ANOVA Midterm Review

t tests, One-Way ANOVA, & Factorial ANOVA

PSYC 308b

t-tests

Compares two group means to determine whether they are significantly different.

General

When should you use a t-test?

- One Independent Variable with two levels
- One Dependent Variable

An independent t-test compares:

*differences between two independent samples*

A dependent t-test compares:

*differences between two related samples*

Assumptions

Independent Samples t-test

- The DV is _*continuous*_ & _*normally distributed*_
- The IV is _*categorical*_
- Variance is _*homogenous*_
- Observations are _*independent*_

Dependent Samples t-test

- The DV is _*continuous*_ & _*normally distributed*_
- The IV is _*categorical*_

Steps for Calculating

- TEST FOR HOMOGENEITY OF VARIANCE: *Levene's test*
  - CORRECTION: *Welch*
- TEST OF SIGNIFICANCE: *t-value* & *p-value*
- EFFECT SIZE: *Cohen's d*
  - Small = 0.25
  - Moderate = *0.50*
  - Large = *0.80*
- POST-HOCS: *none*
  - Why? *There are only 2 means*

What is Cohen's d?

The **mean** difference between two groups in standard deviation units.

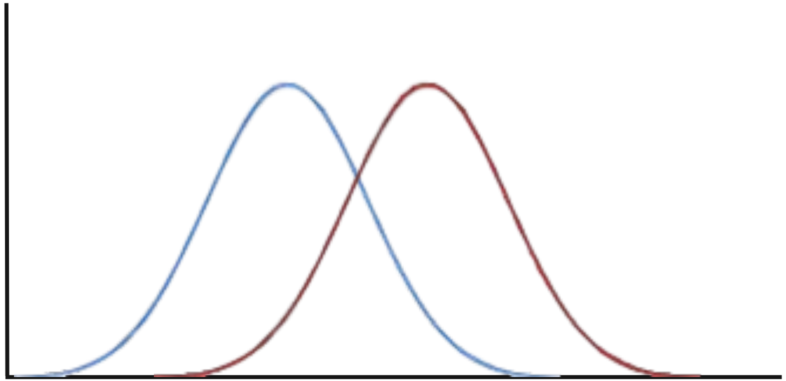
Dependent t-test: $d = M_d / SD_d$

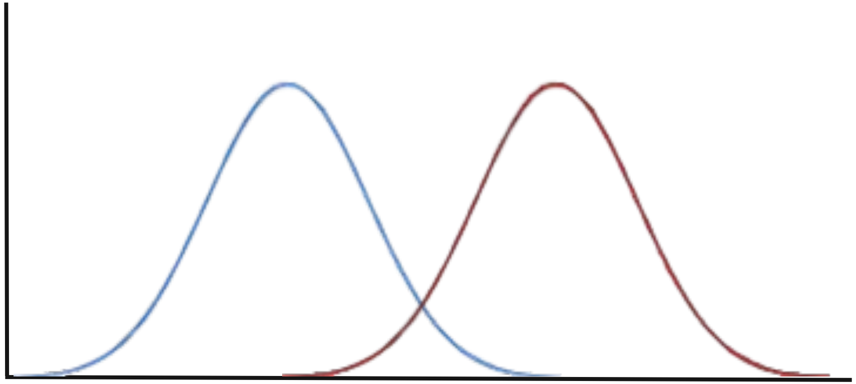
Independent t-test: $d = (M_1 - M_2) / SD_{pooled}$

$SD_{pooled} = (SD_1 - SD_2)/2$

Things that affect significance

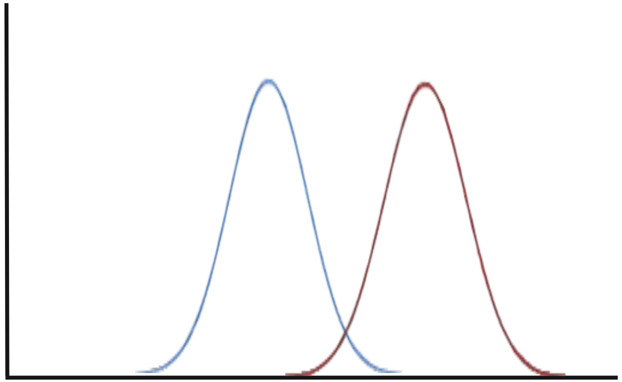
- Differences in group means

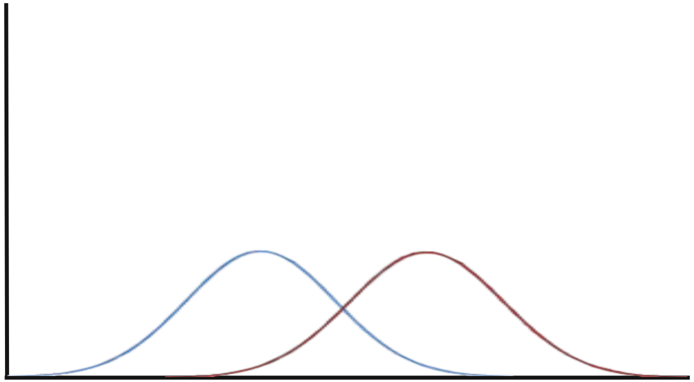




Things that affect significance

- Differences in group means
- Differences in standard deviation





t-tests

Example 1

- You are a math tutor for middle school students. Whenever you get a new student, you give them a pre-test of math problems to establish a base-line. After ten tutoring sessions, you give them a post-test of math problems to measure their improvement.

