} = 1.00$$

- Now, calculate the **z-score** (z) using the formula.

$$Z = \frac{M - \mu}{\sigma_M} = \frac{43 - 40}{1.00} = +3.00$$

Step 4: Make a decision about the null



Our z-score is <u>inside</u> the critical region, which means we **<u>reject the null hypothesis</u>** ($p < 0.05$).

Step 5: **Interpret results** (in basic APA Style)

A one-sample z-test showed that students (n = 36) who use social media had significantly **higher anxiety** than the national average, **z = 3, p < .05.**

Please include an interpretation like this every time you do a NHST/z-test! It interprets your finding in plain English.

ICA 8: Teach the class! Let's do the ICA together

Researchers test whether a new ADHD medication changes hyperactivity in children.

- **H_0:** Children taking the medication will have the same hyperactivity as the general population ($\mu = 30$, $\sigma = 10$).
- **H_1:** Children taking the medication will differ in hyperactivity from the population ($\mu \neq 30$).

They test a sample of n = 40 children who are given the medication. Their mean hyperactivity score is M = 37. Use a two-tailed test, $\alpha = 0.05$.

What's your decision given this data?

Exam 1 Information

SOME EXAM 1 KEY TOPICS

Make sure you know these concepts well!

Levels of Measurement

Distribution Shape & Modality

Mean

Median

Bar chart, box plot, histogram

Variability (SS, variance, SD)

Z-Scores, Meaning/Interpretation

Standard Error, Meaning/Interpretation

Reading the Z-table

Sampling Distribution

Central Limit Theorem

Hypothesis Testing Procedure

Identify Null & Alternative Hypothesis

Conduct Z-test, look up z critical value & p-value

Make Decision, Interpretation & APA write-up

Note: see **learning objectives** for each lecture as a "study guide"

EXAM 1 INFORMATION & FORMAT

- Exam will cover content on **lectures 1 – 8**
- Exam 1 will occur in class and will last **75 minutes**
- You are allowed **one double-sided hand-written page of notes**
- It will be mostly **short-answer questions**
- You will be given a **formula sheet** and **z-table**
- You should ***show all your work,*** including formulas
- Round to **two decimal places** (e.g., 10.21, 120.43)

Bring your calculator!!!