

One-Sample t-test

Lecture 9

Emma Ning, M.A.

Your achievements so far..

- Z-Score
- Sampling distribution
- Standard error
- NHST
- Steps of NHST & z-Test
- Infer what our sample says about the population!



Your achievements so far...

You can already be somewhat a scientist and testing your hypotheses with experiments/studies! EXCEPT...

- Sampling distribution

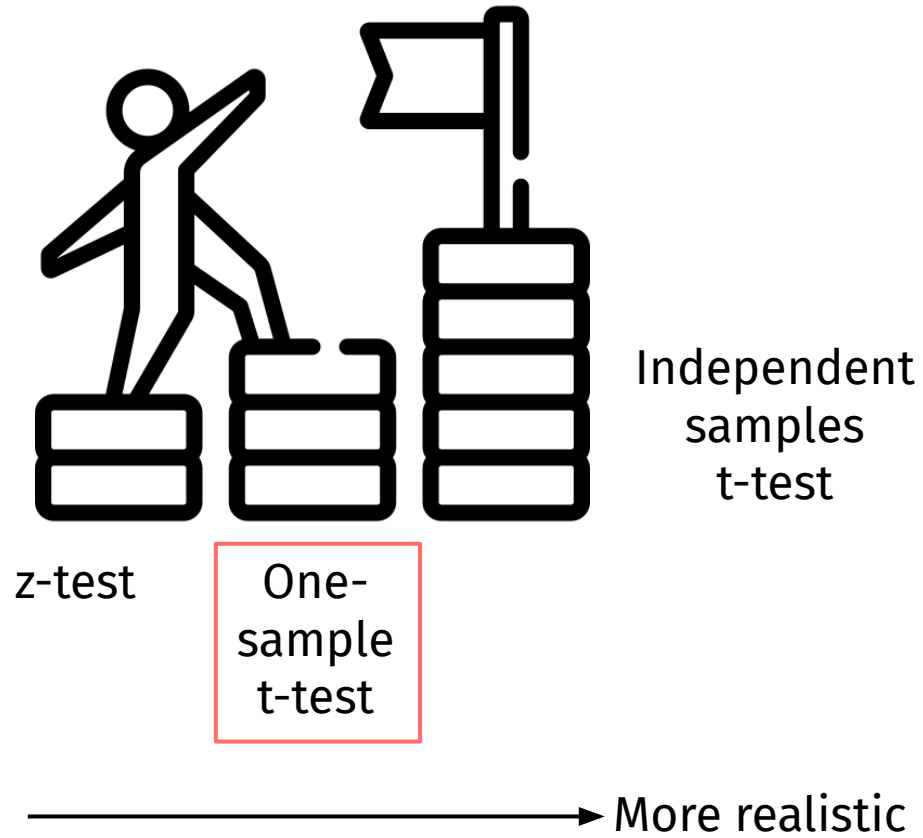
- Standard error

Remember the formula for z-test... doesn't that still

- **involve population SD? How do we get that?**

- Infer what our sample says about the population!

We don't. That's why we have today & rest of the semester.



TODAY'S PLAN

01

Moving From Z to T

02

Worked Example

03

Effect Size

04

Type I & Type II Errors

05

Wrap Up

Learning objectives

- **Differentiate** between when to use a z-test and a **one sample t test**.
- Calculate the **degrees of freedom, estimated standard error, t statistic** for a one sample t test.
- **Conduct** a one sample t test using the NHST steps.
- **Articulate and interpret** the results of a one sample t test in APA style.
- Differentiate **Type I and Type II errors**, and understand how choosing α affects the tradeoff between false alarms (Type I) and misses (Type II).
- **Understand** and be able to **explain** Type 1 and Type II errors (involving the concept of **chance**)



Moving From Z to T

Recap: Intuition behind z-test

Looks scary?
You've already
mastered it!

$$z = \frac{M - \mu}{\sigma_M} = \frac{M - \mu}{\frac{\sigma}{\sqrt{n}}}$$

This is how different our
sample mean is from the
population mean.

Altogether, this formula says:
**How far is my sample mean from the
population mean, measured in units of
standard error?**

Similar to "\$5 per pound"!

This is the standard error
of the mean.

Recap: Intuition behind z-test

This is how different our sample mean is from the population mean.

Then, all we need to do is to find out the p -value associated with our z-score, and make a decision!

Altogether, this formula says:
How far is my sample mean from the population mean, measured in units of standard error?

Similar to "\$5 per pound"!

This is the standard error of the mean.

In reality, **we often do not know**
the population standard
deviation (σ), so we must use a
different type of statistical test.

z test  **t test**

In reality, we often do not know

There are different types of t-tests.

Today, we will focus on one-sample t-test.

We will cover other types of t-tests later.

z test → **t test**

Comparison between z-test & one-sample t-test

$$z = \frac{M - \mu}{\boxed{\sigma}_M} = \frac{M - \mu}{\frac{\boxed{\sigma}}{\sqrt{n}}}$$

$$t = \frac{M - \mu}{\boxed{s}_M} = \frac{M - \mu}{\frac{\boxed{s}}{\sqrt{n}}}$$

	z-test	One-sample t-test
μ	✓	✓
σ	✓	✗

What's the difference in the denominator?

Since we do not know the population σ , we must **estimate** it with our **sample standard deviation**.

We call this estimated standard error

$$t = \frac{M - \mu}{s_M} = \frac{M - \mu}{\frac{s}{\sqrt{n}}}$$

Estimated Standard Error (s_M)

We **estimate** the standard error using our *sample* standard deviation when the value of σ is unknown.

$$S_M = \frac{\text{sample standard deviation}}{\sqrt{\text{sample size}}}$$

$$s_M = \frac{s}{\sqrt{n}} \qquad t = \frac{M - \mu}{s_M} = \frac{M - \mu}{\frac{s}{\sqrt{n}}}$$

Estimated Standard Error (s_M)

We **estimate** the standard error using our *sample* standard deviation when the value of σ is unknown.

Because we're estimating SE from the sample SD (instead of using σ), we introduce extra uncertainty.