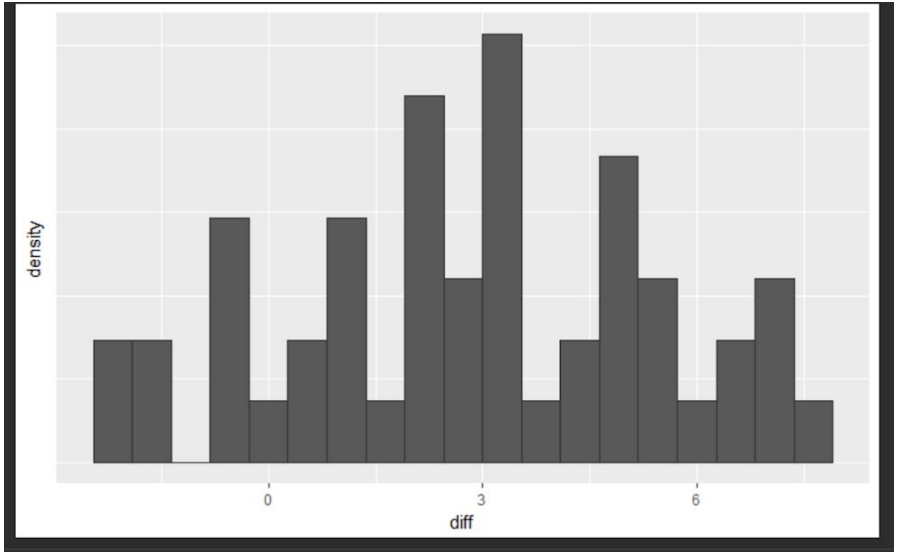
- To build your clientele, you decide to use statistics to show that you are effective!

- Given the above information, what type of test will you run?

Check Assumptions: Normality of the Difference Scores

Descriptives

	diff
N	50
Missing	0
Mean	2.87
Median	2.95
Standard deviation	2.63
Minimum	-2.19
Maximum	7.62
Standard error	0.372
Skewness	-0.139
Kurtosis	2.14



Example 1

PAIRED SAMPLES T-TEST

Paired Samples T-Test

			statistic	df	p	Cohen's d	Lower	Upper
Before_CorrectProblems	After_CorrectProblems	Student's t	-7.71	49.0	< .001	-1.09	-3.61	-2.12

Descriptives

	N	Mean	Median	SD	SE
Before_CorrectProblems	[redacted]	4.98	4.99	2.375	0.336
After_CorrectProblems	[redacted]	7.85	7.75	0.891	0.126

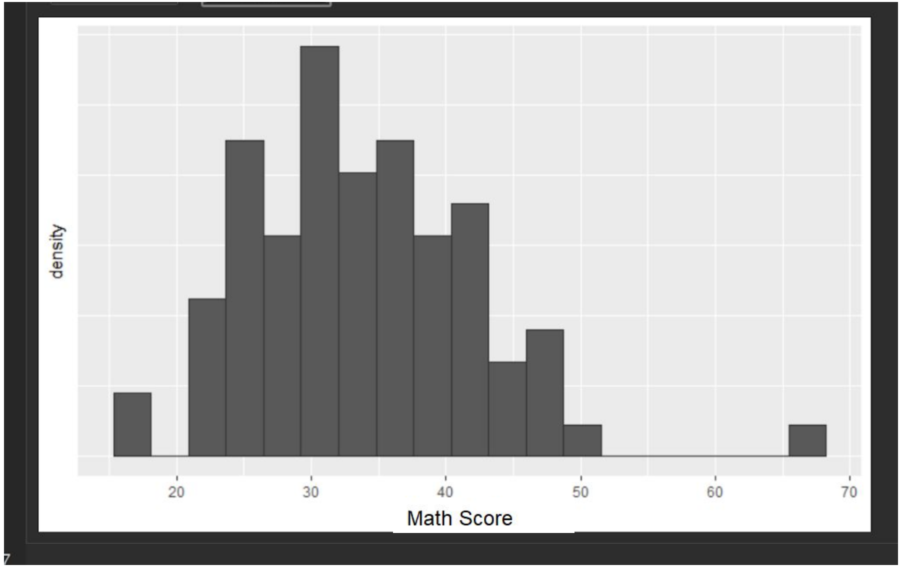
Example 2

- You, the math tutor, realize that there may be some confounding variables that impacted the results from your first study (such as maturation).
- As an effort to gain more evidence of your awesomeness, you decide to conduct a second study. You go to a local middle school and offer to tutor a random sample of one half of a math class all semester for free. At the end of the semester, you give a math test to all students in the class.
- What type of test will you run this time?

Check Assumptions: Normality of the DV

Descriptives

	Math Score
N	80
Missing	0
Mean	33.8
Median	33.2
Standard deviation	8.24
Minimum	17.2
Maximum	67.3
Standard error	0.922
Skewness	0.767
Kurtosis	4.88



Example 2

Independent Samples T-Test

		statistic	df	p	Cohen's d	Lower	Upper
MathScore	Student's t	-5.40	78.0	< .001	-1.21	-11.7	-5.39

ASSUMPTIONS

Test of Equality of Variances (Levene's)

	F	df	p
MathScore	2.29	1	0.134

Note. A low p-value suggests a violation of the assumption of equal variances

Group Descriptives

	Group	N	Mean	Median	SD	SE
MathScore	Cont	40	29.6	29.3	5.41	0.855
	Math	40	38.1	37.8	8.42	1.33

Sample Write-Up

This study examined the effects of tutoring on math scores using an independent samples t-test. There was a significant difference in math scores between the tutoring condition (M = 38.10) and the control condition (M = 29.60), $t(78) = -5.40$, $p < .001$. This was a large effect of tutoring on math scores, $d = -1.21$. The mean of the tutoring condition was 1.21 standard deviations higher that the mean of the control condition.

One-way ANOVA

Compares the means of two or more levels of an independent variable to determine whether they are significantly different.

General

- What is ANOVA?
  - A type of analysis used to test differences among means by analyzing variance.
  - A type of analysis that uses the F distribution
- When should you use a **one-way** ANOVA?
  - One Independent Variable with at least two levels
  - One Dependent Variable

Assumptions

- The DV is ___continuous___ & ___normally distributed___ for each group
- The IV is ___categorical___
- Variance is ___homogenous___
- Observations are ___independent___

Source	SS	df	MS	F
treat	$n\,\Sigma(X_j - X_T)^2$	k - 1	SS_{treat}/df_{treat}	$MS_{treat}\,/MS_{error}$
error	$\Sigma(X_{ij} - X_j)^2$	k(n -1)	SS_{error}/df_{error}	-----
Total	$\Sigma(X_{ij} - X_T)^2$	N - 1	-----	-----

One-way ANOVA

ANOVA: Under the Hood

- But what does it actually do?
  - Tests the null hypothesis that the groups are drawn from populations with the same mean values.
- How does it do that?
  - It estimates two different kinds of variance and compares them

What is F?

