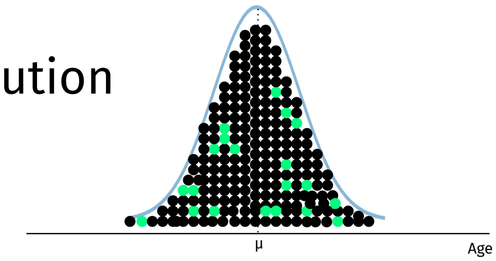
Sampling & Inference, Introduction to Hypothesis Testing

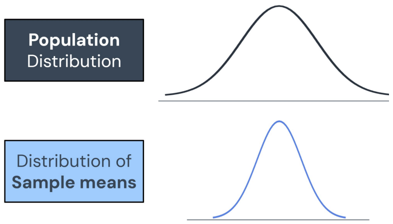
Lecture 7

Emma Ning, M.A.

From our last lecture...

- Sampling distribution

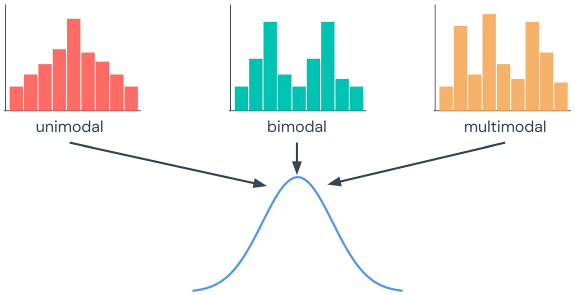




- Standard error
(Standard deviation of the sampling distribution)

$$SE = \frac{\sigma}{\sqrt{n}}$$

- Central Limit Theorem



TODAY'S PLAN

01

Z-Score for Sample Means

02

Introduction to NHST

03

Steps of NHST, z-test Example

04

Wrap Up

Learning objectives

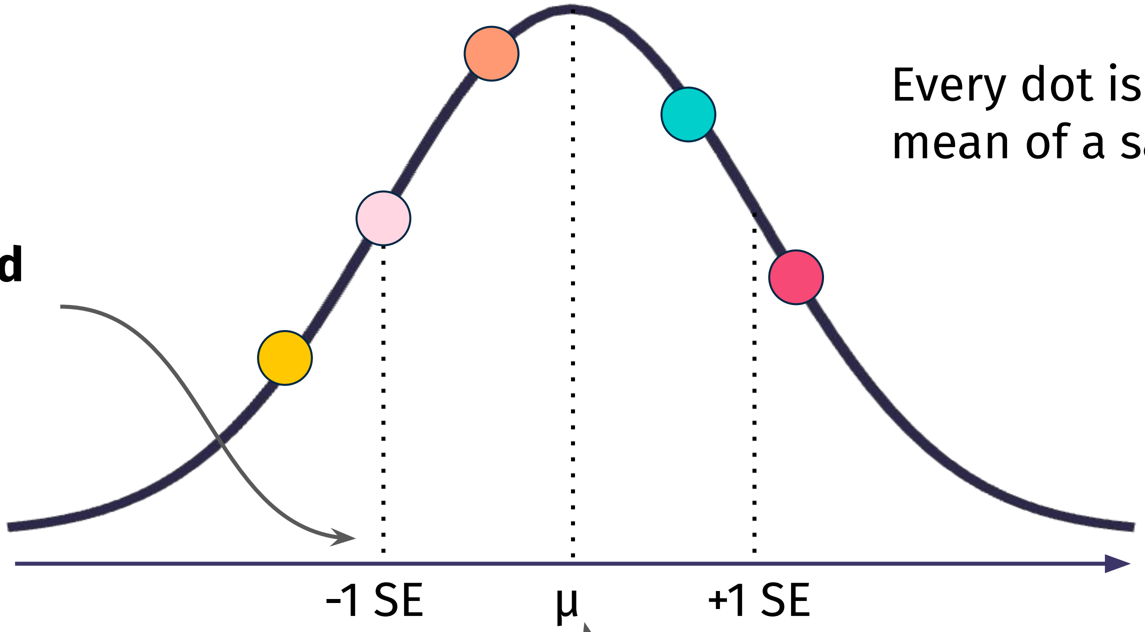
- Calculate and interpret the **z-score for a sample mean**
- Describe the **purpose** of **hypothesis testing**
- Describe the **basic logic and steps** behind Null Hypothesis Significance Testing (**NHST**)
- Understand and can explain what **chance** means

01

Z-Score for Sample Means

Recap: Sampling Distribution of the Mean

The SD of the sampling distribution is called **standard error (σ_M).**



Every dot is the mean of a sample.

Sampling distribution of the mean is centered at the population mean.

Remember we said last class:
"Standard error tells us how much we can trust that one snapshot/sample to infer about our population."

How do we put what we learned all together?

Our goal is always <u>inference</u>

In other words, we want to use our **sample** to say something about the **population.**

Why do we care? How is this used in life?

An example

Why do we care? How is this used in life?

UIC is applying for federal funding to support minority-serving institutions. One of the claims is that UIC students are more linguistically diverse than students at other R1 universities.

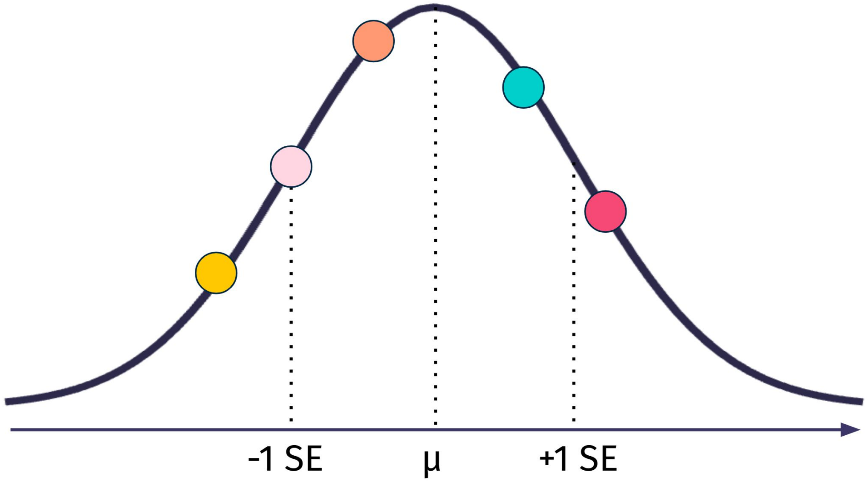
- National data show that the average number of languages spoken fluently by college students is 1.1 (σ = 1).

- We collect a sample of 40 UIC students, and they report an average of 1.5 languages spoken fluently.

