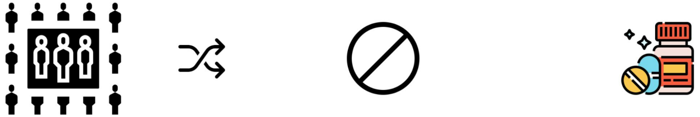
Independent Samples t-test
(part 2)

Lecture 11

Emma Ning, M.A.

From our last lecture...

- Independent samples t-test
  - What it does



  - Compare to z-test & one-sample t-test

	z-test	One-sample t-test	Independent Samples t-test
μ	✅	✅	❌
σ	✅	❌	❌
M	✅	✅	✅
s	✅	✅	✅

  - How to calculate
  - Pooled variance/SD/SE
  - NHST steps

$$t = \frac{M_1 - M_2}{s_{M_1 - M_2}} = \frac{M_1 - M_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

From our last lecture...

- Independent samples t-test
    - What it does
    - Compare to z-test & one-sample t-test
    - How to calculate
    - Pooled variance/SD/SE
    - NHST steps

If you think of an independent-samples t-test as a decision making process...

✅we know if two groups from our sample are different more than chance

🔁but remember, we don't care about samples, we want to use that as a snapshot to say something about the population, right?

TODAY'S PLAN

01

Confidence Interval

02

Statistical Power &
t-test Assumptions

03

Worked Example

04

Wrap Up

Learning objectives

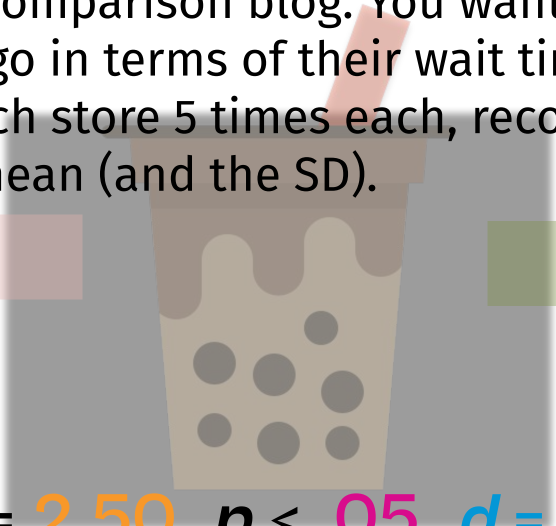
- **Describe** what a confidence interval (CI) is and **explain** why we use these in addition to statistical significance
- Use the data from two separate samples to **compute a 95% CI** describing the size of the mean difference
- **Define** statistical power and **explain** how it relates to the probability of a Type II error
- **Identify** the factors that influence **power** and explain how power is affected by each factor
- **Know and understand** why we need **assumptions** for all statistical tests

01

Confidence Interval

Recap: What we learned so far

Let's say you are very into boba. You not only love boba, but you run a data-driven boba comparison blog. You want to compare the 2 boba shops in Chicago in terms of their wait time to get a drink. You have been to each store 5 times each, recorded the wait times, and calculated the mean (and the SD).

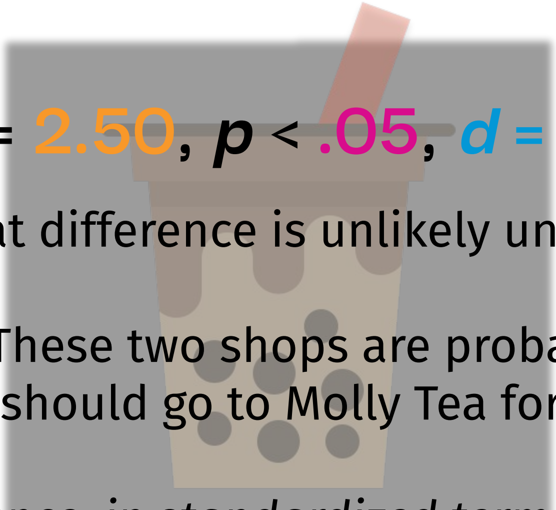


	Molly Tea	**Gathers Tea Bar**
Average wait time (mins)	4 mins	9 mins

$$t(8) = 2.50,\ p < .05,\ d = 0.40$$

Recap: What we learned so far

	Molly Tea	**Gathers Tea Bar**
Average wait time (mins)	4 mins	9 mins



$t(8) = 2.50,\ p < .05,\ d = 0.40$

- Since p < .05, that difference is unlikely under the null → we reject the null.
- → So we infer: "These two shops are probably different in the population, you should go to Molly Tea for faster service!"