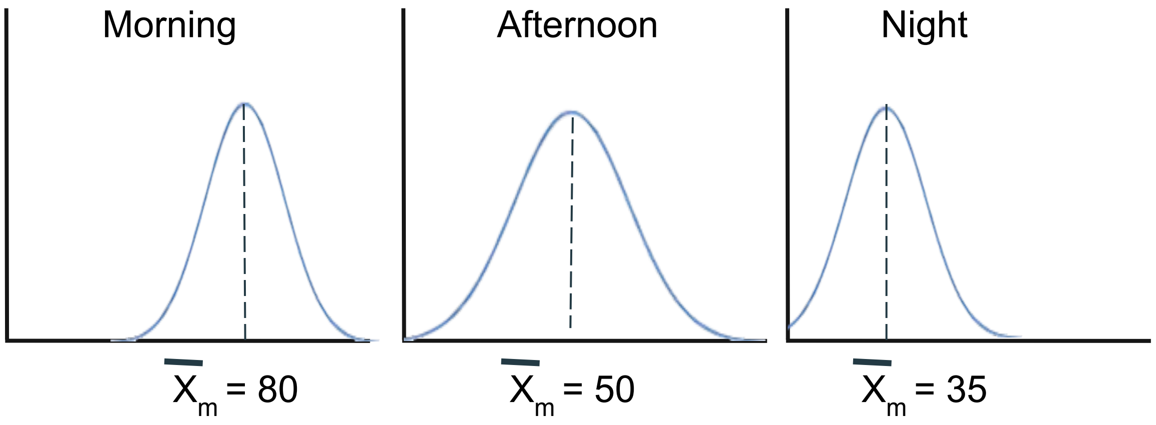
Differences between group means.

$$F = \frac{\text{Between Group Variance}}{\text{Within Group Variance}}$$

Differences among scores within each group.

What is F?

Example: Is there an effect of time of class (morning, afternoon, or night) on final exam scores (0-100)?



$n = 100$
$N = 300$ (100 in each group)

Morning Mean Score: 80
Afternoon Mean Score: 50
Night Mean Score: 35

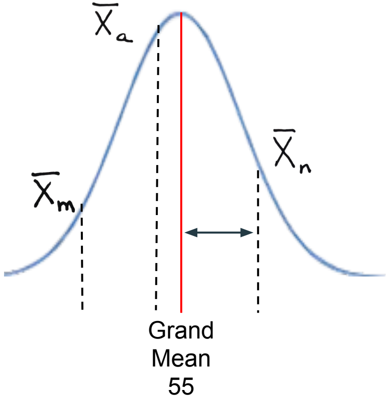
GRAND MEAN: 55

One-way ANOVA

What is F: Sums of Squares

Example: Is there an effect of time of class (morning, afternoon, or night) on final exam scores (0-100)?

Treatment
(Between Group Variance)

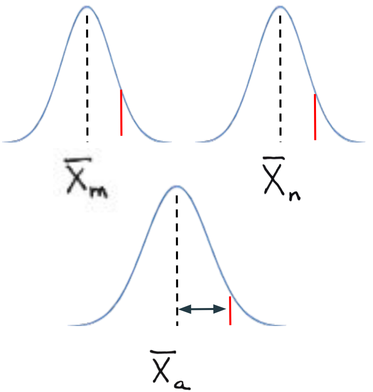


Sums of Squares Between

How much do the group (j) means vary around the grand (T) mean?

$$100\left[(35-55)^2 + (50-55)^2 + (80-55)^2\right] = 105{,}000$$

Error
(Within Group Variance)



Sums of Squares Error (Within)

How much does each observation (i) vary around its group (j) mean?

$$\sum(M-M)^2 + \sum(A-A)^2 + \sum(N-N)^2 = 50{,}850$$

What is F: Sums of Squares

Example: Is there an effect of time of class (morning, afternoon, or night) on final exam scores (0-100)?

Treatment
(Between Group Variance)

Sums of Squares Between
How much do the group (j) means vary around the grand (T) mean?

105,000

÷

Degrees of Freedom
(k-1)

2

=

Mean Squares Between
Estimate of the population variance based on how the group means vary.

52,500

Error
(Within Group Variance)

Sums of Squares Error (Within)
How much does each observation (i) vary around its group (j) mean?

50,850

÷

Degrees of Freedom
k(n-1) or N-k

297

=

Mean Squares Error (Within)
Estimate of the population variance based on the variances of each group

171.21

What is F?

What is the chance of getting this F ratio assuming the groups are not different?

Depends on how many people we're talking about (i.e. df)

$$F = \frac{\text{Between Group Variance}}{\text{Within Group Variance}} = \frac{52{,}500}{171.21} = 303.72$$

$$\boxed{F\,(2,\,297) = 303.72,\ p < .001}$$

The ratio of between group variance to within group variance

The chance you'd get this F value assuming the null is true (i.e. the chance you'd get those group means assuming there is no difference between the groups)

Calculate *F* Given This Output

ANOVA

	Sum of Squares	df	Mean Square	F	p	$\eta^2 p$
Style			105.66	?		
Residuals			8.02			

One-way ANOVA

Calculate *F* Given This Output

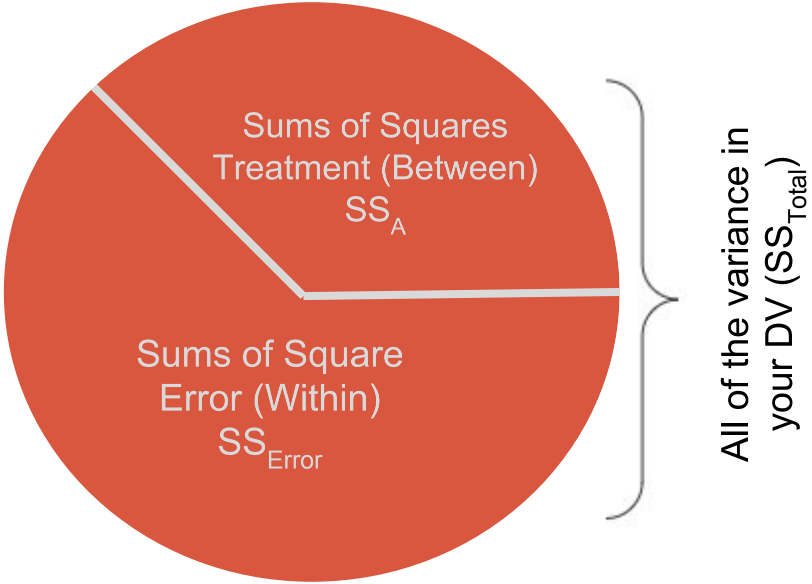
ANOVA

	Sum of Squares	df	Mean Square	F	p	η^2p
Style	211	2		?		
Residuals	457	57				

Source	SS	df	MS	F
treat	$n\ \Sigma(X_j - X_T)^2$	$k - 1$	SS_{treat}/df_{treat}	$MS_{treat}\ /MS_{error}$
error	$\Sigma(X_{ij} - X_j)^2$	$k(n-1)$	SS_{error}/df_{error}	-----
Total	$\Sigma(X_{ij} - X_T)^2$	$N - 1$	-----	-----

One-way ANOVA

What about Eta?



