Paired Samples t-test
(part 2)

Lecture 13

Emma Ning, M.A.

From our last lecture...

- Within-subjects vs. between-subjects design
    - When to use which?
    - Pros & Cons

- Paired samples t-test

$$t = \frac{M_D}{S_{MD}}$$

TODAY'S PLAN

01

In-Class Experiment!

02

Putting All t-tests Together

03

Going into ANOVA

04

Wrap-Up

Learning objectives

- Collect **real data**, perform a paired-samples t-test, and interpret the findings
- **Compare and contrast** the three types of t tests in terms of their **uses, formulas, and interpretations**
- Determine the appropriate type of t-test to use based on a given research design
- Understand and can explain **family-wise error rate** and the harm of running many t-tests

01

In-Class Experiment!

Today, we are going to **collect our own data** as a class and analyze our results.

We want to know whether our **dominant hand** can significantly outperform our **non-dominant hand**.

To test this, we will draw this star

And record our time that it takes to draw it.

Task Information

You will draw the star **two times**: once with your dominant hand, once with your non-dominant hand. You will record the time it takes to draw the star each time.

We hypothesize that drawing with our dominant hand is significantly faster than drawing with our non-dominant hand.

Experimental Procedure

1	**Get in pairs**.
2	Choose one person to be the **timer** (the other will do the task).
3	Follow the **Randomization order** indicated on the G Sheet. Draw the star and have your partner time you.
4	**Record your time <u>in seconds</u>** on the G Sheet; switch roles and repeat.

[Google Sheet Link]

I C A 1 3

As a table, do a 1) paired sample *t*-test AND 2) independent-samples *t*-test with our class data. Was there a significant difference in the time it took us to draw the star with our dominant vs. non-dominant hand? Is there a difference in conclusion depends on which test? (**α = 0.05**, two-tailed)

paired sample *t*-test

M_D	s_D	s_{MD}	n	df
X	X	X	X	X

Two-sample *t*-test

sample 1 mean	sample 2 mean	<u>total</u> sample size	degrees of freedom	pooled variance	estimated standard error
M_{dom}	$M_{non\text{-}dom}$	n	df	s_p^2	s_{M1-M2}
X	X	X	X	X	X

What do you notice?

Did you reach different conclusions based on which test you ran?

02

Putting All t-tests Together

Intuition Behind All t-tests

All t-tests use a signal-to-noise ratio that takes on this form:

$$t = \frac{\text{difference in means}}{\text{Some form of SE}}$$

This ratio is asking the question: **Is the difference I'm observing bigger than what I'd expect just from random noise?**

At the core of NHST and t-tests is the idea that random noise can make us see effects that aren't really there.

This is really important, understand this sentence well.

Intuition Behind All t-tests

All t-tests use a signal-to-noise ratio that takes on this form:

$$t = \frac{\text{difference in means}}{\text{Some form of SE}}$$

At the core of NHST and t-tests is the idea that random

This is really important, understand this sentence well.

This is why we talked about Type 1 and Type 2 errors, power, alpha-level (0.05), and effect size (Cohen's d).

In essence, our goal is to design studies where we have a high chance of detecting <u>real</u> effects — and a low chance of being misled by noise.