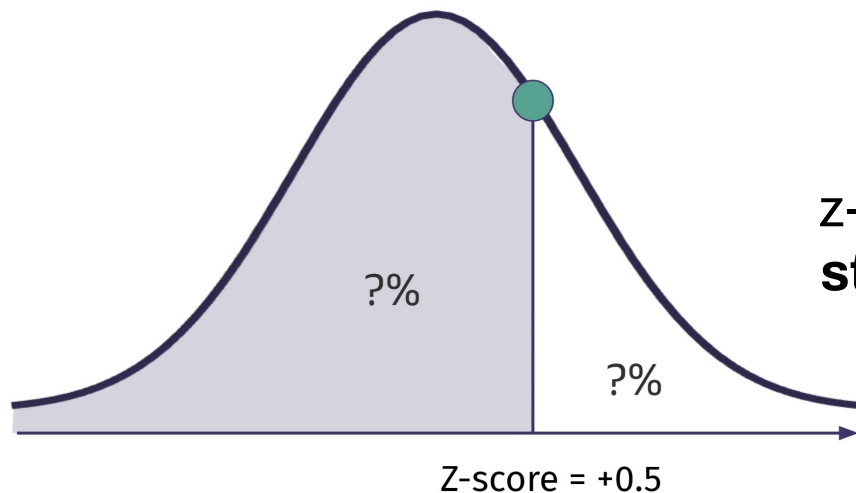
So, instead of using z-tables, we need to use t-tables, which account for that extra uncertainty.

$$s_M = \frac{s}{\sqrt{n}} \quad t = \frac{M - \mu}{s_M} = \frac{M - \mu}{\frac{s}{\sqrt{n}}}$$

We are using the t-distribution instead of the z-distribution

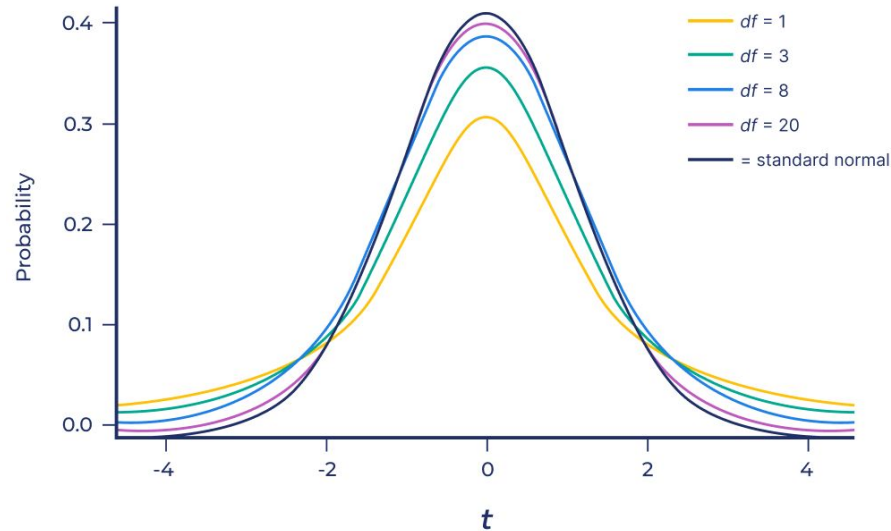
What does this mean?

Remember, previously, we calculate our z-score, then look up the z-table to find our p -value:



This curve is called the z-distribution, aka **standard normal** distribution.

The idea is the same here.
However, instead of using the z-table, we are using the **t-table** that corresponds to the **t-distribution**.



Df is back! Do you remember what it is?

df

n - 1

For a one-sample t test, the formula for *df* is just ***n - 1***.

Remember we introduced df when calculating sample variance?

I used the sudoku example there. This is the same idea here.

Example

Suppose you randomly sample **10** UIC students and measure their daily TikTok usage. You calculate the sample mean to be **1.5 hours**. You use a one-sample t test to determine whether the mean daily TikTok usage of UIC students is different than the national average.

How many **degrees of freedom** do we have?

$$df = n - 1$$

$$df = 10 - 1$$

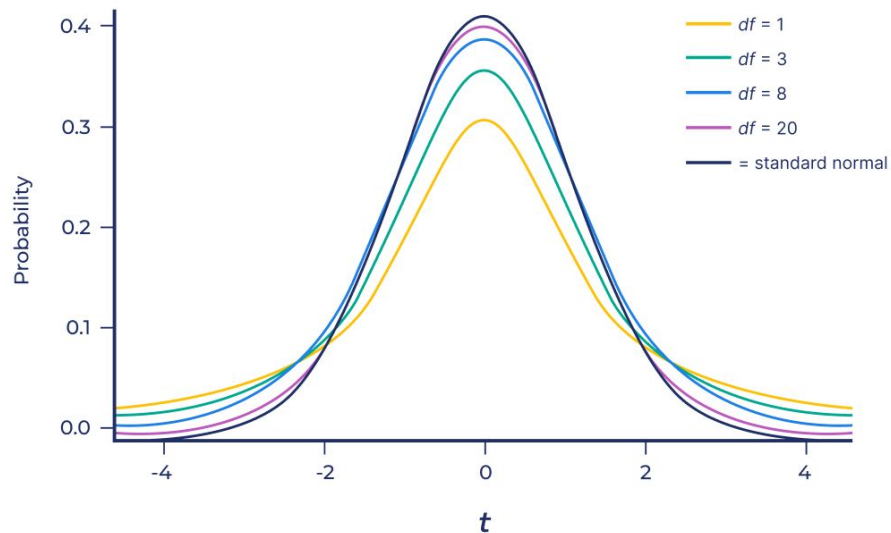
$$\mathbf{df = 9}$$

Example

We need degrees of freedom (df) in a one-sample t-test (and many other tests we'll see this semester).

This is because we are estimating something (like the population standard deviation σ) from our sample data. Every time we estimate, we lose some "freedom" because part of the data has already been used to calculate the estimate.

$$df = 9$$



Notice how each df corresponds to a different curve?
These are all t-distributions with varying **dfs**.

Because $df = n - 1$, the larger the df, the more similar the t-distribution is to the standard normal distribution.

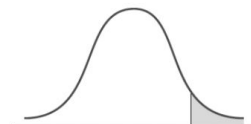


So if we calculate a t-score, just like a z-score, how do we look up our p -value from the t-table?

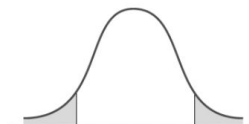
Notice how each df corresponds to a different curve?
These are all t-distributions with varying **dfs**.