Eta (η^2)

What percentage of the variance in your DV is explained by the variance in your IV?

SS_A / SS_{Total}

Example: Is there an effect of time of class (morning, afternoon, or night) on final exam scores (0-100)?

η^2 = *What percentage of the variance in your exam scores can be explained by what time students took the test?*

$SS_A = 52{,}500$ $SS_{Error} = 171.21$

What is the effect of time of day on exam scores?

What is eta-squared?

The variance explained

ANOVA

	Sum of Squares	df	Mean Square	F	p	$\eta^2 p$
Style	211	2	105.66	13.2	< .001	?
Residuals	457	57	8.02			

Steps for Calculating

- CHECK ASSUMPTIONS
  - NORMALITY CUT OFFS _*Skew (-3 to 3) and Kurtosis (-10 to 10)*_
  - TEST FOR HOMOGENEITY OF VARIANCE: _*Levene's test*_
- TEST OF SIGNIFICANCE: _*F-value*_ & _*p-value*_
- EFFECT SIZE: _*Eta-squared*_
  - Small = _0.01_
  - Moderate = _*0.06*_
  - Large = _*0.15*_
- POST-HOCS: _*Tukey or Bonferroni*_

One-way ANOVA

T Test v. Tukey's v. Bonferonni

Multiple T Tests	Tukey's HSD	Bonferroni
Alpha (chance of Type I error) increased each time you run a T test (p + p + p)	Controls for the number of pairwise comparisons that will be done based on the number of groups in your study. Adjusts p value	Controls for all possible combinations of pairwise comparisons, regardless of group number. MOST CONSERVATIVE Might use when you can't afford a Type I error.

Example

- A new memory-enhancing drug has emerged and is being tested on a group of stats students! Students were given the new drug, a placebo, or no treatment prior to taking their final exam. You want to test whether the new memory drug increases test scores.

Example

ANOVA

	Sum of Squares	df	Mean Square	F	p	η^2p
Treatment	267	2	133.42	40.2	< .001	0.480
Residuals	289	87	3.32			

ASSUMPTION CHECKS

Test for Homogeneity of Variances (Levene's)

F	df1	df2	p
1.78	2	87	0.174

Example

Post Hoc Comparisons - Treatment

Treatment		Treatment	Mean Difference	SE	df	t	p-tukey
Drug	-	NoTr	-4.13	0.470	87.0	-8.78	< .001
	-	Placebo	-2.82	0.470	87.0	-5.99	< .001
NoTr	-	Placebo	1.31	0.470	87.0	2.78	0.018

Descriptives

Treatment	N	Mean	SD
Drug	30	32.1	1.99
NoTr	30	36.3	1.91
Placebo	30	35.0	1.53

Sample Write Up: Journal

A one way between groups ANOVA was used to determine the effect of a new memory-enhancing drug on statistics students final exam scores. There was a significant difference in exam scores based on drug condition, $F(2,87) = 40.20$, $p < .001$, $\eta^2 = .48$.. There was a large effect, with 48% of the variance in exam scores explained by a student's drug condition. Post-hoc pairwise comparisons using the Tukey's HSD procedure demonstrated a significant difference between students in the drug condition and those in the no treatment condition, $M1\text{-}M2 = -4.13$, $p < .001$. There was also a significant difference on exam scores between students in the drug condition and those in the placebo condition, $M1\text{-}M3 = -2.82$, $p < .001$. There was also a significant difference in exam scores between those in the no treatment group and those in the placebo group, $M2\text{-}M3 = 1.31$, $p = .018$. Overall, students in the no treatment condition did better on the exam ($M2 = 36.3$, $SD = 1.53$) compared to the drug condition group ($M1 = 32.10$, $SD = 1.99$) and the placebo group ($M3 = 35.00$, $SD = 1.91$).

Sample Write Up: Lay Audience

The new working memory drug made students do worse on the exam. Even students who thought they were taking the drug did worse than students who didn’t take anything, potentially because they were anxious about side effects. If you want your students to do better, don’t give them any kind of drugs - just tell them to study!

Factorial ANOVA

Compares the means across levels of two or more independent variables to determine whether they are significantly different.

General

When should you use a Factorial ANOVA?

- Two Independent Variables with at least two levels
- One Dependent Variable

Assumptions

- The DV is ___*continuous*___ & ___*normally distributed*___
- The IV is ___*categorical*___
- Variance is ___*homogenous*___
- Observations are ___*independent*___

Steps for Calculating

- TEST FOR HOMOGENEITY OF VARIANCE: _*Levene's test*_
  - CORRECTION: _*Welch*_
- TEST OF SIGNIFICANCE: _*F-value*_ & _*p-value*_
  - _2_ Main effects
  - _1_ Interaction effect
- EFFECT SIZE: _*Partial Eta-squared*_
- POST-HOCS FOR MAIN EFFECTS: _*Tukey or Bonferroni*_
- POST-HOCS FOR INTERACTION:
  *Simple effects analysis (ANOVA) & Tukey or Bonferroni if significant*

Factorial ANOVA

Main & Interaction Effects

- Main effect: *The effect of 1 IV averaged across the levels of the other IV*
- Interaction effect: *The effect of 1 IV depends on the other IV*
- Simple effect: *The effect of 1 IV at a particular level of the other IV*

Partitioning Sums of Squares (Variance) for our Dependent Variable

Whole circle = All variance in your DV

Systematic Variance:

SS_A = Main Effect of IV A

SS_B = Main Effect of IV B

SS_{AxB} = Interaction

Unsystematic Variance:

SS_{Error} = Error

Example

You are prepping a course lecture in statistics and developing an example to give your students. However, you let your kids decide what you want to be measured. One says they like ice cream... the other one is just staring at the sun and smiling. You decide to have a little fun and randomly assign some participants to either **experiencing a sunny or rainy day** (you control the weather…) and also either give them **strawberry, chocolate, or vanilla ice cream**. You are interested in whether there are **differences in joyfulness** based on these variables.