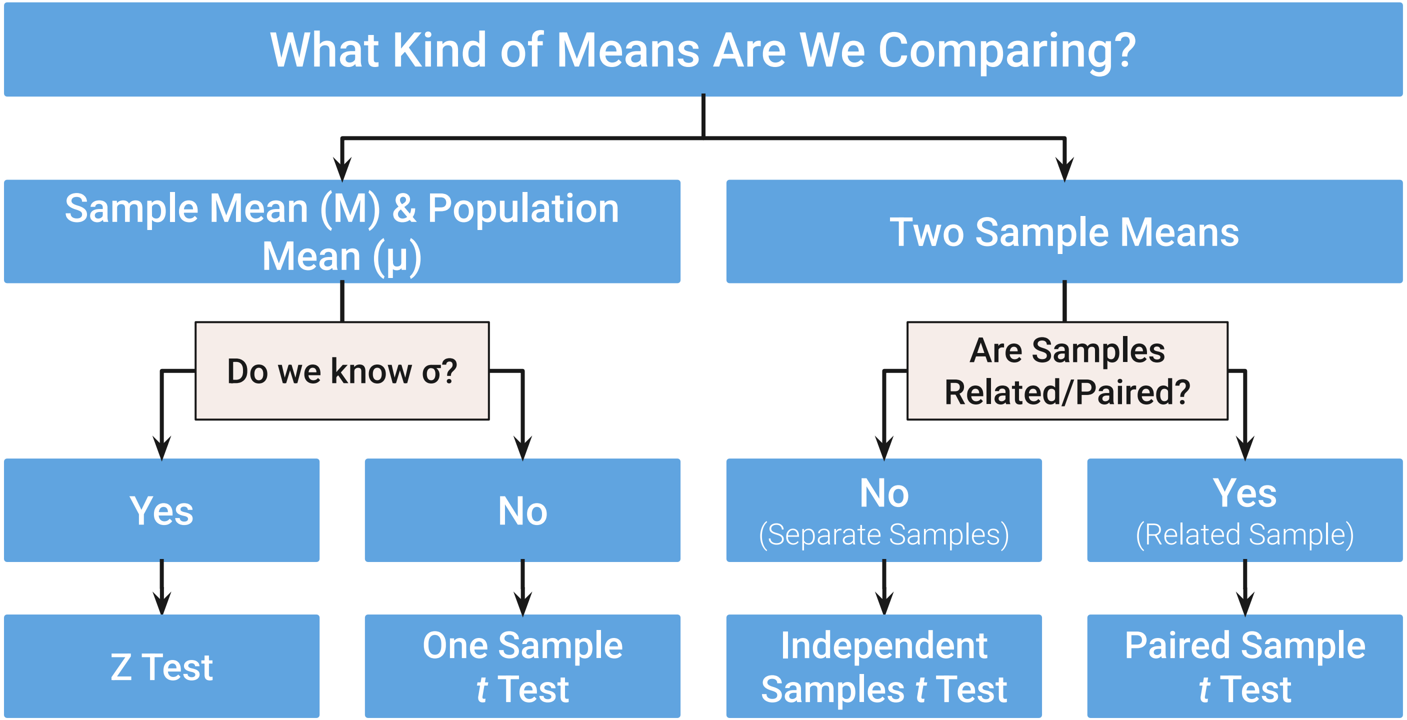
We will come back to this idea later.

Comparing our tests

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

Since you almost never know population statistics, you are almost always dealing with the latter two t-tests.

Which Test to Use?



t statistic and effect size formulas

The *t* Statistic Formula

$$t = \frac{\text{difference in means}}{\text{estimated standard error}}$$

Cohen's *d* Formula

$$d = \frac{\text{difference in means}}{\text{standard deviation}}$$

Keep this formula sheet

Test Statistic	Sample Variance or SD	Estimated Standard Error	*t* statistic	Cohen's *d*	95% CI
One Sample *t* statistic	$s = \sqrt{\frac{SS}{df}}$	$s_M = \frac{s}{\sqrt{n}}$	$t = \frac{M - \mu}{s_M}$	$d = \frac{M - \mu}{s}$	$95\%\ CI = (M - \mu) \pm t_{crit}(s_M)$
Independent Samples *t* statistic	$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$ (given)	$s_{M_1 - M_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$ (given)	$t = \frac{M_1 - M_2}{s_{M_1 - M_2}}$	$d = \frac{M_1 - M_2}{s_p}$	$95\%\ CI = (M_1 - M_2) \pm t_{crit}(s_{M1-M2})$
Paired Sample *t* statistic	$s_D = \sqrt{\frac{SS_D}{df}}$ (given)	$s_{M_D} = \frac{s_D}{\sqrt{n}}$	$t = \frac{M_D}{s_{M_D}}$	$d = \frac{M_D}{s_D}$	$95\%\ CI = M_D \pm t_{crit}(s_{MD})$

Keep this formula sheet

Test Statistic	Sample Variance or SD	Degrees of Freedom
One Sample *t* statistic	$s = \sqrt{\frac{SS}{df}}$	df = n - 1
Independent Samples *t* statistic	$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$ (given)	$df_1 = n_1 - 1$ $df_2 = n_2 - 1$ df = n - 2
Paired Sample *t* statistic	$s_D = \sqrt{\frac{SS_D}{df}}$ (given)	df = n - 1

This is the df you use for the test: you look up your t-table with this number, and you report this number in your APA write-up.