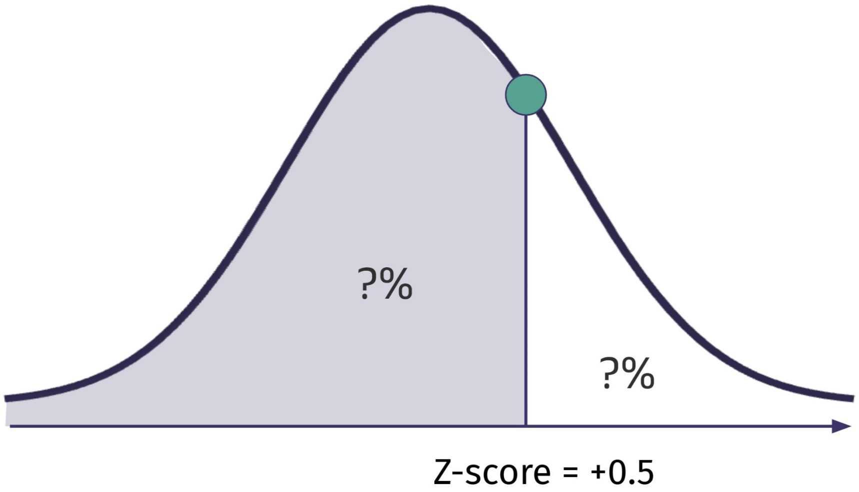
Is our sample unusually high compared to the national average, or could this difference just be due to chance?

Combining Z-Scores & Probability

We can calculate a z-score for every sample mean.



And figure out the probability of observing that sample mean.

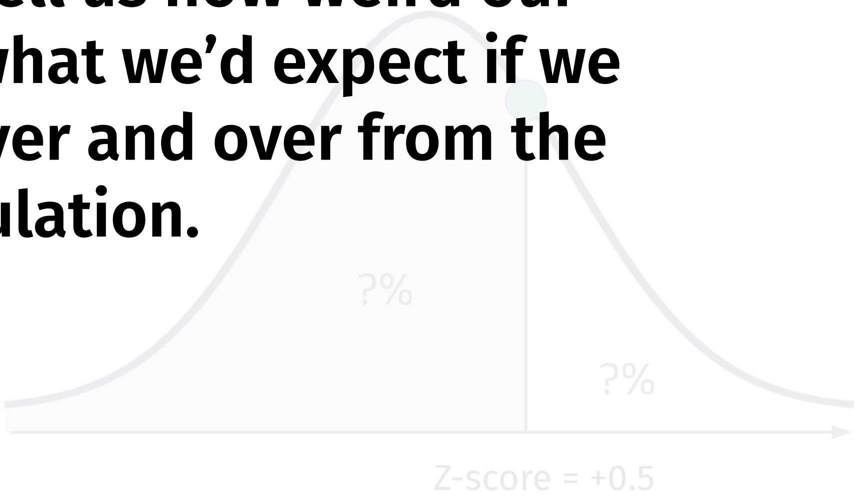


Combining Z-Scores & Probability

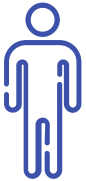
We can calculate a z-score for every sample mean.

And figure out the probability of observing that sample mean.

The probability would tell us how weird our sample is, compared to what we'd expect if we kept drawing samples over and over from the same population.

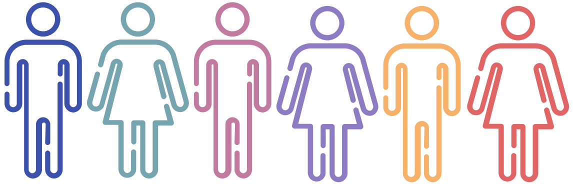


A *Slightly New* **Formula**



For a **single score** *(X)*

$$Z = \frac{\boxed{X} - \mu}{\boxed{s}}$$



For a **sample mean** *(M)*

$$Z = \frac{\boxed{M} - \mu}{\boxed{\sigma_M}}$$

Calculating a Z-Score for Sample Mean

1. Calculate **Standard Error** (using formula)

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

2. Calculate **Z-Score for the sample mean** (using formula)

$$Z = \frac{M - \mu}{\sigma_M}$$

3. Look up **probability** (using z-table)

Example

National data suggests that **μ = 1.1** (**σ = 1**). If I took a random sample of *n* = 40 students at UIC, how unusual would it be to get a sample mean ***greater than what we observed, M = 1.5***?

STEP 1: Calculate Standard Error	STEP 2: Calculate Z-Score	STEP 3: Find Probability

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_M = \frac{1}{\sqrt{40}}$$

$$\sigma_M = \mathbf{0.16}$$

$$Z = \frac{M - \mu}{\sigma_M}$$

$$Z = \frac{1.5 - 1.1}{0.16}$$

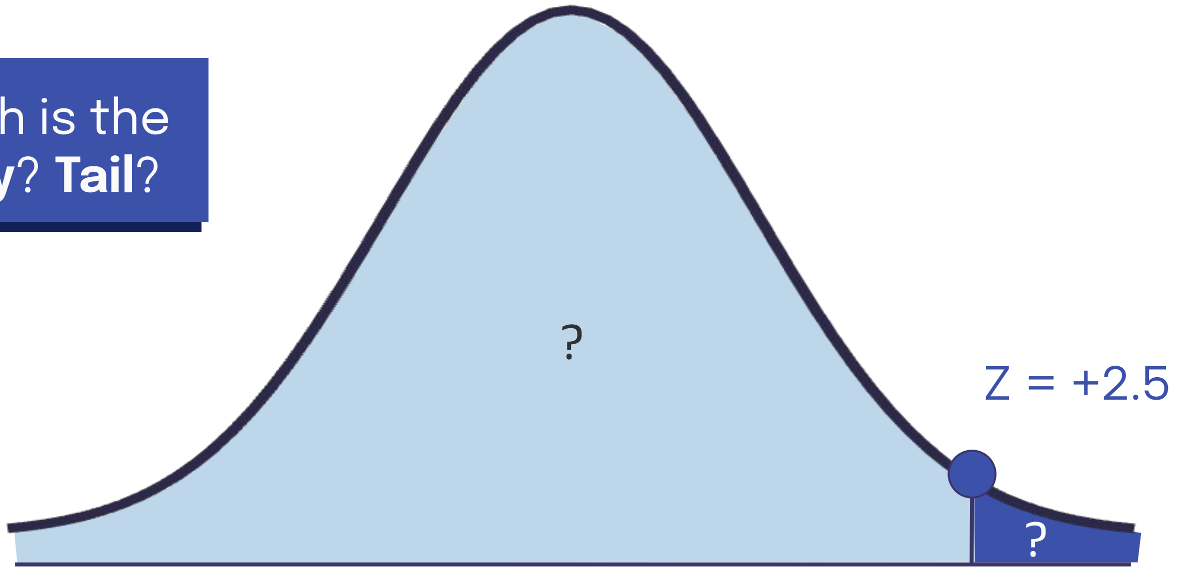
$$Z = \mathbf{2.5}$$

(look up **Z = 2.5** in Z–Table and find probability in **tail**)