Main Analyses

ANOVA

ANOVA

	Sum of Squares	df	Mean Square	F	p	$\eta^2 p$
Icecream	90.150	2	45.075	37.590	< .001	0.397
Weather	0.300	1	0.300	0.250	0.618	0.002
Icecream:Weather	78.050	2	39.025	32.545	< .001	0.363
Residuals	136.700	114	1.199			

1. Is there a main effect of ice cream, averaged across the levels of weather, on joyfulness?
2. Is there a main effect of weather, averaged across the levels of ice cream, on joyfulness?
3. Is there an interaction of weather and ice cream on joyfulness?

Main Effects for Joyfulness

Is there a main effect of ice cream, averaged across the levels of weather, on joyfulness? **(Yes)**

	Strawberry Ice Cream	Chocolate Ice Cream	Vanilla Ice Cream
Average Weather			

Is there a main effect of weather, averaged across the levels of ice cream, on joyfulness? **(No)**

	Average Ice Cream
Sunny Weather	
Rainy Weather	

Is there an effect of 1 IV averaged across the levels of the other IV?

Main Effect Follow-Up: Post-Hoc

Because there is a **main effect of ice cream**, we should do post-hoc analyses for ice cream.

POST HOC TESTS

Post Hoc Comparisons - Icecream

Icecream		Icecream	Mean Difference	SE	df	t	p-bonferroni
Chocolate	-	strawberry	-0.0750	0.245	114	-0.306	1.000
	-	vanilla	1.8000	0.245	114	7.351	< .001
strawberry	-	vanilla	1.8750	0.245	114	7.657	< .001

Interaction for Joyfulness: 2 ways to look at this

The effect of 1 IV depends on the other IV **(we have one!)**

	Strawberry Ice Cream	Chocolate Ice Cream	Vanilla Ice Cream
Sunny Weather			
Rainy Weather			

	Strawberry Ice Cream	Chocolate Ice Cream	Vanilla Ice Cream
Sunny Weather			
Rainy Weather			

Interaction Follow-Up: Strawberry One-Way ANOVA

```
> anova(data = dat_A.1, dep = 'Joyfulness', factors = c('Weather'), effectSize = 'eta', postHoc = 'Weather', postHocCorr = 'bonf')
```

ANOVA

ANOVA

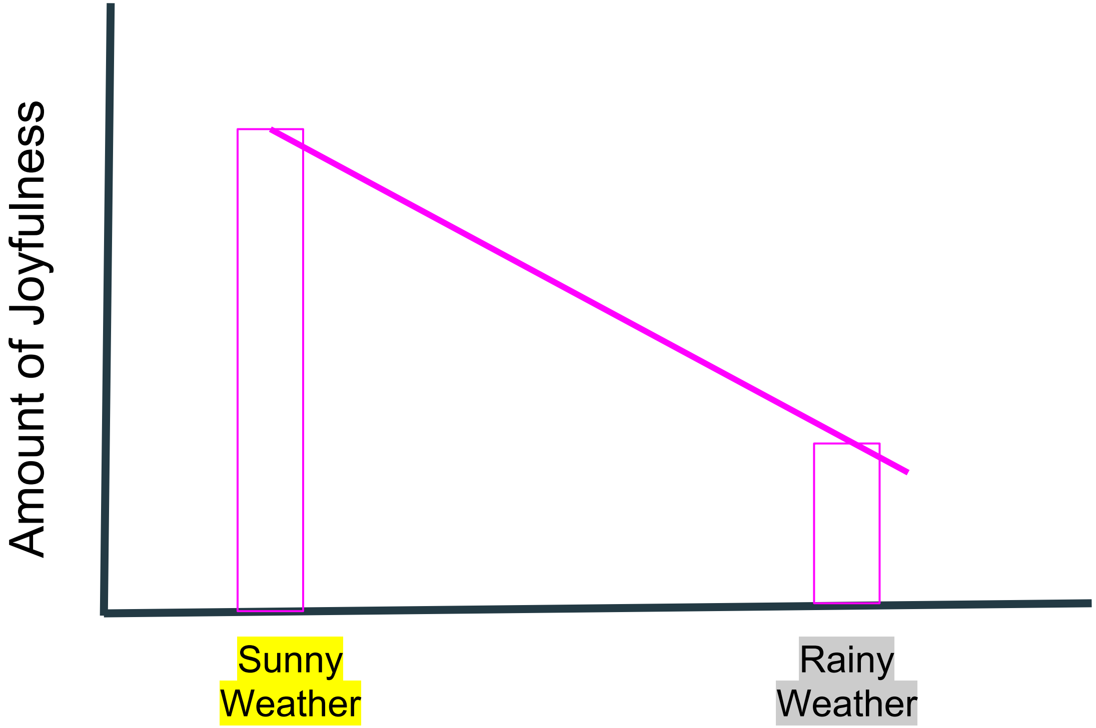
	Sum of Squares	df	Mean Square	F	p	η^2
Weather	36.1	1	36.10	33.5	< .001	0.468
Residuals	41.0	38	1.08			

POST HOC TESTS

Post Hoc Comparisons - Weather

Weather		Weather	Mean Difference	SE	df	t	p-bonferroni
cloudy	-	sunny	-1.90	0.328	38.0	-5.78	< .001

Interaction



Interaction Follow-Up: Chocolate One-Way ANOVA

```
> anova(data = dat_A.2, dep = 'Joyfulness', factors = c('Weather'), effectSize = 'eta', postHoc = 'Weather', postHocCorr = 'bonf')
```