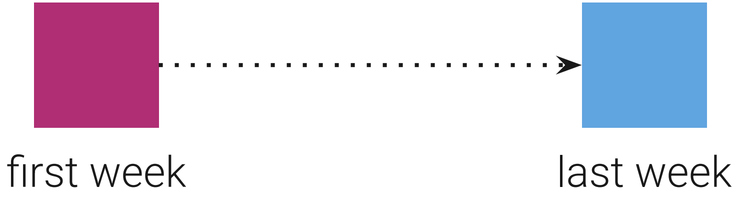
Keep this formula sheet

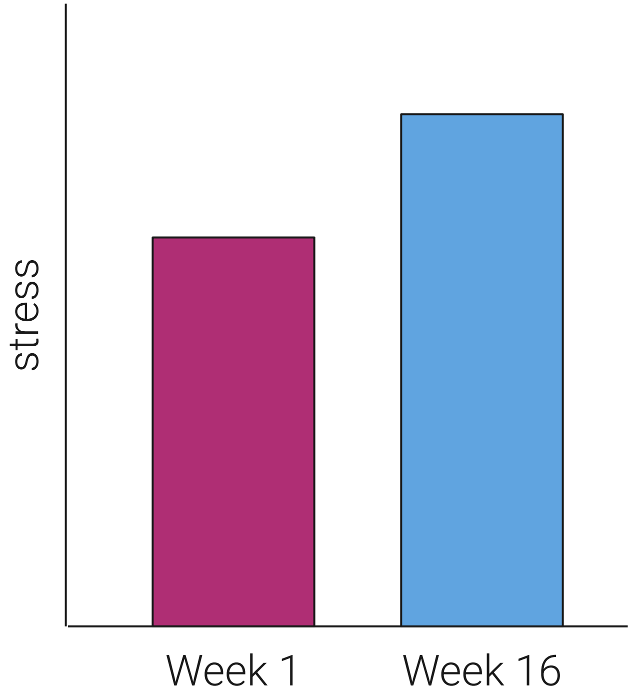
Ultimately, which test to run depends on:

1. **What do we know? (population or sample?)**
2. **Does each person contribute more than 1 data point?**

Example 1

Professor Ning wants to know if students feel more stressed at the beginning of the semester or at the end. She assess students' stress levels the first and last week of classes.





What type of *t* test should she use?

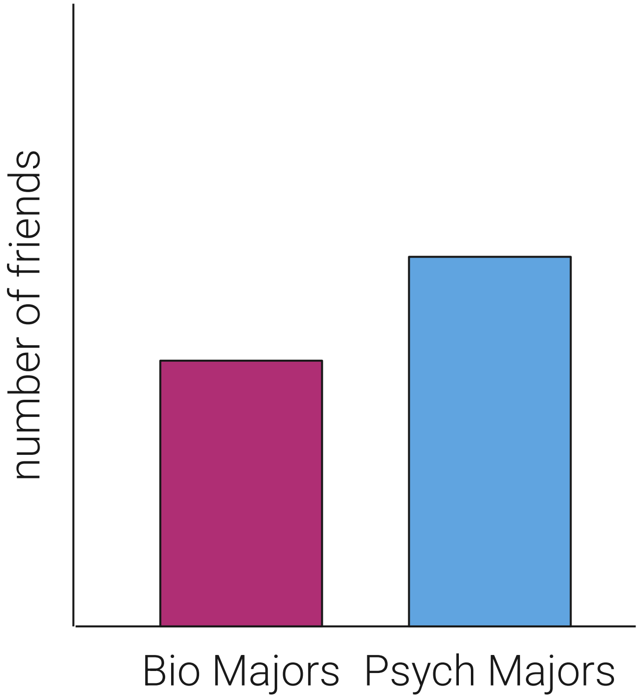
Example 2

A UIC research team wants to know if psychology majors or biology majors have more social connections. They compare the average number of friends for psychology and biology majors.