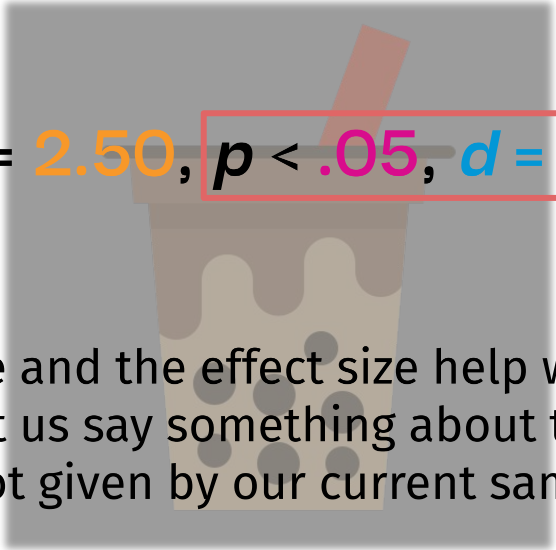
How big is the difference, in standardized terms (aka pooled SD units)?

- A 0.4 SD difference is moderate — meaningful but not huge.

Inference from samples

	Molly Tea	Gathers Tea Bar
Average wait time (mins)	4 mins	9 mins

$t(8) = 2.50$, $\boxed{p < .05,\ d = 0.40}$



- Both the *p*-value and the effect size help with **inference**
  - Aka, they let us say something about the population, based on the snapshot given by our current sample

This is all good… However, you really want to run a good blog and not be called a bad scientist. So there's more considerations.

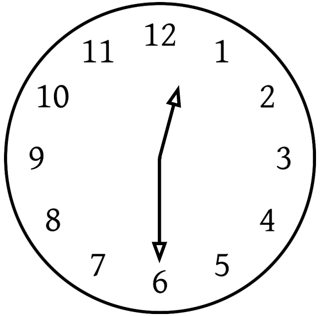
You also need to make sure your conclusion is:

- **Precise**
- **Trustworthy (no fine print?)**
- **Do you have enough data to notice a real difference**

How precise is the mean difference?

We can calculate the **confidence interval (CI)** of the mean difference.

Rather than only focused on retaining/rejecting the null, I can just describe this difference.



H_0: The true time is exactly 7:00 pm.
H_1: The true time is not 7:00 pm.

If you reject the alternative hypothesis, do you know what time it actually is?

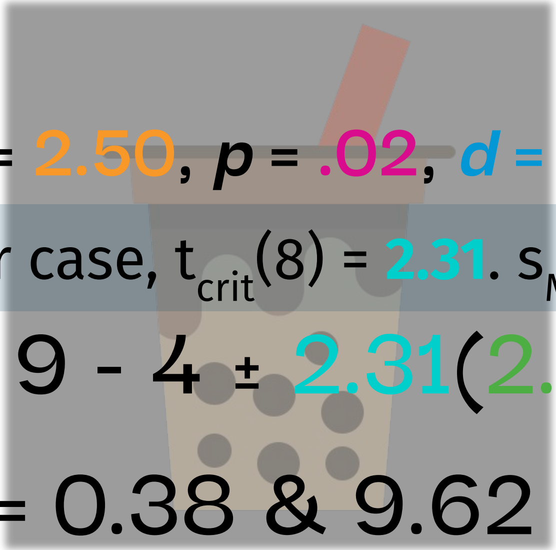
How precise is the mean difference?

$$CI = M_1 - M_2 \pm t_{crit}(s_{M1\text{-}M2})$$

This gives us 2 numbers (one where you added and one where you subtracted). They correspond to the lower and upper bounds.

Example

	Molly Tea	Gathers Tea Bar
Average wait time (mins)	4 mins	9 mins



$t(8) = 2.50$, $p = .02$, $d = 0.40$

In our case, $t_{crit}(8) = 2.31$. $s_{M1-M2} = 2$.

$$CI = 9 - 4 \pm 2.31(2.00)$$

$$CI = M_1 - M_2 \pm t_{crit}(s_{M1-M2})$$

$$= 0.38 \ \& \ 9.62$$

95% CI [0.38, 9.62]

APA STYLE FOR CIs

df p-value Cohen’s d

$t(8) = 2.50$, $p > .05$, $d = 0.40$, 95% CI [0.38, 9.62]

t-test t-statistic α Confidence Interval

SIGNIFICANCE & CIs

If the confidence interval **contains (or "passes") zero**, then we can conclude there is **not** a significant difference/effect.