Example

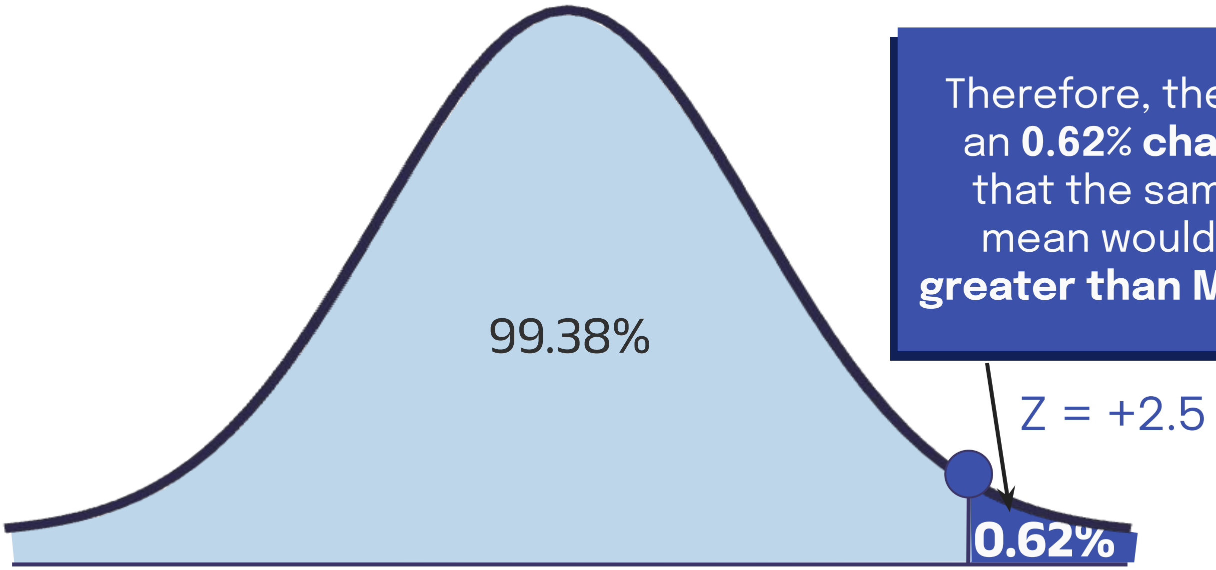
National data suggests that **μ = 1.1** (**σ = 1**). If I took a random sample of *n* = 40 students at UIC, how unusual would it be to get a sample mean ***greater than what we observed, M = 1.5***?

Which is the **body**? **Tail**?



Example

National data suggests that **μ = 1.1** (**σ = 1**). If I took a random sample of *n* = 40 students at UIC, how unusual would it be to get a sample mean ***greater than what we observed, M = 1.5***?



Therefore, there is an **0.62% chance** that the sample mean would be **greater than M = 1.5**.

From our results, we are basically saying:

UIC is “weird”, students at UIC do speak more languages than the national average, statistically.

Therefore, give us the funding!

This is very similar to calculating z-scores, finding probability—except now it’s about a <u>sample mean</u>, not an individual person.

It’s not super likely we’d get a sample mean above 1.5 if the true population mean is 1.1.

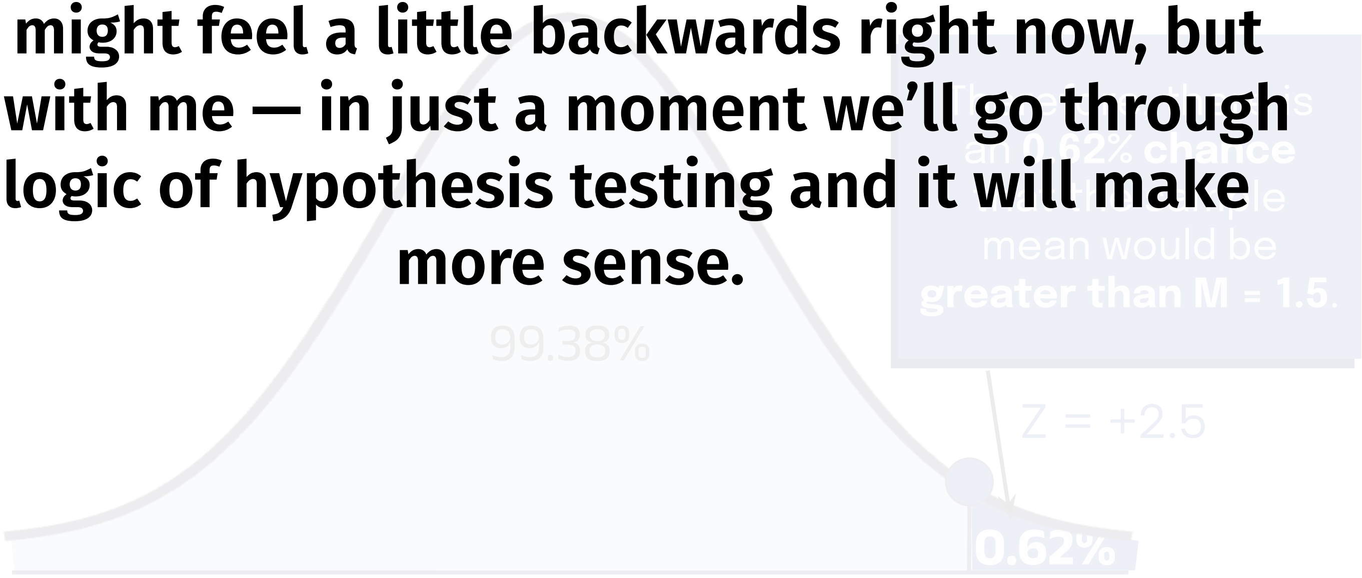
Similarly, if I take a random sample, and my sample mean is 1.8, maybe our original assumption (that μ = 1.1) needs to be questioned.

This is the logic of using one sample to “infer” about the population.

Example

National data suggests that **μ = 1.1** (**σ = 1**). If I took a random sample of *n* = 40 students at UIC, how unusual would it be to get a sample mean ***greater than what we observed, M = 1.5***?

This might feel a little backwards right now, but stick with me — in just a moment we'll go through the logic of hypothesis testing and it will make more sense.



02

Introduction to NHST

What is NHST?