Sample A

Sample B

What type of *t* test should they use?



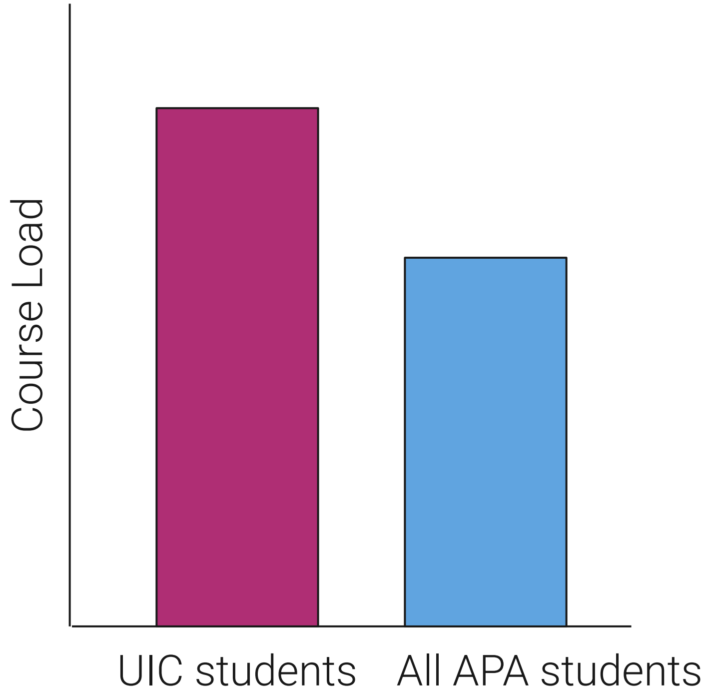
Example 3

UIC clinical program wants to figure out whether clinical students at UIC take more classes than what's needed for APA accreditation. The director compares UIC students course load to everyone else in APA accreditation programs.

Sample

Population

What type of *t* test should they use?



03

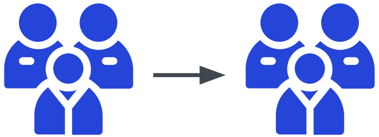
Going into ANOVA

For the past few weeks, we focused on comparing the means between **two groups**.

One Sample *t*-Test

Population Sample

Paired Sample *t*-Test



Sample (time 1) Sample (time 2)