ANOVA

ANOVA

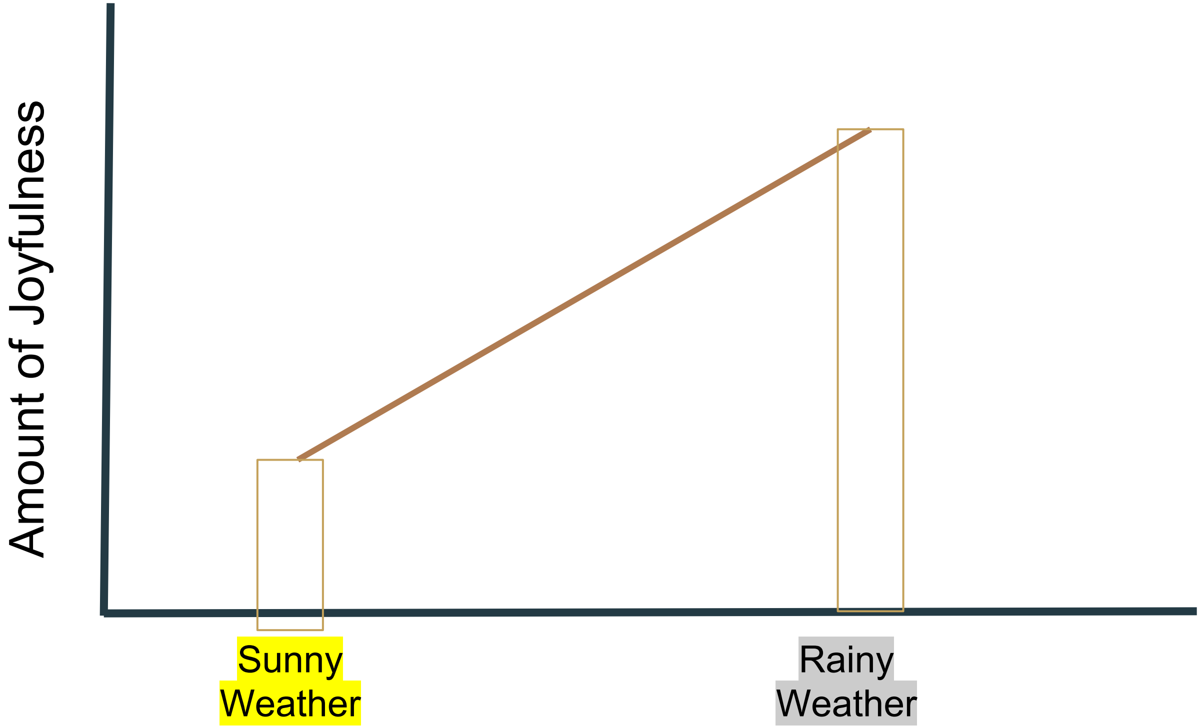
	Sum of Squares	df	Mean Square	F	p	η²
Weather	42.0	1	42.03	30.2	< .001	0.442
Residuals	53.0	38	1.39			

POST HOC TESTS

Post Hoc Comparisons - Weather

Weather		Weather	Mean Difference	SE	df	t	p-bonferroni
cloudy	-	sunny	2.05	0.373	38.0	5.49	< .001

Interaction



Interaction Follow-Up: Vanilla One-Way ANOVA

```
> anova(data = dat_A.3, dep = 'Joyfulness', factors = c('Weather'), effectSize = 'eta', postHoc = 'Weather', postHocCorr = 'bonf')
```

ANOVA

ANOVA

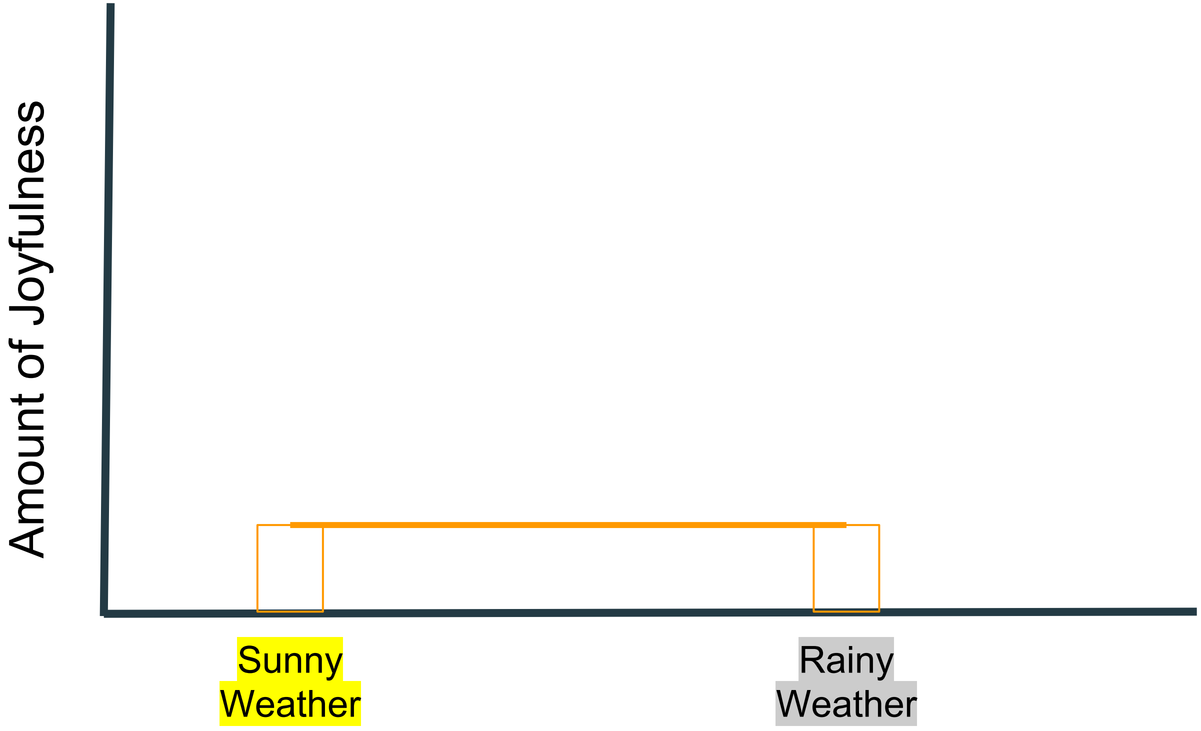
	Sum of Squares	df	Mean Square	F	p	η^2
Weather	0.225	1	0.225	0.200	0.657	0.005
Residuals	42.750	38	1.125			