Significant ($p < 0.05$)

95% CI [2.45, 5.21]

does not contain zero

Non-Significant ($p > 0.05$)

95% CI [-0.82, 1.23]

contains (crosses) zero

Note: And yes, even though it may seem somewhat "redundant" to the p-value, it's important to report the 95% confidence interval.

SIGNIFICANCE & CIs

If the confidence interval **contains (or "passes") zero**, then we can conclude there is **<u>not</u>** a significant difference/effect.

Significant ($p < 0.05$)

95% CI [2.45, 5.21]

does not contain zero

Non-Significant ($p > 0.05$)

95% CI [-0.82, 1.23]

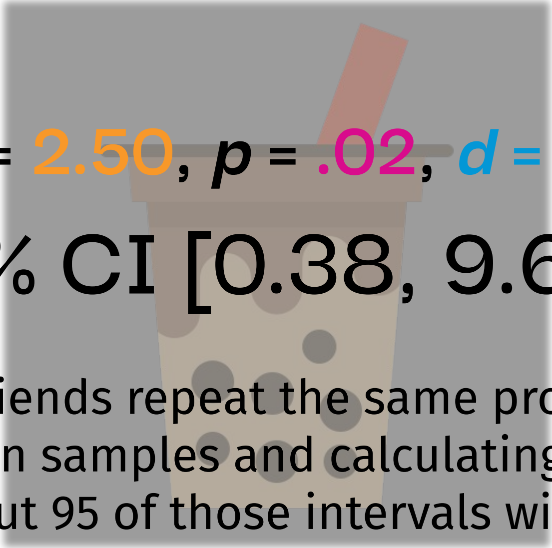
contains (crosses) zero

Note: And yes, even though it may seem somewhat "redundant" to the p-value, it's important to report the 95% confidence interval.

So what does the CI actually mean?

Meaning of 95% CI

	Molly Tea	Gathers Tea Bar
Average wait time (mins)	4 mins	9 mins



$t(8) = 2.50$, $p = .02$, $d = 0.40$

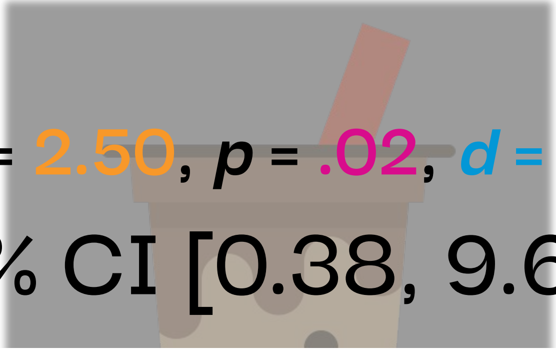
95% CI [0.38, 9.62]

If each of your 100 friends repeat the same process you went through — collecting their own samples and calculating their own confidence intervals — then about 95 of those intervals will contain the true average wait time difference between the two shops.

The more formal, accurate way to conceptualize CI

Meaning of 95% CI

	Molly Tea	Gathers Tea Bar
Average wait time (mins)	4 mins	9 mins



$t(8) = 2.50, p = .02, d = 0.40$

95% CI [0.38, 9.62]

More intuitive:
Think of a confidence interval as the range of values that are **reasonable guesses for the true effect**, given the data you happened to collect. It's saying: 'Based on what we saw, the real difference could realistically be anywhere from X to Y.' Anything outside that range is unlikely to be the true value, because it doesn't line up well with your sample.