Independent Samples *t*-Test

Sample X Sample Y

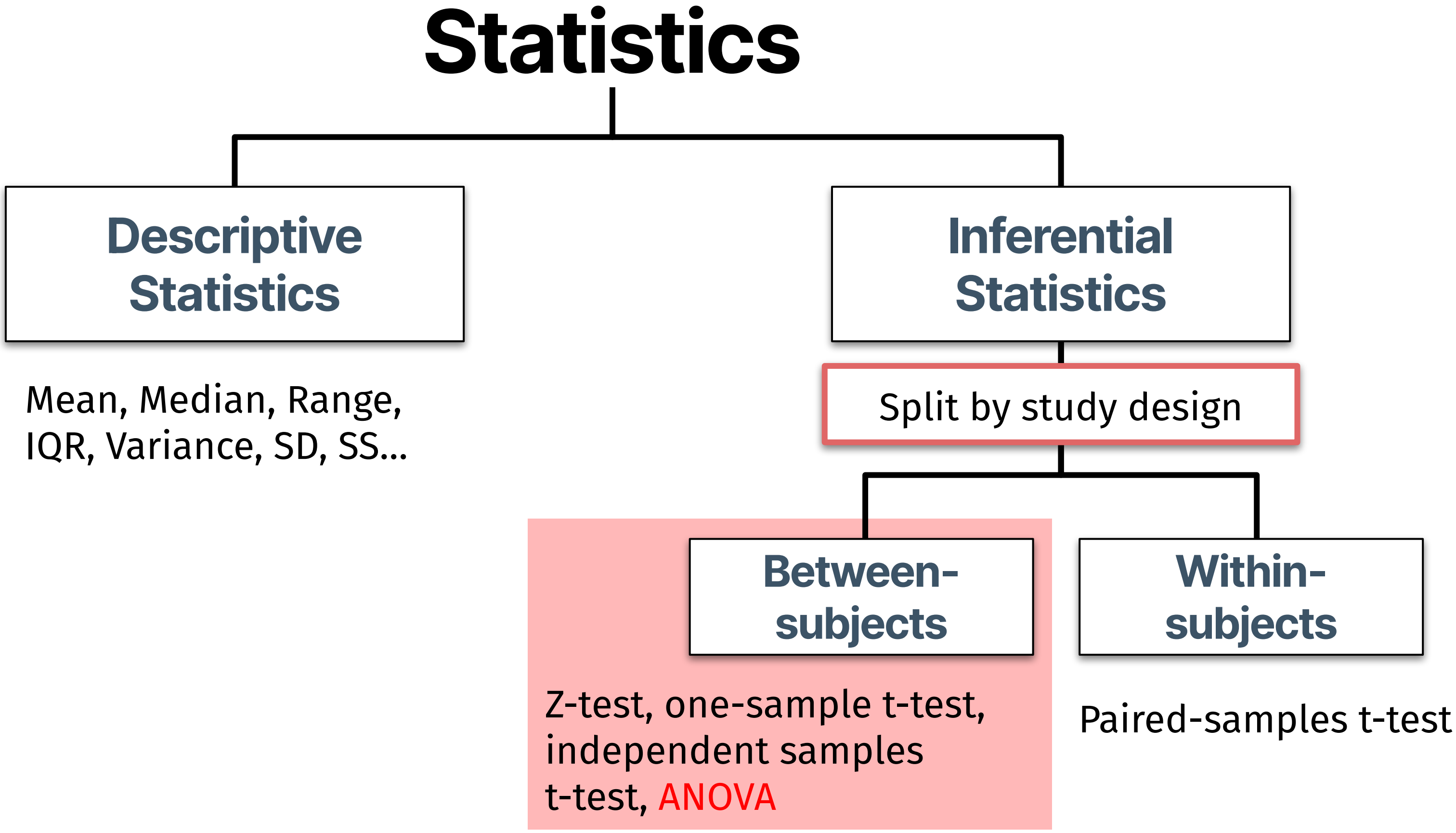
What if we have <u>more</u> than two groups?