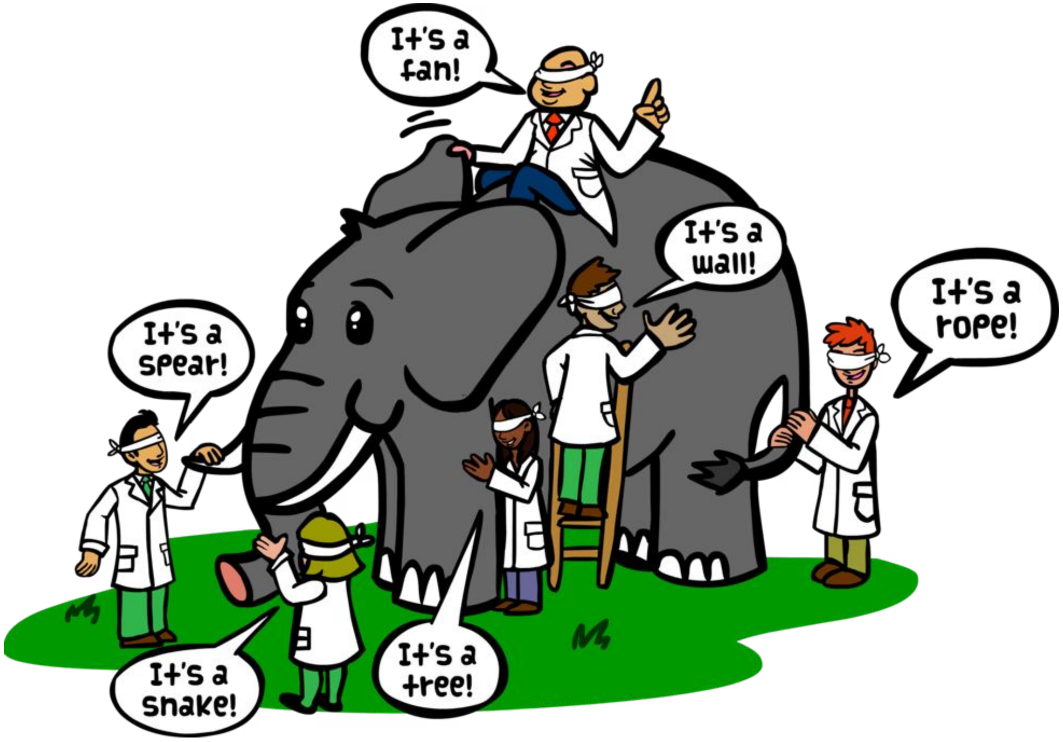
Null Hypothesis Significance Testing is a common method in psychology and inferential statistics where we use a **sample** to **test** whether our results are strong enough to **challenge a default assumption**. It helps us decide if what we found is likely due to chance or if **something real** might be going on.

N ull
H ypothesis
S ignificance
T esting

People often refer to Null Hypothesis Significance Testing (NHST) simply as “hypothesis testing.”

In most psychology and behavioral science contexts, the two terms are used interchangeably.

Intuition behind NHST



Population = the full elephant (truth we can't directly see)

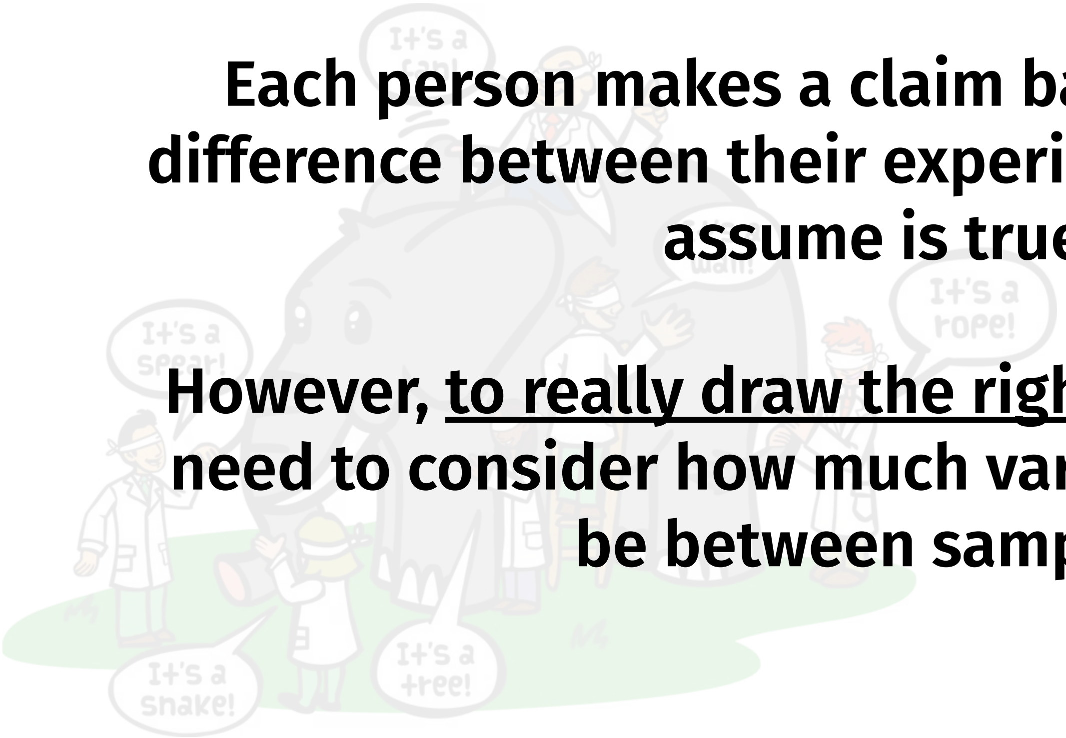
Sample = what each blindfolded person experiences

Null hypothesis = our default belief about the elephant

Data = the part we touch

Inference = using what we felt to decide what we think is out there

Intuition behind NHST



Each person makes a claim based only on the difference between their experience and what they assume is true.

However, <u>to really draw the right conclusion</u>, they need to consider how much variation there could be between samples.

03

Steps of NHST, z-test Example

Intuition: The Basic Steps of NHST

01 “My experience super feels like a snake. So I believe the population is a snake. This represents the **alternative hypothesis**. The **null** would be: The population is not a snake (or: it could be anything — tree, wall, rope, etc.).”

02 “If very few people felt what I did, and we’re assuming the elephant isn’t a snake, then our experience is too unusual to just be chance (i.e., sampling error). So I start to doubt the assumption that the population is not a snake.”

03 “Let me see whether my experience is very different from what the others experience.”

04 “Given everyone's experiences, should I stick with my belief that it's a snake — or do I now think that conclusion doesn’t hold up?”

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

Example

A psychology researcher is interested in **whether students at UIC are getting less sleep** than the national average for college students, which is 7.5 hours per night with a standard deviation of 1.5 hours.

Population	**UIC Sample**
(all college students)	(one class)
$\mu = 7.50$	$M = 6.50$
$\sigma = 1.50$	$n = 60$

01 Restate your research question as hypotheses

There are **two types** of hypotheses:

Alternative Hypothesis

H_A

The alternative hypothesis is what you **expect** or **predict** to happen.