POST HOC TESTS

Post Hoc Comparisons - Weather

Weather		Weather	Mean Difference	SE	df	t	p-bonferroni
cloudy	-	sunny	0.150	0.335	38.0	0.447	0.657

Interaction

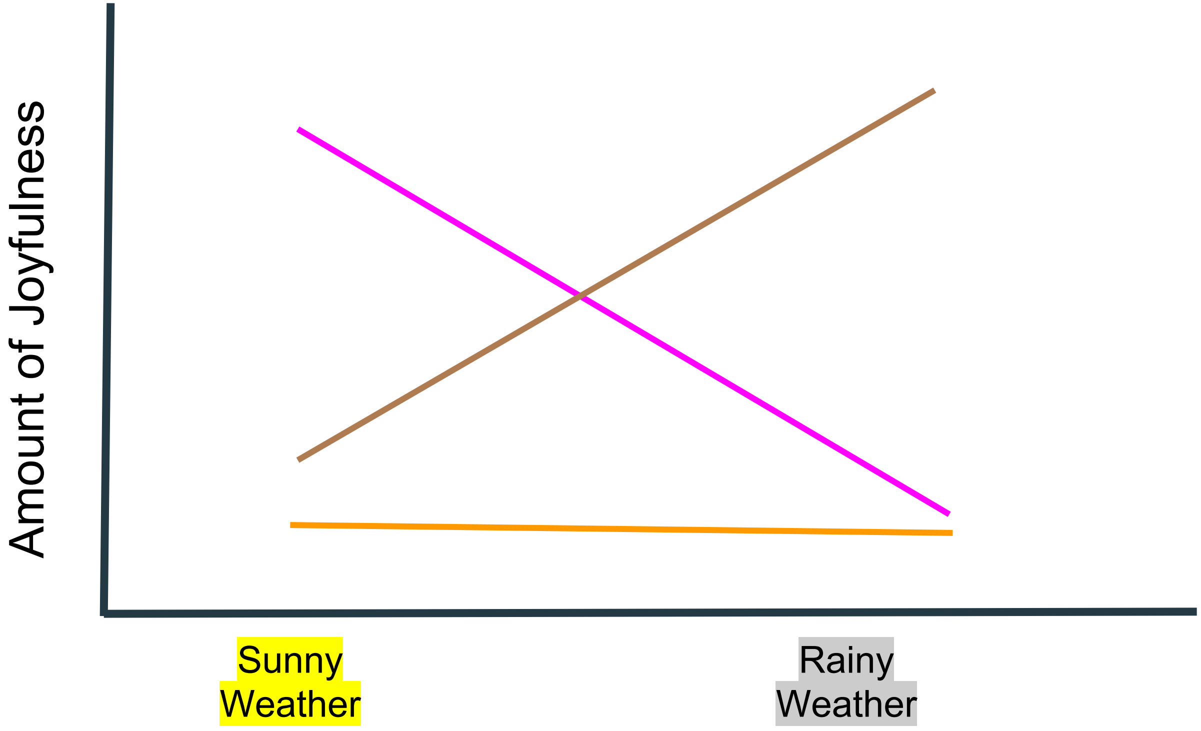


Joyfulness Simple Effects

	Strawberry Ice Cream	Chocolate Ice Cream	Vanilla Ice Cream
Sunny Weather			
Rainy Weather			

The effect of 1 IV at a particular level of the other IV

Interaction



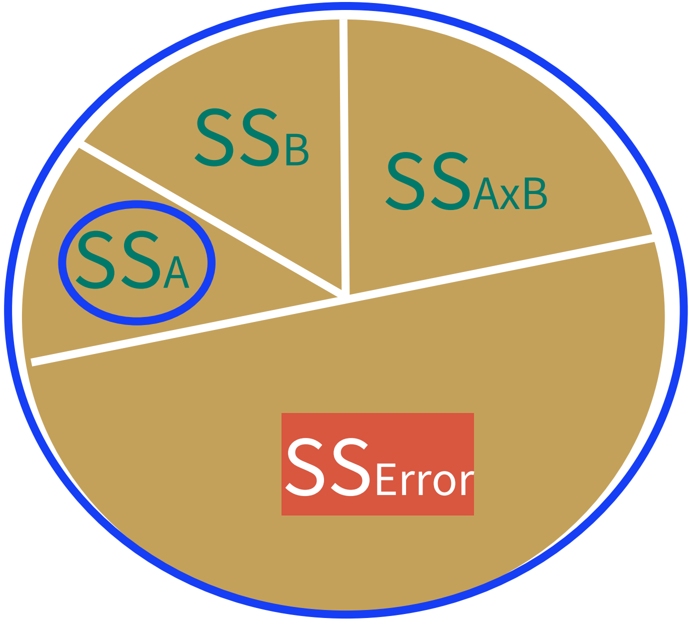
Sample Write-Up

This study examined the effects of ice cream (strawberry, chocolate, vanilla) and weather (rainy, sunny) on joyfulness using a 2 x 3 Factorial ANOVA. **There was a significant main effect of ice cream** averaged across the weather, $F(2, 114) = 45.08$, $p < .001$, $\eta^2 p = .40$. This is a large effect size, indicating that 40% of the variance is accounted for by ice cream. In this main effect, chocolate ice cream had greater levels of joyfulness than vanilla ice cream, $M1 - M3 = 1.80$, $p < .001$. Additionally, strawberry ice cream had greater levels of joyfulness than vanilla, $M2 - M3 = 1.88$, $p < .001$. **There was no significant main effect of weather.** There was also an **significant interaction of ice cream and weather on joyfulness**, $F(2, 114) = 39.03$, $p < .001$, $\eta^2 p = .36$. This is a large effect size, indicating that 36% of the variance is accounted for by the interaction. Using bonferroni's post hoc test, there were two significant findings. First, strawberry ice cream had significantly higher levels of joyfulness in sunny weather compared to cloudy weather, $M1 - M2 = -1.90$, $p < .001$. Conversely, chocolate ice cream resulted in significantly higher levels of joyfulness in cloudy weather compared to sunny weather, $M1 - M2 = 2.05$, $p < .001$. There was no difference with vanilla ice cream in sunny or rainy weather.