E.g., is the wait time different between Gathers, Molly Tea, and Tiger Sugar

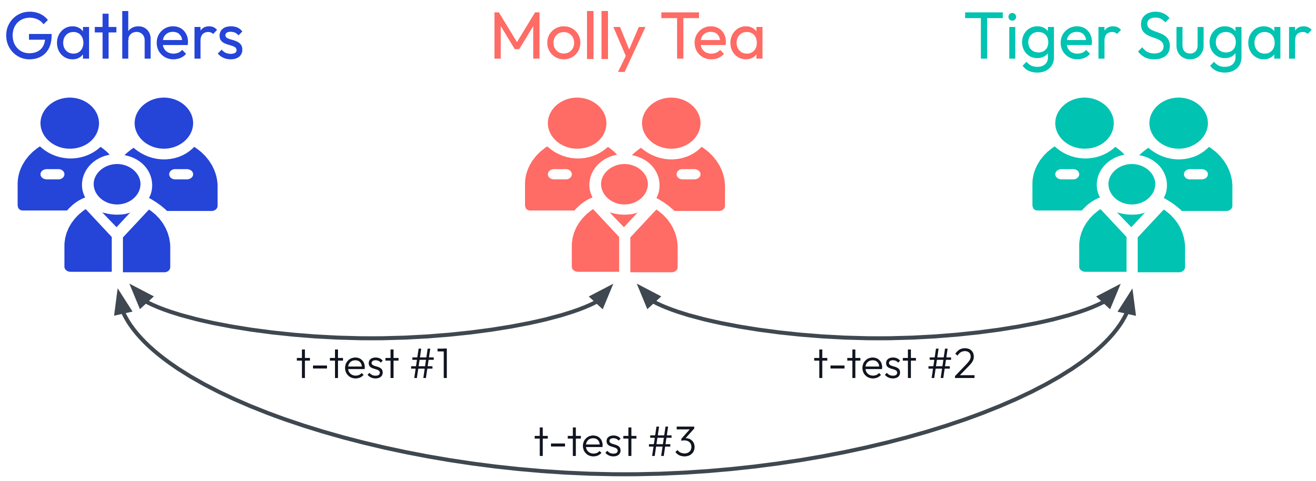
ANOVA

ANalysis **O**f **VA**riance

Where is ANOVA in this map

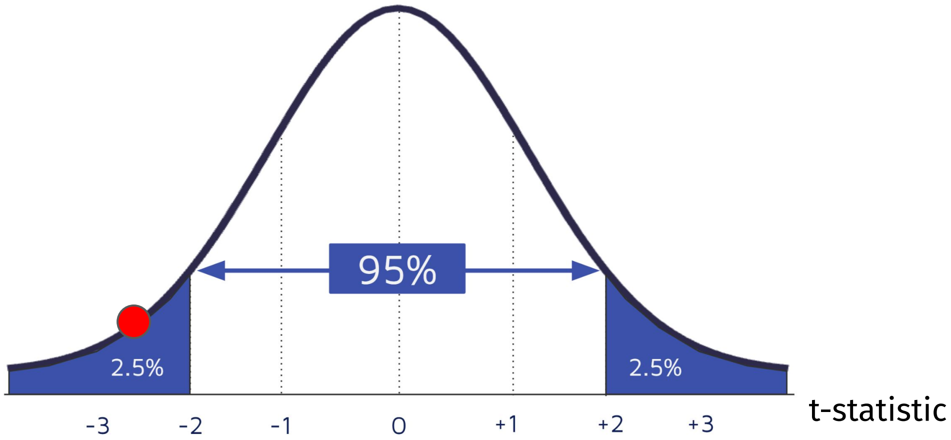


Why can't we just run a lot of *t* tests?



What is a potential issue with doing this?

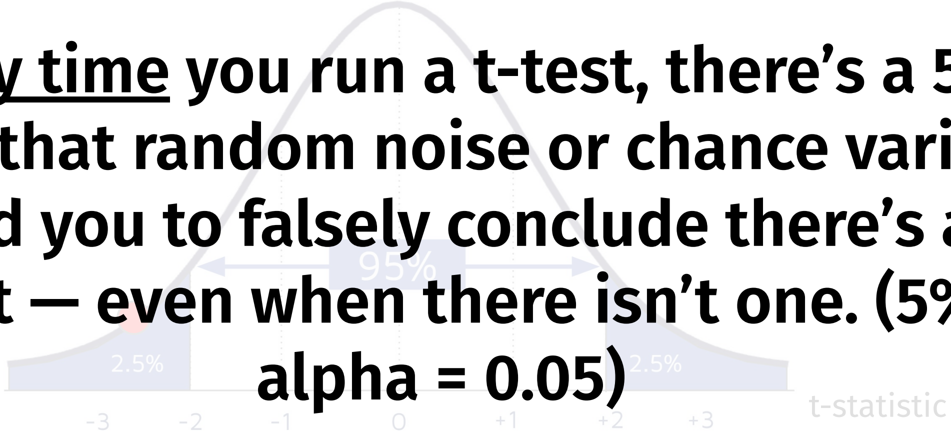
Remember this?



In a two-sample t-test, the observed difference between groups lands somewhere on a normal curve.

Even if there's no real difference (i.e., the null is true), sometimes the observed value lands in the tail, just by chance.

Remember this?



<u>Every time</u> you run a t-test, there's a 5% chance that random noise or chance variation will lead you to falsely conclude there's a real effect — even when there isn't one. (5% if alpha = 0.05)