Null Hypothesis

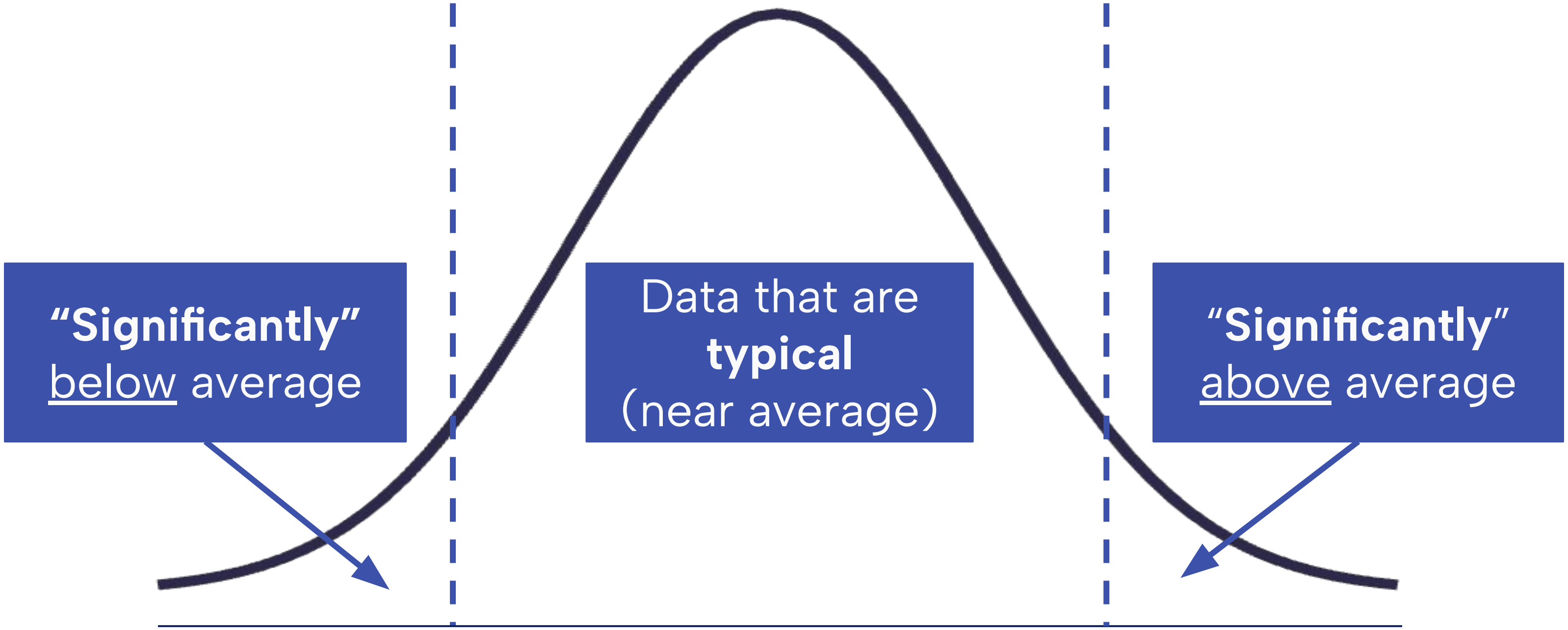
H_0

The null hypothesis claims that there is **no effect, relationship,** or **difference**; it is the opposite of the alternative hypothesis.

Example

Alternative Hypothesis	Null Hypothesis
H_A	H_0
The amount of sleep UIC students get is **significantly different** the national average.	The amount of sleep UIC students get is the **same** as the national average.

02 Decide what is considered “extreme” or significant”



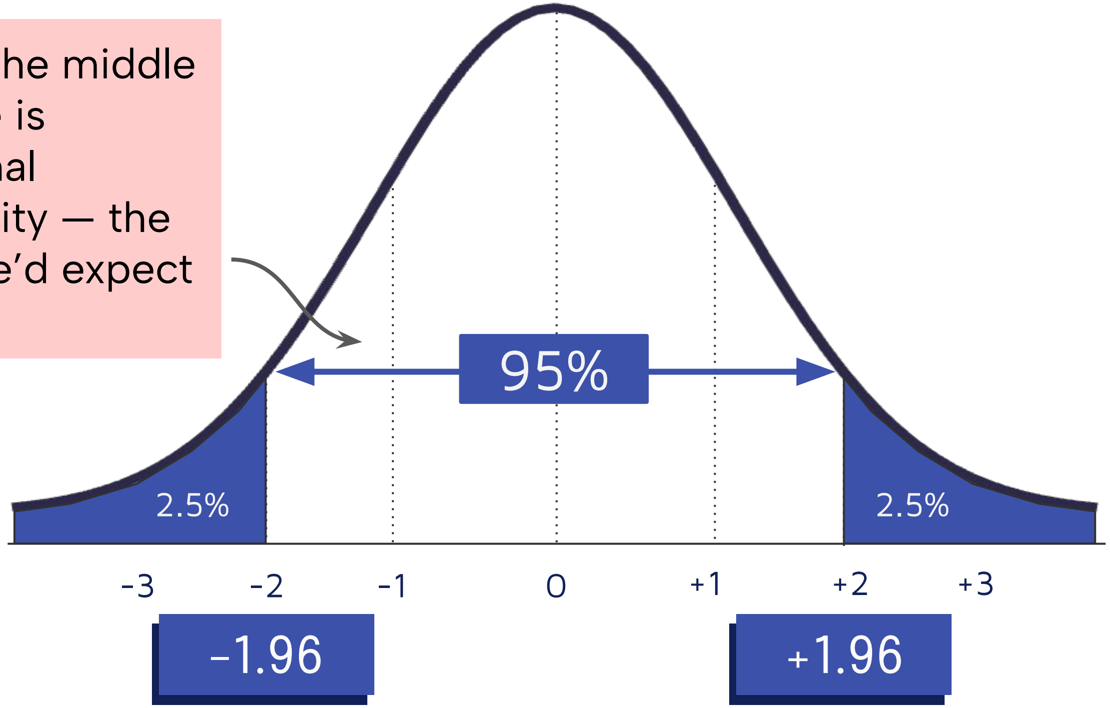
Note: this is called a “two-tailed test”

The results of NHST provide you with a **probability** (percentage). We call this a **p-value**.

You typically need a **p-value** of **less than 5%** to declare "**statistical significance**."