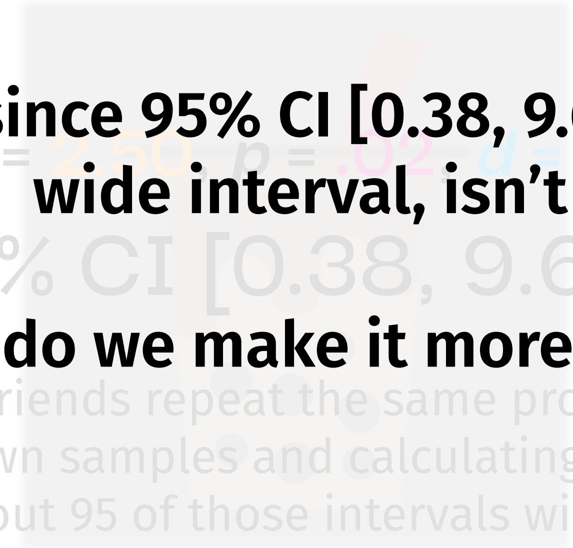
The confidence interval gives you a sense of what the real, long-term difference might be — if you had unlimited time and tracked everyone’s wait.

This is the essence of inference: using a small sample to say something about the bigger picture.

Of course, *p*-value and effect size both serve as tools of inference, but 95% CI goes further by telling you the range of values that are plausible for the population difference.

Meaning of 95% CI

	Molly Tea		Gathers Tea Bar
Average wait time (mins)	4 mins		9 mins



$t(8) = 2.50$, $p = .02$, $d = 0.40$

95% CI [0.38, 9.62]

In this case, since 95% CI [0.38, 9.62], that's a quite wide interval, isn't it?

How do we make it more precise?

If each of your 100 friends repeat the same process you went through — collecting their own samples and calculating their own confidence intervals — then about 95 of those intervals will contain the true average wait time difference between the two shops.

Learning to find clues in the equation!

Let's go through this equation together! It's your secret recipe book:

$$CI = \boxed{M_1 - M_2} \pm \boxed{t_{crit}}(\boxed{s_{M1-M2}})$$

The sample mean difference don't change.

Hmm, this is the t-crit from the t-table with df = 8.
If this number is **smaller**, our interval is **narrower**.

Similarly, if this number is **smaller**, our interval is **narrower**.

	Proportion (a) in One tail				
	.05	.025	.01	.005	.0005
	Proportion (a) in Two tails combined				
df	.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

As df goes up, our t-crit value gets smaller!

And df = n - 2 for independent samples t-test.

So, we need to increase n!

Similarly, for our SE of mean difference...

$$\boxed{s_{M1 - M2}} = \sqrt{\frac{s_p^2}{\boxed{n_1}} + \frac{s_p^2}{\boxed{n_2}}}$$

So we want anything under the square root to be small.

We want this number to be small for a narrower 95% CI.

For each division, if we keep s_p^2 the same, increasing n would give us smaller numbers.

Learning to find clues in the equation!

Let's go through this equation together! It's your secret recipe book:

$$CI = \boxed{M_1 - M_2} \pm \boxed{t_{crit}} (\boxed{s_{M1-M2}})$$

So, the solution for narrowing our 95% CI is pretty consistent – increase our sample size!

5 visits per group is not enough!

The sample mean difference don't change.

Hmm, this is the t-crit from the t-table with df = 8. If this number is **smaller**, our interval is **narrower**.

Similarly, if this number is **smaller**, our interval is **narrower**.

02

Statistical Power & t-test Assumptions

What else you need to not be called a bad scientist

✅Precise (95% CI)
❌Trustworthy (Assumptions)
❌Do you have enough data to notice a real difference

What is Power?

Power is the ability to detect a **real** effect.

TYPE I ERROR	
False Positive	α
We reject the null hypothesis but it is actually true.	

TYPE II ERROR	
False Negative	β
We fail to reject the null hypothesis but it is actually false.	

What is Power?

Power is the ability to detect a **real** effect.

TYPE II ERROR

False Negative	β

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

Since β represents:

- When there is a real effect
- But we didn't find it

Power = 1 - β, representing:

- When there is a real effect
- We found it!

What is Power?

Power is the ability to detect a **real** effect.

TYPE II ERROR	
False Negative	β
We fail to reject the null hypothesis but it is actually false.	

If Power ↓, Type II Error ↑

If Power ↑, Type II Error ↓

Essentially, power is about making sure your study is strong enough to detect a real effect — if one exists.

That often means trying to get the largest sample size you reasonably can, within your budget and effort.

Because if you don’t find an effect, you don’t want to be left wondering, “Oops… maybe I just didn’t try each bubble tea place enough.”
(That’s what happens when your sample size is too small — aka, low power.)

You want to be confident that if you didn’t find a difference, it’s because there really isn’t one — not because your study wasn’t strong enough to detect it.