Because $df = n-1$, the larger the df, the more similar the t-distribution is to the standard normal distribution.

t Distribution Table (t table)



One tail
(either right or left)

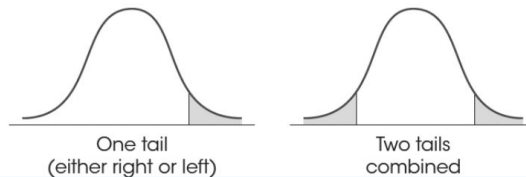


Two tails
combined

df	Proportion (a) in <u>One</u> tail				
	.05	.025	.01	.005	.0005
	Proportion (a) in <u>Two</u> tails combined				
	.10	.05	.02	.01	.001
1	6.314	12.706	31.821	63.657	636.578
2	2.920	4.303	6.965	9.925	31.600
3	2.353	3.182	4.541	5.841	12.924
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.869
6	1.943	2.447	3.143	3.707	5.959
7	1.895	2.365	2.998	3.499	5.408
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965

df	Proportion (a) in <u>One</u> tail				
	.05	.025	.01	.005	.0005
	Proportion (a) in <u>Two</u> tails combined				
	.10	.05	.02	.01	.001
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.768
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.689
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.660
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.290

t Distribution Table (t table)



df	Proportion (a) in <u>One</u> tail				
	.05	.025	.01	.005	.0005
	Proportion (a) in <u>Two</u> tails combined				
	.10	.05	.02	.01	.001

1	6.314	12.708	31.821	63.657	636.576
2	2.920	4.308	6.965	9.925	31.699
3	2.353	3.182	5.841	8.163	12.941
4	2.132	2.776	5.208	7.171	10.215
5	2.015	2.576	4.779	6.608	9.588
6	1.943	2.447	4.501	6.314	9.246
7	1.895	2.365	4.343	6.163	9.001
8	1.860	2.306	4.221	6.024	8.828
9	1.833	2.262	4.133	5.965	8.718
10	1.812	2.228	4.064	5.909	8.623
11	1.796	2.201	4.015	5.862	8.541
12	1.782	2.179	3.982	5.821	8.473
13	1.771	2.160	3.959	5.789	8.417
14	1.761	2.145	3.938	5.762	8.371
15	1.753	2.131	3.919	5.740	8.333
16	1.746	2.120	3.902	5.721	8.301
17	1.740	2.110	3.888	5.704	8.273

df	Proportion (a) in <u>One</u> tail				
	.05	.025	.01	.005	.0005
	Proportion (a) in <u>Two</u> tails combined				
	.10	.05	.02	.01	.001

18	1.734	2.101	3.878	5.688	8.242
19	1.729	2.093	3.869	5.679	8.216
20	1.725	2.086	3.861	5.671	8.192
21	1.721	2.080	3.854	5.664	8.169
22	1.718	2.075	3.848	5.658	8.148
23	1.715	2.071	3.843	5.653	8.128
24	1.713	2.067	3.839	5.649	8.109
25	1.711	2.064	3.836	5.646	8.091
26	1.709	2.061	3.833	5.643	8.074
27	1.703	2.052	3.823	5.633	8.059
28	1.701	2.048	3.819	5.629	8.046
29	1.699	2.045	3.816	5.626	8.034
30	1.697	2.042	3.813	5.623	8.023
40	1.684	2.021	3.787	5.598	7.999
60	1.671	2.000	3.767	5.577	7.973
120	1.658	1.980	3.745	5.556	7.947
∞	1.645	1.960	3.726	5.536	7.929

First, we need to know if we are doing a **one** or **two-tailed test** and our **alpha level**.

t Distribution Table (t table)

Second, we need to know our **df**.

Proportion (a) in <u>Two tails combined</u>						Proportion (a) in <u>Two tails combined</u>					
df	.10	.05	.02	.01	.001	df	.10	.05	.02	.01	.001
1	6.314	12.706	31.821	63.657	636.578	18	1.734	2.101	2.552	2.878	3.922
2	2.920	4.303	6.965	9.925	31.600	19	1.729	2.093	2.539	2.861	3.883
3	2.353	3.182	4.541	5.841	12.924	20	1.725	2.086	2.528	2.845	3.850
4	2.132	2.776	3.747	4.604	8.610	21	1.721	2.080	2.518	2.831	3.819
5	2.015	2.571	3.365	4.032	6.869	22	1.717	2.074	2.508	2.819	3.792
6	1.943	2.447	3.143	3.707	5.959	23	1.714	2.069	2.500	2.807	3.768
7	1.895	2.365	2.998	3.499	5.408	24	1.711	2.064	2.492	2.797	3.745
8	1.860	2.306	2.896	3.355	5.041	25	1.708	2.060	2.485	2.787	3.725
9	1.833	2.262	2.821	3.250	4.781	26	1.706	2.056	2.479	2.779	3.707
10	1.812	2.228	2.764	3.169	4.587	27	1.703	2.052	2.473	2.771	3.689
11	1.796	2.201	2.718	3.106	4.437	28	1.701	2.048	2.467	2.763	3.674
12	1.782	2.179	2.681	3.055	4.318	29	1.699	2.045	2.462	2.756	3.660
13	1.771	2.160	2.650	3.012	4.221	30	1.697	2.042	2.457	2.750	3.646
14	1.761	2.145	2.624	2.977	4.140	40	1.684	2.021	2.423	2.704	3.551
15	1.753	2.131	2.602	2.947	4.073	60	1.671	2.000	2.390	2.660	3.460
16	1.746	2.120	2.583	2.921	4.015	120	1.658	1.980	2.358	2.617	3.373
17	1.740	2.110	2.567	2.898	3.965	∞	1.645	1.960	2.326	2.576	3.290

df	Proportion (α) in <u>One tail</u>				
	.05	.025	.01	.005	.0005
	Proportion (α) in <u>Two tails combined</u>				
	.10	.05	.02	.01	.001
1	6.314	12.706	31.821	63.657	636.578
2	2.920	4.303	6.965	9.925	31.600
3	2.353	3.182	4.541	5.841	12.924
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.869
6	1.943	2.447	3.143	3.707	5.959
7	1.895	2.365	2.998	3.499	5.408
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965