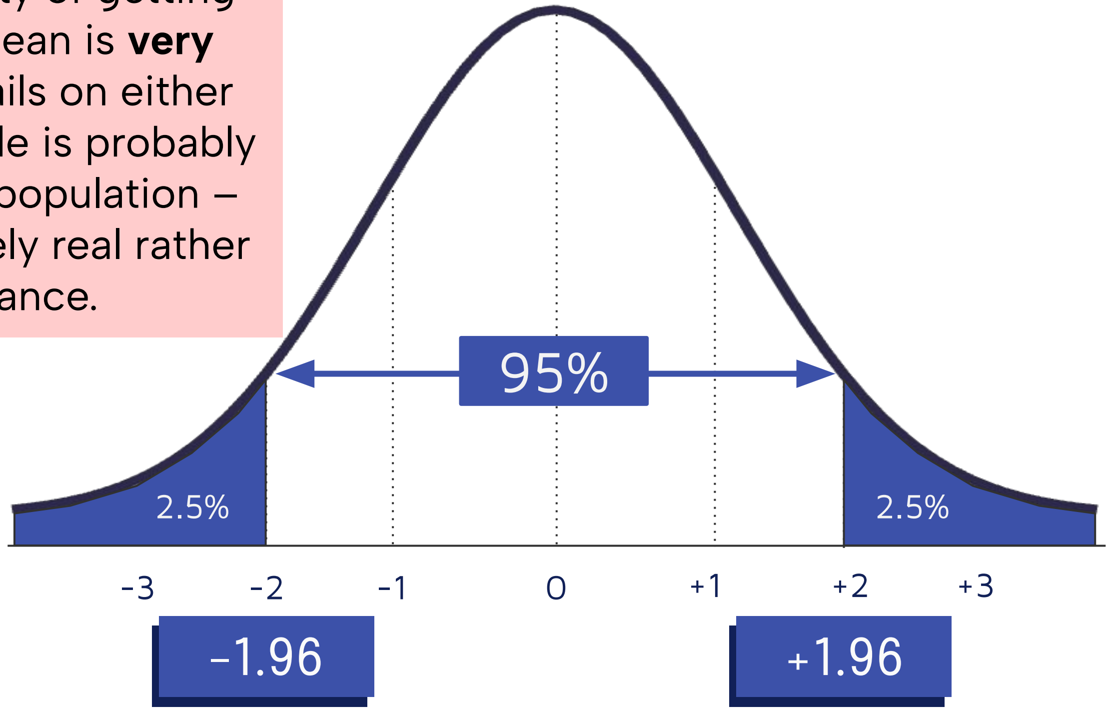
How do we determine what is "**significant**?"

Anything within the middle 95% of the curve is considered normal sampling variability — the kind of results we'd expect just by chance.



How do we determine what is “**significant**?”

If the probability of getting our sample mean is **very small** (in the tails on either side), the sample is probably different from population – the effect is likely real rather than chance.



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

STEP 1:
Calculate *Standard Error*

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_M = \frac{1.5}{\sqrt{60}}$$

$$\sigma_M = \mathbf{0.19}$$

STEP 2:
Calculate *Z-Score*

$$Z = \frac{M - \mu}{\sigma_M}$$

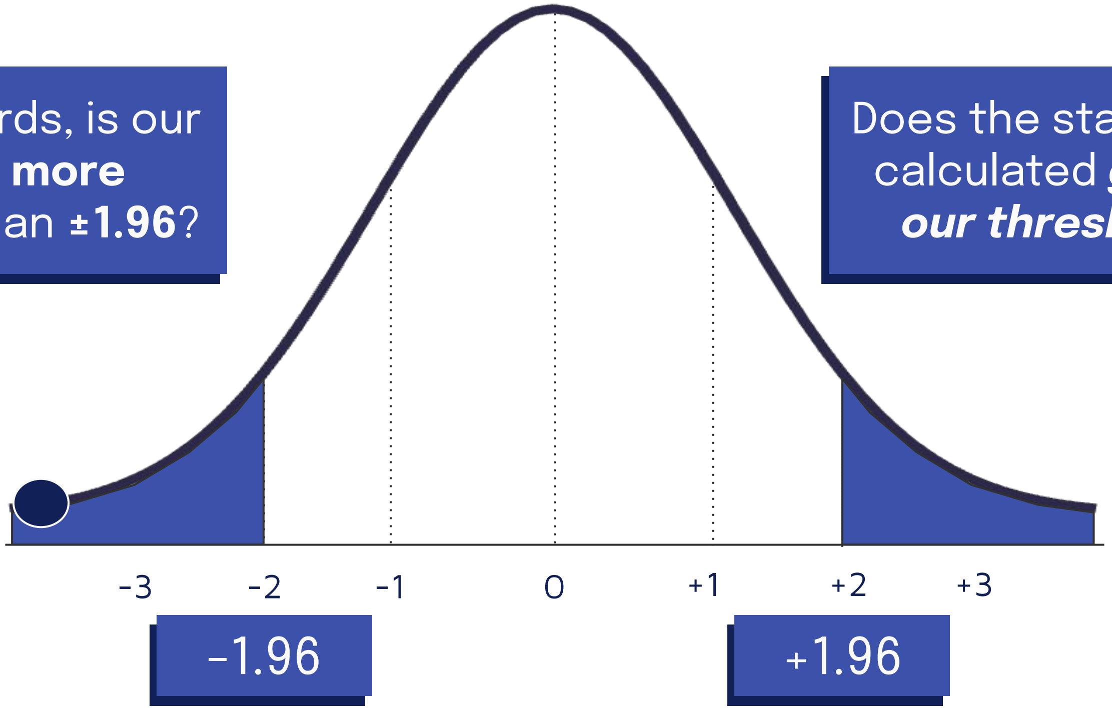
$$Z = \frac{6.5 - 7.5}{0.19}$$

$$Z = \mathbf{-5.28}$$

04 Make a final decision

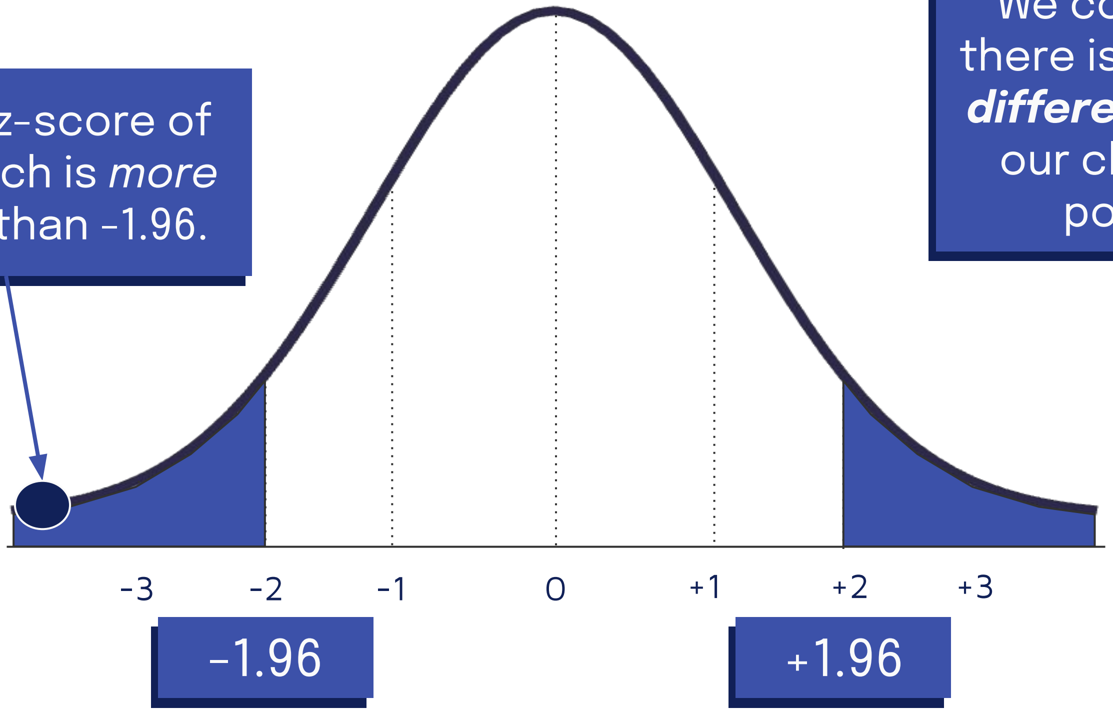
In other words, is our z-score **more extreme** than **±1.96**?