WHAT **INCREASES** OUR **POWER**?

↑ Sample Size (*n*)

The *most common* way we increase our power is by **increasing our sample size**.

↑ Effect Size (*d*)

Studies with **larger effect sizes** will be more powerful because it is easier to detect “big” differences.

↑ Alpha (*α*)

If we **raise our alpha level** (e.g., from 0.05 to 0.10), our power will also increase.

Use One-Tailed Test

One-tailed tests are more powerful than two-tailed tests.

WHAT INCREASES OUR POWER?

There's a specific type of analysis called "<u>Power Analysis</u>" to determine the minimum sample size required to detect an effect.

You won't be asked to do it. However, you will need to know the sample size needs to big enough for our tests to be <u>powerful</u> to detect the effects.

And you should be able to explain what power is!

↑ Effect Size (*d*)

Studies with **larger effect sizes** will be more powerful because it is easier to detect "big" differences.

Use One-Tailed Test

One-tailed tests are more powerful than two-tailed tests.

What else you need to not be called a fraud...

✅Precise (95% CI)
❌Trustworthy (Assumptions)
✅Do you have enough data to notice a real difference

Making trustworthy inferences

Even though we didn’t talk about it, every statistical test, from z-test, one-sample t-test, to independent samples t-test, comes with its own **assumptions**.

The assumptions are there to make sure the **inferences** we draw are **trustworthy** and **scientifically backed.**

And as we move from the z-test to the one-sample and then to the independent samples t-test, the assumptions become **increasingly more complex**, since we are using sample to estimate population statistics.

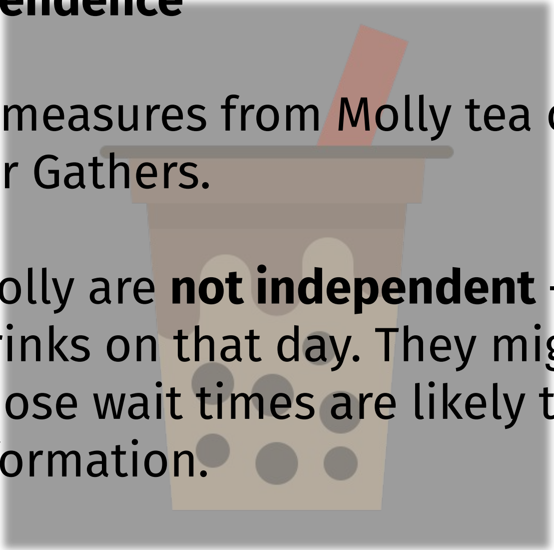
Assumptions for the Independent Samples *t* Test

1. The observations within each sample must be **independent**.
2. The two populations from which the samples are selected must be **normal**.
3. The two populations from which the samples are selected must have **equal variances** *(homogeneity of variances).*

Making assumptions more intuitive

Assumption #1: Independence

Imagine we take all 5 measures from Molly tea on the same day, but do it on different days for Gathers.

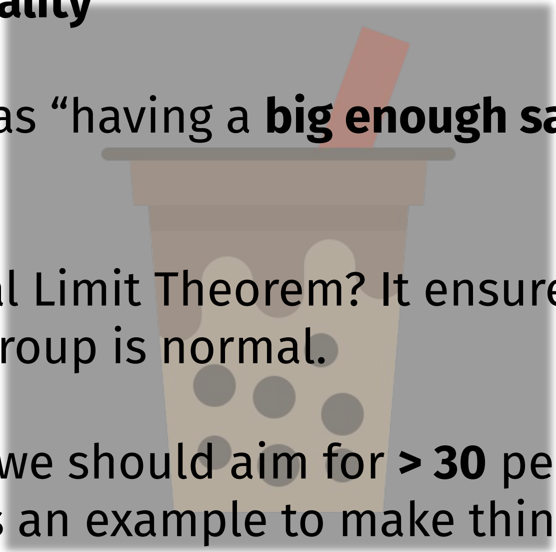


The data points for Molly are **not independent** – the same staff might be handling all our drinks on that day. They might be consistently faster or slower. So those wait times are likely to be similar, not truly separate pieces of information.

We want to make sure there are no associations or hidden connections between our data points.