EXAMPLE

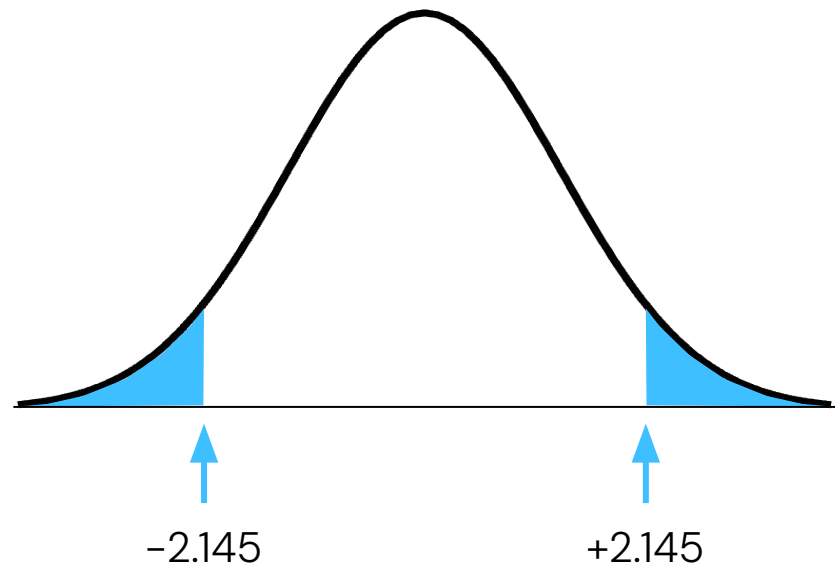
Let's imagine we have a sample of **15 people** and we want to do a **two-tailed test ($\alpha = 0.05$)**.

$$df = 14$$

$$t_{crit} = \pm 2.145$$

df	Proportion (a) in <u>One</u> tail				
	.05	.025	.01	.005	.0005
	Proportion (a) in <u>Two</u> tails combined				
	.10	.05	.02	.01	.001
1	6.314	12.706	31.821	63.657	636.578
2	2.920	4.303	6.965	9.925	31.600
3	2.353	3.182	4.541	5.841	12.924
4	2.132	2.776	3.747	4.604	8.610
5	2.015	2.571	3.365	4.032	6.869
6	1.943	2.447	3.143	3.707	5.959
7	1.895	2.365	2.998	3.499	5.408
8	1.860	2.306	2.896	3.355	5.041
9	1.833	2.262	2.821	3.250	4.781
10	1.812	2.228	2.764	3.169	4.587
11	1.796	2.201	2.718	3.106	4.437
12	1.782	2.179	2.681	3.055	4.318
13	1.771	2.160	2.650	3.012	4.221
14	1.761	2.145	2.624	2.977	4.140
15	1.753	2.131	2.602	2.947	4.073
16	1.746	2.120	2.583	2.921	4.015
17	1.740	2.110	2.567	2.898	3.965

EXAMPLE



df	Proportion (α) in <u>One</u> tail				
	.05	.025	.01	.005	.0005
	Proportion (α) in <u>Two</u> tails combined				
	.10	.05	.02	.01	.001
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.768
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.689
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.660
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.290

PRACTICE AT YOUR TABLE:

PROBLEM 1

You have a sample of **26** people and we want to do a **two**-tailed test ($\alpha = 0.05$).

PROBLEM 2

You have a sample of **19** people and we want to do a **one**-tailed test ($\alpha = 0.01$).

PROBLEM 3

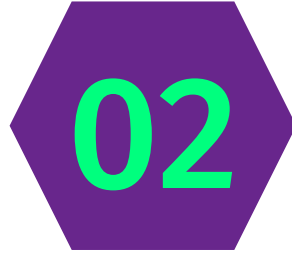
You have a sample of **46** people and we want to do a **two**-tailed test ($\alpha = 0.05$).

Write down your t_{crit} !

Summary

	z test	one sample <i>t</i> test
Test Statistic	z-statistic	<i>t</i> statistic
Standard Error	$\sigma_M = \frac{\sigma}{\sqrt{n}}$	$s_M = \frac{s}{\sqrt{n}}$
Test Statistic Formula	$Z = \frac{M - \mu}{\sigma_M}$	$t = \frac{M - \mu}{s_M}$
Cutoff (critical value)	$\pm 1.96^*$	varies (changes with <i>df</i>)

*assuming $\alpha = 0.05$, two-tailed

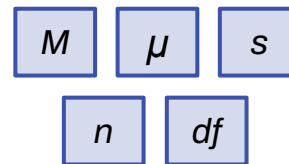


Worked Example

NHST STEPS

1

State the **null** and **alternative hypotheses**.



2

Set your **cutoff score** (find t_{crit}).

*Use a **t-table**!*

3

Calculate your **test statistic** (i.e., t-statistic) and **effect size**.

4

Make your **decision**.

Reject the null

or

Fail to reject the null

5

Interpret your results (in APA style).

NHST STEPS

1

State the **null** and **alternative hypotheses**.

M

μ

s

n

df

2

Set your **cutoff score** (find t_{crit})

Use t -table!

As you can see, the steps look really similar to z-test.

We will go through all the steps first, then talk about effect size in a moment.

4

Make your **decision**.

Reject the null

or

Fail to reject the null

5

Interpret your results (in APA style).

A Few Minor Changes

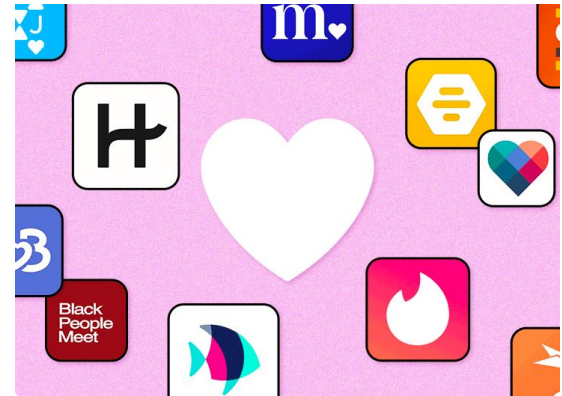
Our process of NHST for a one-sample t test will be similar to our z-test, with a few small changes:

1. Our **formula** for our t statistic has slightly changed (using *estimated* standard error)
2. We will need to calculate and use **degrees of freedom**.
3. We will use a **t table** – not a z-table – to find our critical values (t_{crit}) and critical region.
4. We add **effect size** in our reporting.