Does the statistic we calculated ***go past our thresholds***?



04 Make a final decision

We got a z-score of –**5.28,** which is *more extreme* than -1.96.



We conclude that there is a ***significant difference*** between our class and the population.

If the null hypothesis were true, the probability of getting a result as extreme as ours (or more extreme) is very low. So, we reject the null hypothesis.

We conclude that the results **support our hypothesis**. UIC students sleep significantly less than the population.

What you just did is called a <u>z-test</u>.

In a z-test, you know the population mean and population SD.

z-test is just one type of NHST. NHST is a general framework that includes many kinds of tests (like t-tests, ANOVAs, etc.) that you will soon learn about.

Another Example

A psychologist is studying whether college students who regularly attend peer support groups report different depression levels compared to the general nonclinical population. The psychologist administers the Beck Depression Inventory (BDI) to a random sample of **36** students who have participated in a peer support program for at least six months. The population mean for the BDI (nonclinical) is $\mu = 12.60$ and a standard deviation of $\sigma = 9.90$. The sample has an average BDI score of $M = 10.60$.

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

01 Restate your research question as hypotheses

Alternative Hypothesis	Null Hypothesis
H_A	H_0
Students in the peer support program have **different depression levels** than the general population.	Students in the peer support program have **the same depression levels** as the general population.

The Basic Steps of NHST

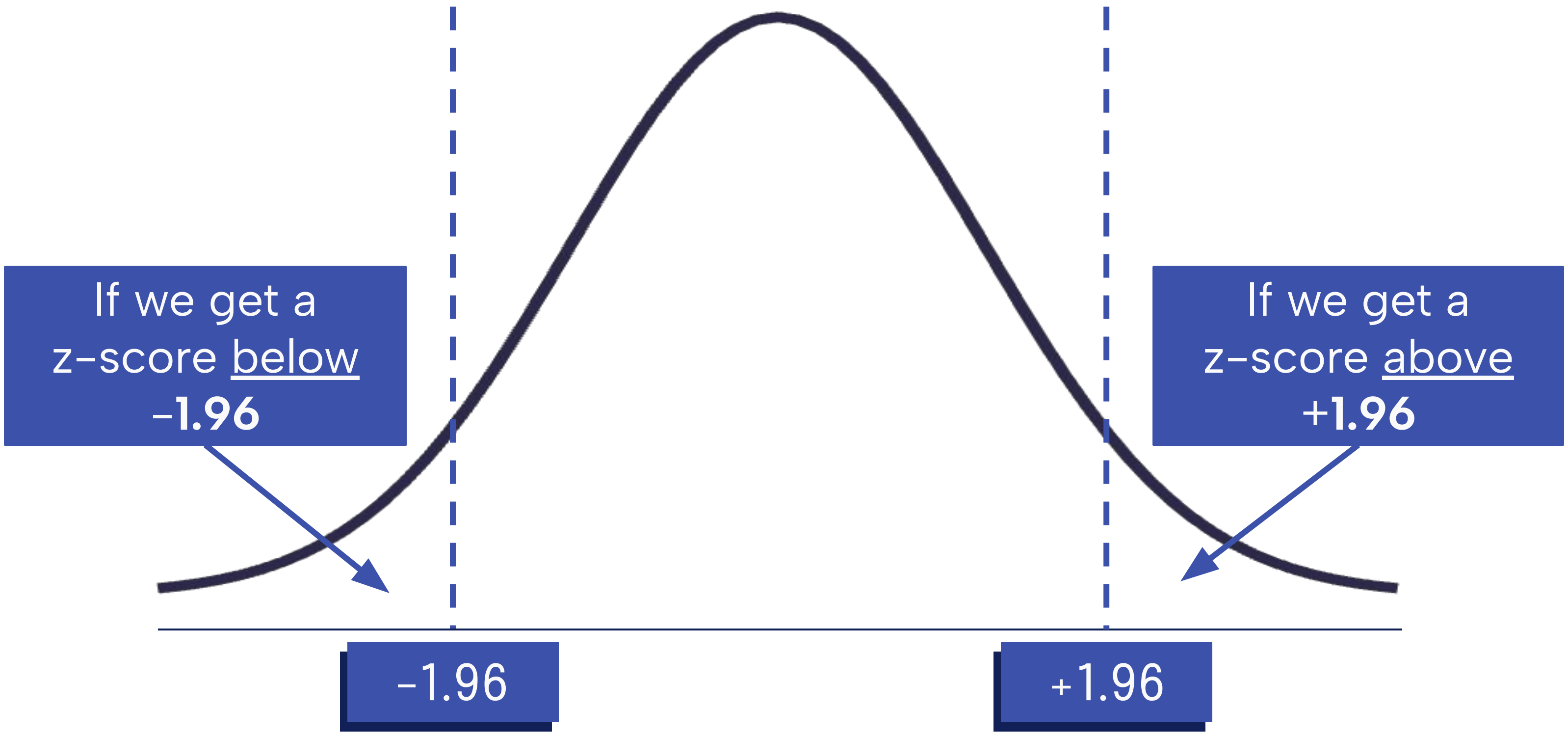
01 Restate your research question as **hypotheses**