Making assumptions more intuitive

Assumption #2: Normality



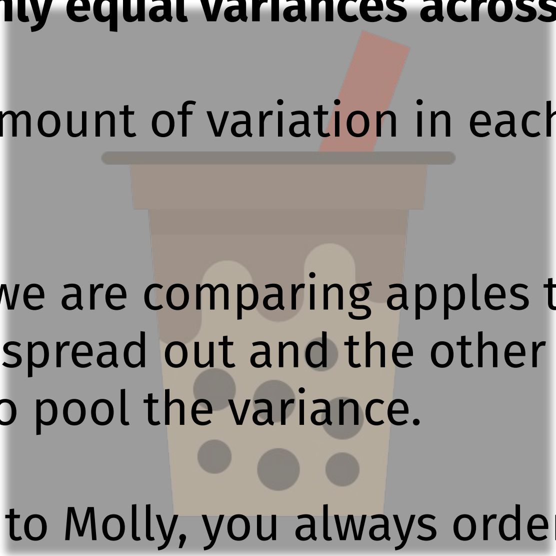
You can think of this as “having a **big enough sample size** in each group”.

Remember the Central Limit Theorem? It ensures that our sampling distribution in each group is normal.

This also means that we should aim for **> 30** people per group. Our tea bar example is just as an example to make things simple. But in real research, small samples can make your results unreliable.

Making assumptions more intuitive

Assumption #3: Roughly equal variances across groups



This just means the amount of variation in each group should be about the same.

This just makes sure we are comparing apples to apples — if one group has data that's super spread out and the other is tightly clustered, it doesn’t make sense to pool the variance.

Imagine when you go to Molly, you always order a simple fruit tea where the prep time is consistently fast, whereas when you go to gathers, you always order a custom matcha with added toppings, and employees have to check if you have all your toppings.

Assumptions for the Independent Samples *t* Test

1. The observations within each sample must be **independent**.
2. The two populations from which the samples are selected must be **normal**.
3. The two populations from which the samples are selected must have **equal variances** (*homogeneity of variances*).

There are ways to test if these assumptions hold, but for this class, we won't be talking about them.

However, you should be able to know what the assumptions are, and explain to others why we need them in the context of <u>inference</u>.

03

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) <u>and **effect size**</u>.

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

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

We are interested in whether there's actually a wait time difference between Molly Tea and Gathers Tea Bar. Since our previous 5 people per group is way too underpowered, you went to Molly Tea for a total of 32 times, and Gathers for a total of 40 times, over the course of 2 years. You hypothesize that the wait time is different (implying a two-tailed test).

Molly Tea	Gathers
M = **6.2**	M = **9.2**

$$s_p^2 = 1.71$$

$$s_{M1 - M2} = 0.31$$

STEP 0: Annotate Your Problem

sample 1 mean	sample 2 mean	total sample size	degrees of freedom	pooled variance	estimated standard error
M_1	M_2	n	df	s_p^2	$s_{M1 - M2}$
6.2	9.2	72	?	1.71	0.31

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) <u>and **effect size**</u>.
4. Make your **decision**. *Reject the null* or *Fail to reject the null*
5. **Interpret** your results (in APA style).

STEP 1: State Hypotheses

Null Hypothesis

H_0

There is **no difference** in wait time between the two stores

$\mu_1 = \mu_2$

Alternative Hypothesis

H_A

There **is a difference** in wait time between the two stores

$\mu_1 \neq \mu_2$

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) <u>and **effect size**</u>.

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

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

STEP 2: Find Cutoff Score (t_{crit})

	Proportion (a) in One tail				
	.05	.025	.01	.005	.0005
	Proportion (a) in Two tails combined				
df	.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:

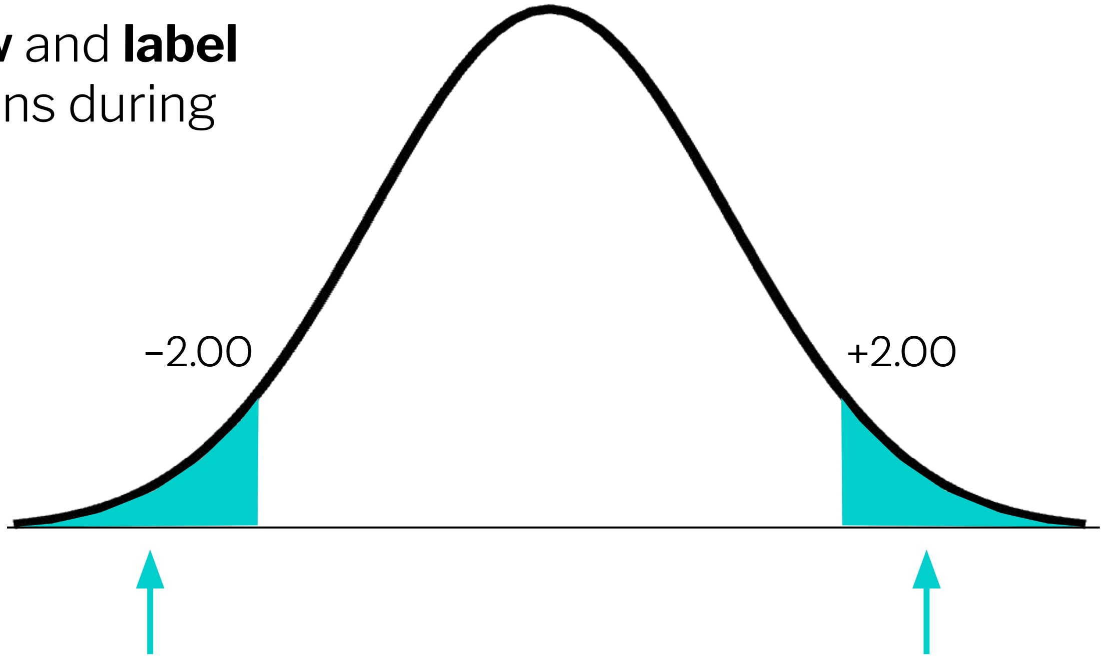
- **df** = 70
- **α** = 0.05
- **two**-tailed test

Our t_{crit} is +**2.00**.

STEP 2: Find Cutoff Score (t_{crit})

Always **draw** and **label** critical regions during this step!

$t_{crit} = \pm\mathbf{2.00}$



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) <u>and **effect size**</u>.

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

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

STEP 3: Calculate *t*, *d*, and 95% *CI*

sample 1 mean	sample 2 mean	total sample size	degrees of freedom	pooled variance	estimated standard error
M_1	M_2	n	df	s_p^2	$s_{M1 - M2}$
6.2	9.2	72	70	1.71	0.31

3A Compute you **t statistic.**

$$t = \frac{M_1 - M_2}{s_{M1 - M2}} = \frac{9.2 - 6.2}{0.31} = \boxed{9.68}$$

3B Compute your ***effect size*.**

$$d = \frac{M_1 - M_2}{\sqrt{s_p^2}} = \frac{9.2 - 6.2}{\sqrt{1.71}} = \boxed{2.29}$$

STEP 3: Calculate *t*, *d*, and 95% *CI*

3C Calculate your **95% confidence interval**.

$$95\%\ CI = M_1 - M_2 \pm t_{crit}(s_{M1-M2})$$

$$= (9.2 - 6.2) \pm 2.00(0.31)$$

$$= 3 \pm 0.62$$

$$= 2.38\ \&\ 3.62$$

95% CI [2.38, 3.62]

NHST STEPS

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

2. Set your **cutoff score** (find t_{crit}). *Use a **t-table**!*

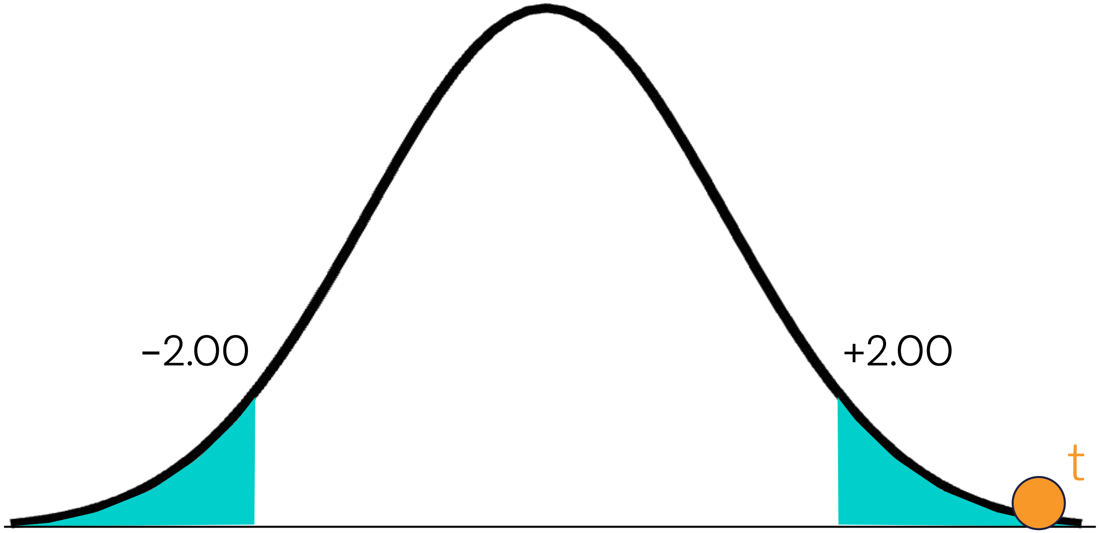
3. Calculate your **test statistic** (i.e., t-statistic) <u>and **effect size**</u>.