Example: Lying on Social Media Apps

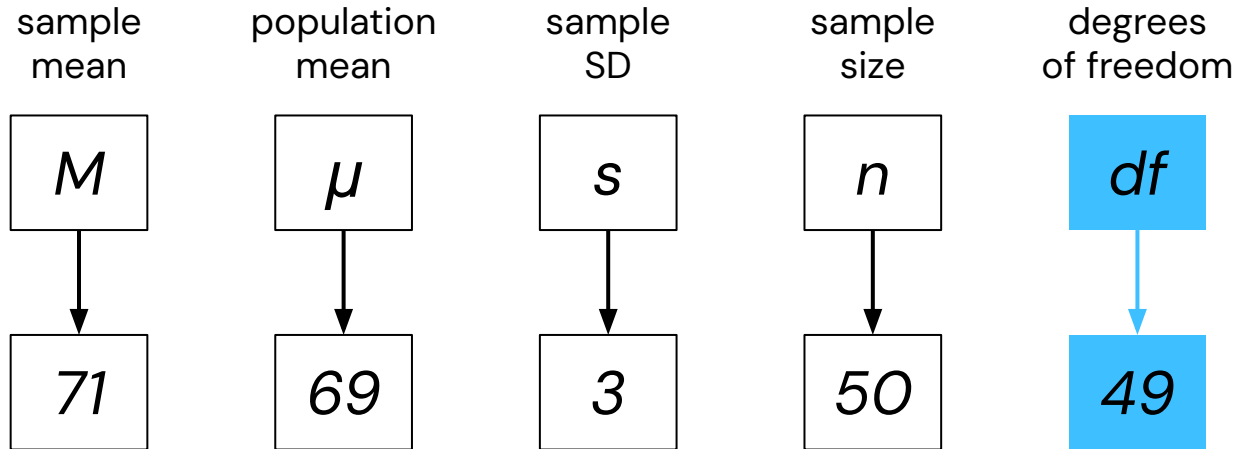
Imagine you are a psychologist studying how people present themselves online, particularly on dating apps. You're curious if males tend to exaggerate their height in their profiles.

To investigate, you collect a sample of **50** self-reported heights from male Tinder users. You know that the average height for adult males in the U.S. is **69.0** inches. In your sample, the average reported height is **71.1** inches with a standard deviation of **3.0** inches.



0

Before you do anything, be sure to annotate the problem and highlight important information.



NHST STEPS



1

State the **null** and **alternative hypotheses**.

M

μ

s

n

df

2

Set your **cutoff score** (find t_{crit}).

*Use a **t-table**!*

3

Calculate your **test statistic** (i.e., t-statistic) and effect size.

4

Make your **decision**.

Reject the null

or

Fail to reject the null

5

Interpret your results (in APA style).

1 State the **null** and **alternative hypotheses**.

H_0 :

Males on Tinder **do not differ** in their reported height compared to the national average.

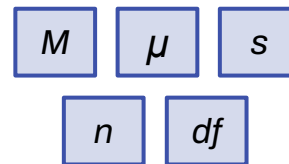
$$\mu = 69$$

H_1 :

Males on Tinder report heights that **significantly differ** from the national average.

$$\mu \neq 69$$

NHST STEPS



1 State the **null** and **alternative hypotheses**.

 **2** Set your **cutoff score** (find t_{crit}).

*Use a **t-table**!*

3 Calculate your **test statistic** (i.e., t-statistic) and effect size.

4 Make your **decision**. *Reject the null* or *Fail to reject the null*

5 **Interpret** your results (in APA style).

2A Find your **cutoff score** (critical value).

df	Proportion (a) in <u>One</u> tail				
	.05	.025	.01	.005	.0005
	Proportion (a) in <u>Two</u> tails combined				
	.10	.05	.02	.01	.001
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.768
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.689
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.660
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.290

What we need to know:

- $df = 49$
- $\alpha = 0.05$
- two-tailed test

What is our t_{crit} ?

2A Find your **cutoff score** (critical value).

df	Proportion (a) in <u>One</u> tail				
	.05	.025	.01	.005	.0005
	Proportion (a) in <u>Two</u> tails combined				
	.10	.05	.02	.01	.001
18	1.734	2.101	2.552	2.878	3.922
19	1.729	2.093	2.539	2.861	3.883
20	1.725	2.086	2.528	2.845	3.850
21	1.721	2.080	2.518	2.831	3.819
22	1.717	2.074	2.508	2.819	3.792
23	1.714	2.069	2.500	2.807	3.768
24	1.711	2.064	2.492	2.797	3.745
25	1.708	2.060	2.485	2.787	3.725
26	1.706	2.056	2.479	2.779	3.707
27	1.703	2.052	2.473	2.771	3.689
28	1.701	2.048	2.467	2.763	3.674
29	1.699	2.045	2.462	2.756	3.660
30	1.697	2.042	2.457	2.750	3.646
40	1.684	2.021	2.423	2.704	3.551
60	1.671	2.000	2.390	2.660	3.460
120	1.658	1.980	2.358	2.617	3.373
∞	1.645	1.960	2.326	2.576	3.290

What we need to know:

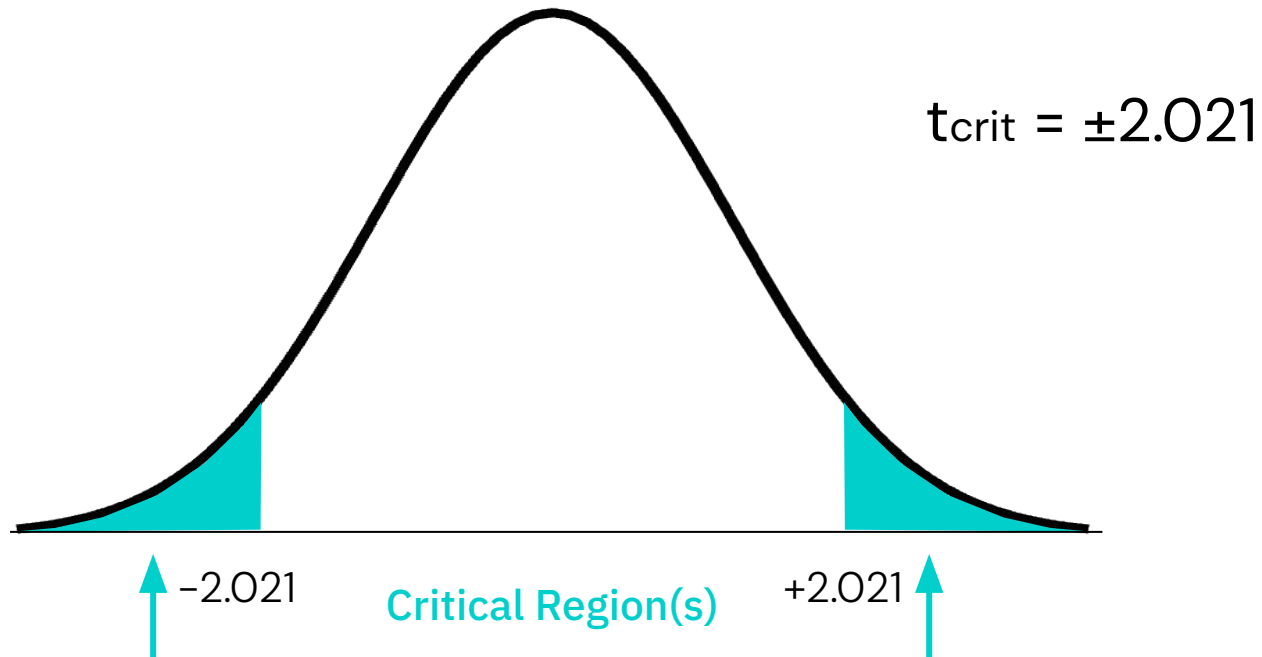
- $df = 49$
- $\alpha = 0.05$
- two-tailed test

Our t_{crit} is **± 2.021**

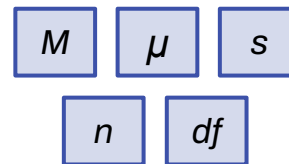
(always **round down** to the lower number)

2B Draw your critical region(s).

You should **always** draw your **critical region** and **label it** during this step.



NHST STEPS



1 State the **null** and **alternative hypotheses**.

2 Set your **cutoff score** (find t_{crit}).

*Use a **t-table**!*

 **3** Calculate your **test statistic** (i.e., t-statistic) and effect size.

4 Make your **decision**. *Reject the null* or *Fail to reject the null*

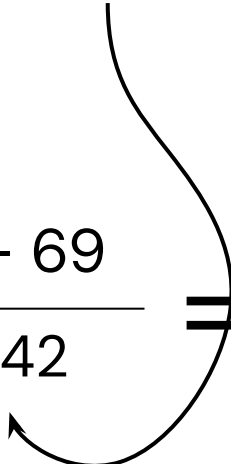
5 **Interpret** your results (in APA style).

3 Compute sample statistics (i.e., t-statistic).

3A Compute estimated standard error.

$$s_M = \frac{s}{\sqrt{n}} = \frac{3}{\sqrt{50}} = 0.42$$

3B Compute your *t* statistic.

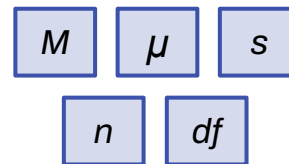
$$t = \frac{M - \mu}{s_M} = \frac{71 - 69}{0.42} = 4.76$$


3 Compute sample statistics (i.e., t-statistic).

3C Compute your ***effect size***.

We will talk about this in the next section! For now, let's continue (:

NHST STEPS



1 State the **null** and **alternative hypotheses**.

2 Set your **cutoff score** (find t_{crit}).

*Use a **t-table**!*

3 Calculate your **test statistic** (i.e., t-statistic) and effect size.

 **4**

Make your **decision**.

Reject the null

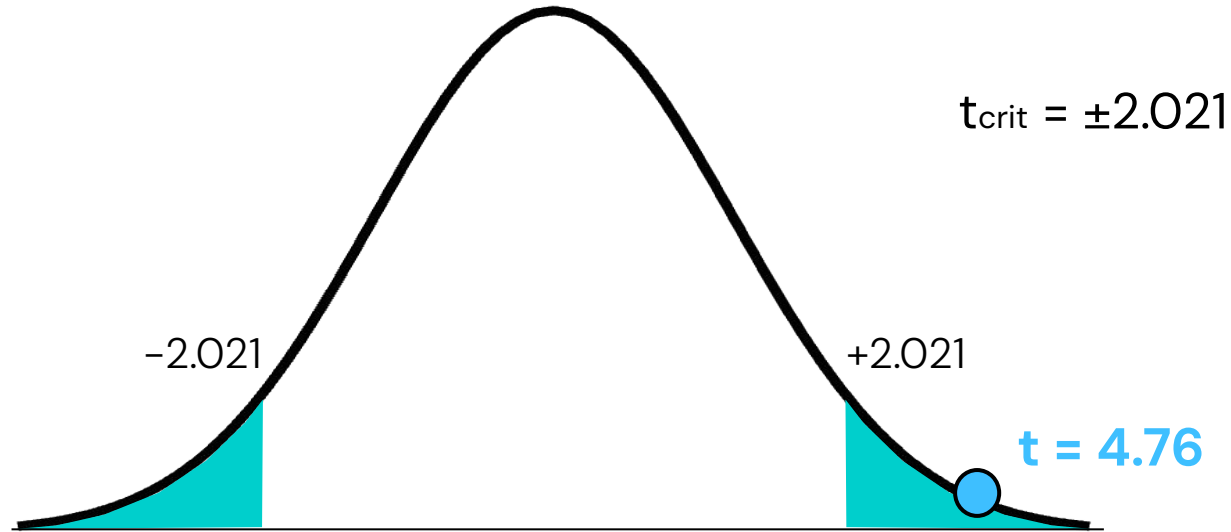
or

Fail to reject the null

5

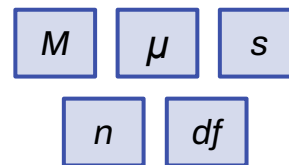
Interpret your results (in APA style).

4 Make your decision.



Our t-score is **4.71**, which falls *in* the critical region.
This means we **reject the null hypothesis**.