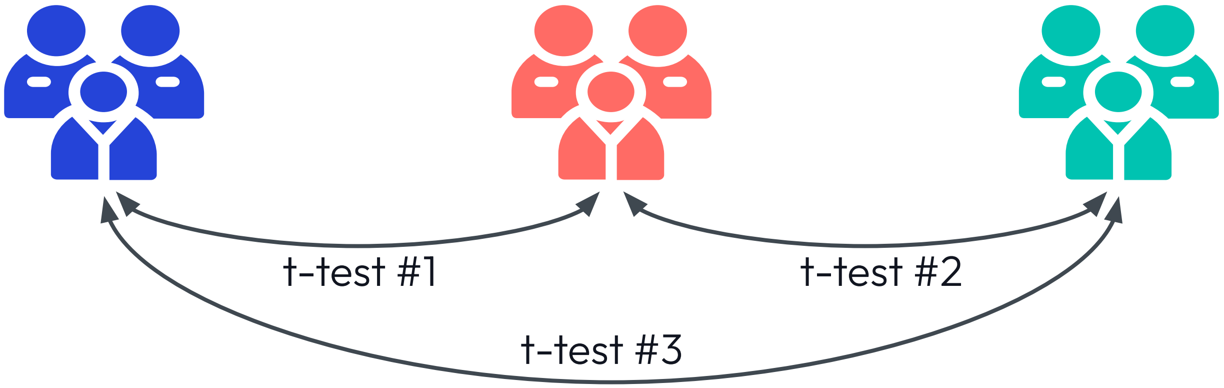
If we run a bunch of t-tests, this adds up.

In a two-sample t-test, the observed difference between groups lands somewhere on a normal curve.

Even if there's no real difference (i.e., the null is true), sometimes the observed value lands in the tail, just by chance.

Family-wise Error Rate

Family-wise error rate (FWER) is the probability of making one or more **false discoveries, or type I errors** when performing **multiple hypotheses tests**.



Family: 3 tests
Alpha: 0.05
FWER: $1 - (1 - 0.05)^3 = 0.143$

(You do not need know how to calculate FWER)

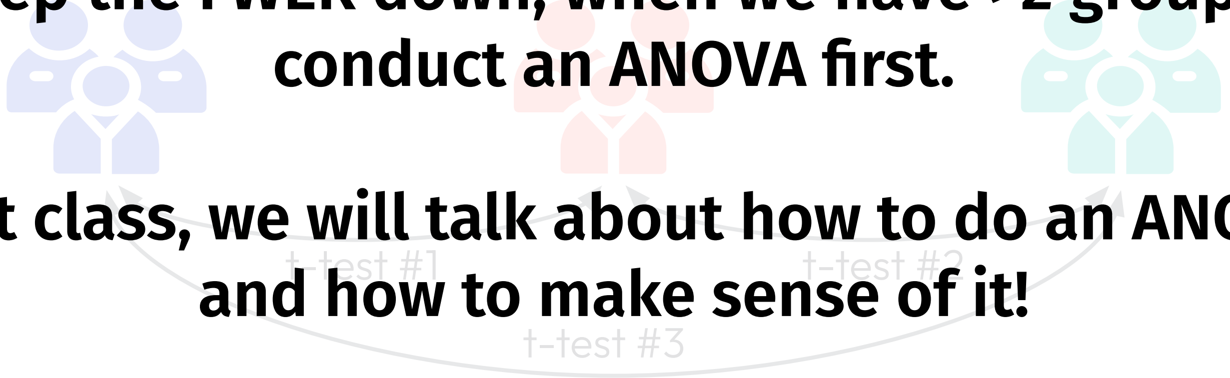
If we conduct 3 separate t-tests, and when there's no real effect, there's about a 14% chance that we'll make at least one false positive — that is, find a "significant" result just due to chance or random noise.

Family-wise Error Rate

Family-wise error rate (FWER) is the probability of making one or more false discoveries, or type I errors when performing multiple hypotheses tests.

To keep the FWER down, when we have >2 groups, we conduct an ANOVA first.

Next class, we will talk about how to do an ANOVA, and how to make sense of it!



Family: 3 tests
Alpha: 0.05
FWER: $1 - (1 - 0.05)^3 = 0.143$

(You do not need know how to calculate FWER)

If we conduct 3 separate t-tests, and when there's no real effect, there's about a 14% chance that we'll make at least one false positive — that is, find a "significant" result just due to chance or random noise.

04

Wrap Up

Key Takeaways

- Collected & analyzed your own data!

- **Compared and contrasted** independent-samples vs. paired-samples t-test**:**
  - Advantage of using paired over independent-samples when you can

- Differentiated when to use which t-test

- Family-wise error rate (FWER) and transition into ANOVA