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

02 Decide what is considered "extreme" or significant"



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

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

STEP 1:
Calculate *Standard Error*

$$\sigma_M = \frac{\sigma}{\sqrt{n}}$$

$$\sigma_M = \frac{9.90}{\sqrt{36}}$$

$$\sigma_M = \mathbf{1.65}$$

STEP 2:
Calculate *Z-Score*

$$Z = \frac{M - \mu}{\sigma_M}$$

$$Z = \frac{10.6 - 12.6}{1.65}$$

$$Z = \mathbf{-1.21}$$

The Basic Steps of NHST

01 Restate your research question as **hypotheses**

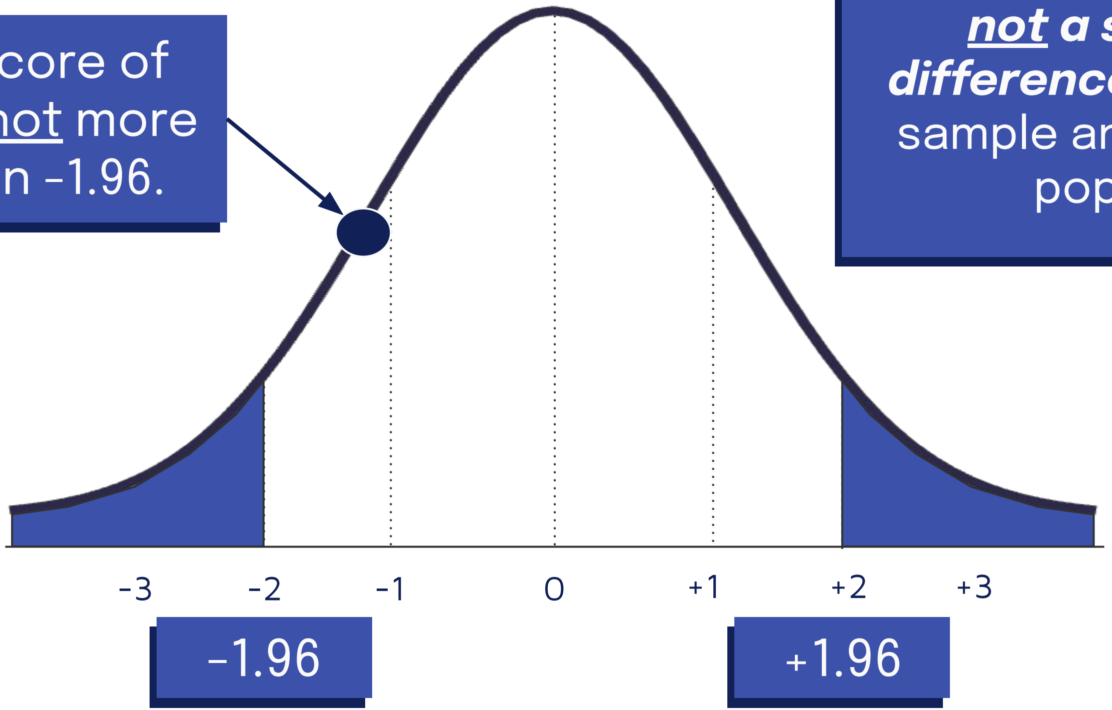
02 Decide what cutoff score is “**extreme**” or “**significant**”

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

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

04 **Make a final decision**

We got a z-score of -**1.21.** That is <u>not</u> more extreme than -1.96.



We conclude that there is ***<u>not</u> a significant difference*** between our sample and the general population.

Since our z-score is not more extreme than ±1.96, the probability of getting a result as extreme as ours (or more extreme) is not low enough to be considered rare.

Therefore, we **fail to reject** the null hypothesis – our result could plausibly be due to chance.

We conclude that the results **do not support our hypothesis**. Students who regularly attended the peer support groups did not report different depression levels compared to the general population.

Since our z-score is not more extreme than ±1.96, the probability of getting a result as extreme as ours (or more extreme) is not low enough to be considered rare.

Therefore, we fail to reject the null hypothesis – our result could plausibly be due to chance.

We will practice this more in the next class.

We conclude that the results **do not support our hypothesis**. Students who regularly attended the peer support groups did not report different depression levels compared to the general population.

In-Class Activity
(ICA 7)

1. I will **assign your group a mean** (M).

2. You will **go through the four basic NHST steps.**
   - State your two hypotheses
   - Your cutoff score will be **Z = ±1.96 *(draw it out)***
   - Calculate your z-score and make your final decision

3. Complete **ICA 7** on Blackboard.

Scenario

Nationwide, first-year college students have an **mean stress level score of 30** (scale of 0–50) with a **standard deviation of 8**. A university counselor at UIC suspects that students at their school experience higher stress than the national average. You get a sample of **25** students and calculate their mean.

$\mu = ?$	$\sigma = ?$	$n = ?$

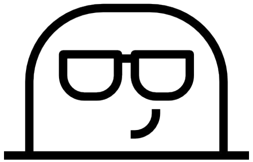
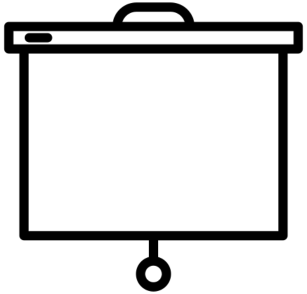
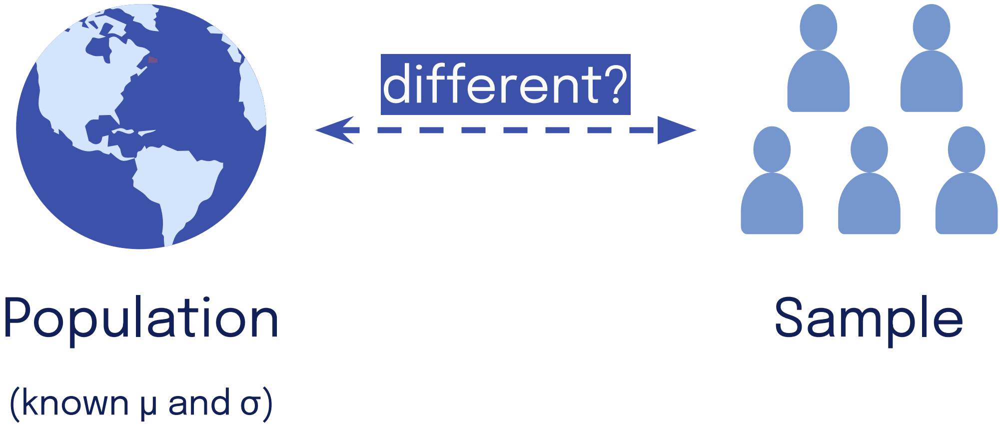
 

Table 5	Table 3	Table 1
$M = 33.3$	$M = 34.8$	$M = 33.6$
Table 6	Table 4	Table 2
$M = 35.0$	$M = 30.5$	$M = 27.4$

What we just did is called a **z-test**, which compares a **sample** to a **population** with a *known* *μ and σ*.



04

Wrap Up

This Friday’s Lab

How would you like us to use the lab session before Exam 1 (Lab 4)? Which option do you think would help you feel most prepared and confident for the exam? Please pick the one you believe would help you succeed on the exam.

Copy chart

15 responses

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

Next Class!

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