NHST STEPS



1 State the **null** and **alternative hypotheses**.

2 Set your **cutoff score** (find t_{crit}).

*Use a **t-table**!*

3 Calculate your **test statistic** (i.e., t-statistic) and effect size.

4 Make your **decision**. Reject the null or Fail to reject the null

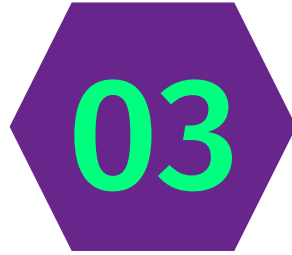
 **5** **Interpret** your results (in APA style).

5 Interpret your results (in **APA style**).

The diagram shows the APA style notation for a t-test result: $t(49) = 4.76, p < .05, d = XX$. Colored arrows point from labels to specific parts of the notation: a purple arrow from 'df' to '(49)', a yellow arrow from 'p-value' to ' $p < .05$ ', an orange arrow from 'effect size' to ' $d = XX$ ', a red arrow from 't-test' to ' t ', a blue arrow from 't-statistic' to '4.76', and a teal arrow from ' α ' to ' $p < .05$ '.

$$t(49) = 4.76, p < .05, d = XX$$

A **one-sample t-test** was conducted to compare the average height of the sample of Tinder males to the average height of U.S. males. Results indicated that the sample of tinder males reported **significantly taller** heights ($M = 71$) compared to the national average ($M = 69$), **$t(49) = 4.76, p < .05, d = XX$** .



Effect Size

Intuition behind effect size

You are trying to improve your sleep quality. You decided to try some off the counter sleep supplements.

You tracked your sleep for a month:



Avg hours slept

6.50

6.60

That's a difference of 0.10 hours — about 6 minutes.

Intuition behind effect size

You are trying to improve your sleep quality. You decided to try some off the counter sleep supplements.

Technically, the supplements improved your sleep.

You tracked your sleep for a month:

But would you actually feel or care about the difference of 6 minutes of extra sleep?

Probably not. It has a tiny effect size.

Avg hours slept

6.50

6.60

That's a difference of 0.10 hours — about 6 minutes.

Just because something is statistically significant, doesn't mean it's meaningful in real life.

We often want to know the *size* or *magnitude* of an effect or difference. We can calculate **effect size** to help us assess the “practical significance” of a finding and go beyond the binary “yes” or “no” of NHST.

Cohen's d is the most common measure of effect size for *t* tests, so we will focus on that one.

Cohen's *d* Formula

(look familiar?)

Notice that the **numerator** is the same as our *z* score and *t* score.

$$d = \frac{\text{sample mean} - \text{population mean}}{\text{standard deviation}}$$

Notice that the **denominator** is **standard deviation** (*not standard error*)

Interpreting Cohen's d

Cohen's d represents the **difference between the two means** in **standard deviations**.

$$d = \frac{M - \mu}{s}$$

Example: I calculate Cohen's d and get **$d = 0.50$** . This means that the population and sample means differ by **0.50 standard deviations**.

Interpreting Cohen's d

$< .20$

very small effect

$.20 - .49$

small effect

$.50 - .79$

medium effect

$\geq .80$

Large effect

3

Compute sample statistics (i.e., t-statistic).

3C Compute your **effect size**.

sample mean	population mean	sample SD	sample size	degrees of freedom
M	μ	s	n	df
71	69	3	50	49

$$d = \frac{M - \mu}{s} = \frac{71 - 69}{3} = 0.67$$

What **size** is our effect?
(very small, small, medium, large)

Remember:
Use **SD** (not
standard error)

5 Interpret your results (in **APA style**).

Diagram illustrating the components of the APA style t-test result: $t(49) = 4.76, p < .05, d = 0.67$. The components are labeled with arrows:

- $t(49)$: t-test (red arrow)
- 49 : df (purple arrow)
- 4.76 : t-statistic (blue arrow)
- $p < .05$: p-value (yellow arrow)
- $d = 0.67$: effect size (orange arrow)
- α (cyan arrow pointing to the p-value)

A **one-sample t-test** was conducted to compare the average height of the sample of Tinder males to the average height of U.S. males. Results indicated that the sample of tinder males reported **significantly different** heights ($M = 71$) compared to the national average ($M = 69$), $t(49) = 4.76, p < .05, d = 0.67$.

ICA 9: One Sample t Test

Starbucks claims their new protein foam contain an average of **15** grams of protein per serving. Your team is skeptical and tests a random sample of **30** orders of coffee with one serving of the protein foam to check if the average protein content differs from the expected value of 15g.

You collect data on the 30 random grande orders with the foam and get a mean of [??] grams and a standard deviation of **1.5** grams. **Conduct a one-sample t test to assess if your sample differs from the population mean.**

($\alpha = 0.05$, two-tailed)

$$\mu = ?$$

$$M = ?$$

$$s = ?$$

$$n = ?$$

$$df = ?$$

Team 1



$M = 15.30\text{g}$

Team 2



$M = 14.80\text{g}$

Team 3



$M = 15.50\text{g}$

Team 4



$M = 16.00\text{g}$

Team 5

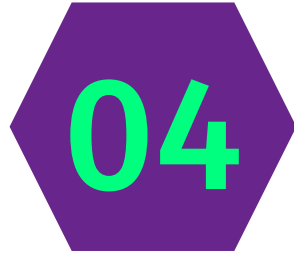


$M = 15.15\text{g}$

Team 6



$M = 14.25\text{g}$



Type I & Type II Errors

Sometimes, we are **wrong**.

Statistics are based on probability, so we sometimes can make the **incorrect conclusion** about our data.

Let's see a watch a quick example to help illustrate this....



A mom says an "everything bagel" caused her to fail a drug test. The hospital reported her to child protective services anyway



Our “**decision**” does not match reality

Reality



Mother eats a bagel (and
does not use opiates)

Test



The test is **positive**
for opiates

Decision



The mother has used
opiates (**not true**)

We made an “**error**” in our conclusion

Type I Error



Mom does not use opiates



Positive drug test result

False Positive

We find a significant effect when really there is not one.

Type II Error



Mom does use opiates



Negative drug test result

False Negative

We did not find a significant effect when really there is one.

A NHST way to look at this

H_0 : The mother does not use opiates.

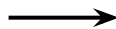
H_a : The mother does use opiates.

False positive (Type I error).



We reject the null, but the null is actually true.

False negative (Type II error).