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

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

STEP 4: Make Our Decision About Null

t_{crit} = ±**2.00**



Our t statistic is in the critical region, so 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) <u>and **effect size**</u>.

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

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

STEP 5: Write results in APA style.

Results from an **independent samples t-test** indicated that **Molly Tea** (M = 6.2) had **significantly shorter** wait time than **Gathers** (M = 9.2) with a large effect size, $\boldsymbol{t(72) = 9.68, p < .05, d = 2.29}$, **95% CI [2.38, 3.62].**

Important to report the direction!

$$\underset{\text{t-test}}{t}(\overset{\text{df}}{72}) = \underset{\text{t-statistic}}{9.68},\ \overset{\text{p-value}}{p} < \underset{\alpha}{.05},\ \overset{\text{Cohen's d}}{d = 2.29}$$

ICA 11

A behavioral scientist is studying whether taking notes by hand versus on a laptop leads to better performance on a short-answer exam. Two groups of students are randomly assigned:

Handwriting (*n* = 40)	**Laptop (*n* = 50)**
$M_1 = \mathbf{78}$	$M_2 = \mathbf{72}$

$s_p^2 = 411$ $s_{M1-M2} = 4.30$

Conduct an independent samples t test ($\alpha = 0.05$, two-tailed test).

	Proportion (a) in One tail				
	.05	.025	.01	.005	.0005
	Proportion (a) in Two tails combined				
df	.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

CALCULATIONS

3A Compute your **t statistic**.

$$t = \frac{M_1 - M_2}{s_{M1 - M2}} = \frac{78 - 72}{4.30} = \boxed{1.40}$$

3B Compute your ***effect size***.

$$d = \frac{M_1 - M_2}{\sqrt{s_p^2}} = \frac{78 - 72}{\sqrt{411}} = \boxed{0.30}$$

CALCULATIONS

3C Calculate your **95% confidence interval**.

$$95\%\ CI = M_1 - M_2 \pm t_{crit}(s_{M1\text{-}M2})$$

$$= (78 - 72) \pm 2.00(4.30)$$

$$= 6 \pm 8.60$$

$$= -2.6\ \&\ 14.6$$

95% CI [-2.6, 14.6]

INTERPRETATION

Although the difference between student performance between the two conditions was of small magnitude (Cohen's d = 0.30), it was not statistically significant, $t(88) = 1.4$, $p > .05$. The 95% CI [−2.6, 14.6] includes zero, indicating that with this sample we cannot know whether taking notes by hand or typing is more effective.

04

Wrap Up

Key Takeaways

- 95% **Confidence Intervals (CI):**
  - Calculation
  - Interpretation
  - Why it’s useful

- Statistical **Power**
  - Why it is important (and what underpowering results in)
  - Factors that affect it

- Independent samples t-test **assumptions**

UIC PSYCHOLOGY UNDERGRADUATE RESEARCH READINESS PROGRAM

APPLYING TO GRAD SCHOOL
Workshop

Thinking about a graduate degree in psychology but feeling overwhelmed by the application process? Join current UIC Psychology graduate students for a hands-on workshop to learn key strategies for crafting a standout application with step-by-step guidance.

12:00 - 1:00 PM

Friday, Oct. 10th

BSB 1076

go.uic.edu/PURR

Mid Semester Feedback Survey

https://forms.gle/ihVykMZaN5eaNRSt5