We fail to reject the null, but it is actually false.

A NHST way to look at this

H_0 : The mother does not use opiates.

H_a : The mother does use opiates.

We'd like to avoid both types of errors...

But in practice, reducing one makes the other more likely. Let's see why.

False negative (Type II error). →

We fail to reject the null, but it is actually false.

Type I Error (α)

Our type I error is **equal to α** (our alpha level)

- Think of it as “seeing something that isn’t there”

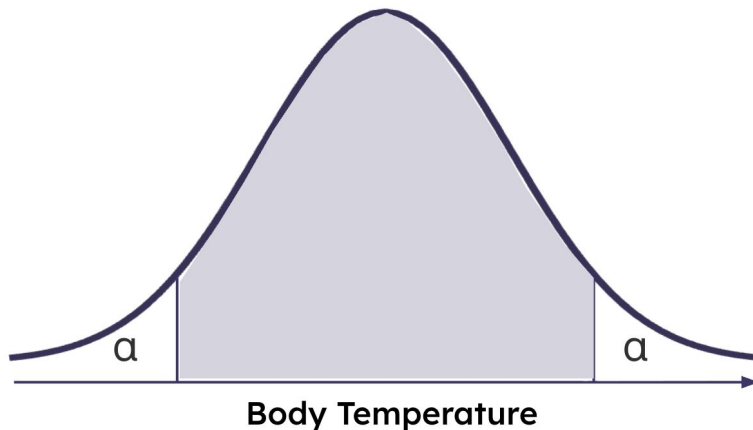
$$\alpha = 0.05$$

We have a **5% chance** of making a type I error.

$$\alpha = 0.01$$

We have a **1% chance** of making a type I error.


Let's say this is our sampling distribution of body temperature:



Imagine we're screening for COVID using body temperature. But people vary — even healthy people. So if we collect samples, we might by chance draw some "hot" people — people whose temperatures are on the extreme right tail of the distribution.

This is what we mean by **Type I error**: If enough of these hot-by-chance people show up in our sample, we might run a test and wrongly conclude there's an effect — like, “aha! they probably have COVID!” — when actually, they're just random weirdos from the tail.

Let's say this is our sampling distribution of body temperature:



As a result, if we shrink our α from 0.05 to 0.01, we are less likely to catch those weirdos, and draw the wrong conclusion.

Imagine we're screening for COVID using body temperature. But people vary — even healthy people have slightly higher body temperatures. "Hot" people — people whose temperatures are on the extreme right tail of the distribution.

Reducing α reduces Type 1 error rate. But...

This is what we mean by **Type I error**: If enough of these hot-by-chance people show up in our sample, we might run a test and wrongly conclude there's an effect — like, "aha! they probably have COVID!" — when actually, they're just random weirdos from the tail.

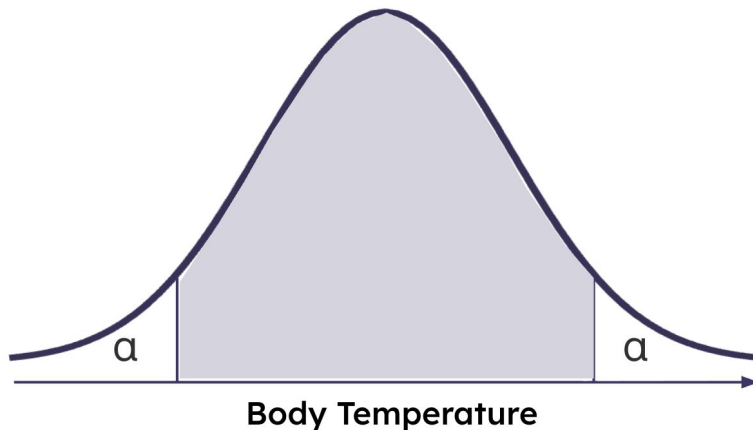
Type II Error (β)

Our type II error is called **Beta (β)**

- Think of it like “fail to detect something that actually exists”
- It is like getting a negative COVID test back even though you actually do have COVID(false negative).
- Higher the Type 1 error, lower the Type 2 error. And vice versa.

This is related to “statistical power**” which is something we will touch on in future weeks.**

Let's say this is our sampling distribution of body temperature:



Sometimes, someone actually has COVID but just doesn't show a high fever — maybe they run cold, or maybe it's early in their illness.

In our sample, we only see normal or low temperatures. So we run our test, and we fail to detect an effect. That's a **Type II error** — we missed something real.





Let's say this is our sample distribution of body temperature:
The key thing is —
No matter how good a test is, it will make mistakes,
simply by chance.

This is why we need to care about how often these
“by chance” events happen, so we don’t draw the
wrong conclusions from our data.





Sometimes, someone actually has COVID but just doesn't show a high fever — maybe they run cold, or maybe it's early in their illness.

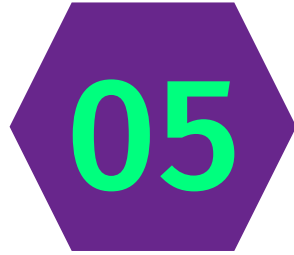
α is the rule we set for how often we're willing to risk
a false alarm (Type I error). But setting α also affects
our ability to avoid misses (Type II error). This
trade-off is why the choice of α matters.

All Possible NHST Decisions

		REALITY	
		Null is True	Null is False
OUR DECISION	Reject Null	 Type I Error	 Correct
	Fail to Reject Null	 Correct	 Type II Error

All Possible NHST Decisions

		REALITY	
		Null is True	Null is False
OUR DECISION	Reject Null	 False Positive	 True Positive
	Fail to Reject Null	 True Negative	 False Negative



Wrap Up

Key Takeaways

1. A **one-sample t test** compares a sample to a population with a known mean and unknown standard deviation.
2. With t tests, we calculate **estimated standard error**, **degrees of freedom**, and our **t statistic** and go through our NHST steps.
3. Effect size (e.g., **Cohen's d**) helps us measure *how much* of a difference there is; it is measured in SDs.
4. We can make the incorrect conclusion with NHST, including **Type I errors** (false positives) and **Type II errors** (false negatives).

Announcements

- Your exam 1 grade will be posted by Friday
- Our Friday lab will be focused on going through Exam 1, remaining time will be practice for this week's content
- Homework 3 due this Sunday at midnight