Complete Eta-Squared vs. Partial Eta-Squared

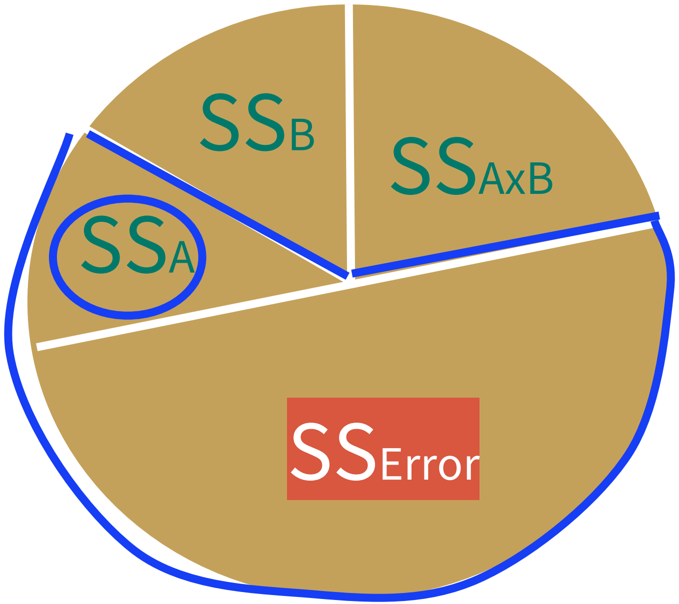
SSeffect / SSTotal

SSA / SSTotal



SSeffect / (SSeffect + SSerror)

SSA / (SSA + SSerror)



PRACTICE TIME

Practice: Which Test Do I Use?

- You want to test the stability of eyewitness memory. To do this, you conduct an experiment in which you show participants a 30-second video clip, then test their memory of specific events in the video. You randomly assign participants to three conditions. In the first, participants immediately respond to a series of questions about the video. In the second, participants wait 10 minutes before responding. In the third condition, participants wait 30 minutes before responding.

Practice: Which Test Do I Use?

- You want to know if holding a review session before a statistics exam improves test scores. You <u>do not</u> hold a review session before administering the midterm. You <u>do</u> hold a review session before administering the final exam.

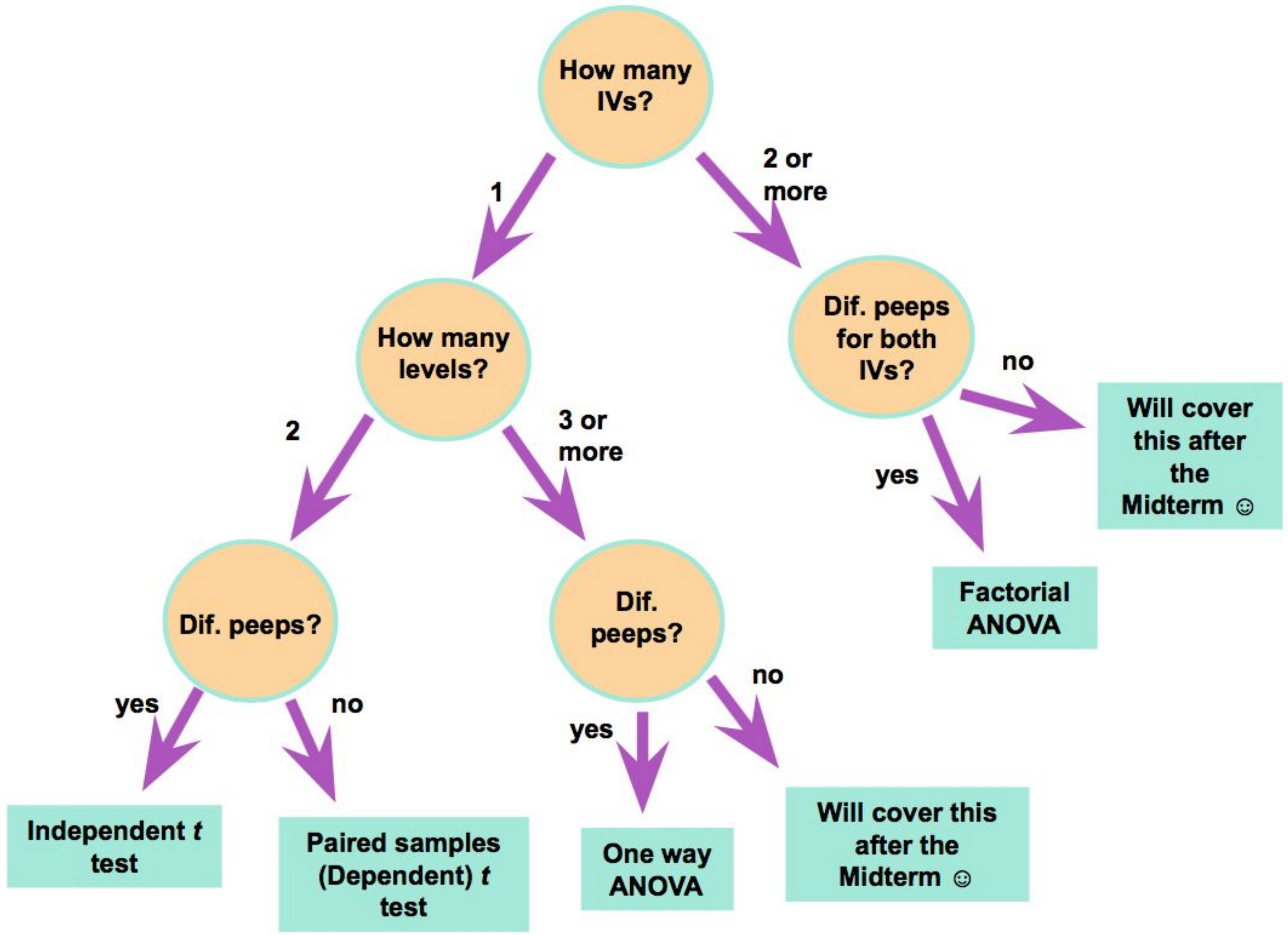
Practice: Which Test Do I Use?

- You want to know if job difficulty and flextime predict job satisfaction. You measure job satisfaction as low, medium and high. Flextime offered is measured as yes or no.

Practice: Which Test Do I Use?

- You want to know if teaching infants sign language helps or hinders the language acquisition process. To test this, you have 50 infants learn sign language in conjunction with their native language. You have another 50 infants learn their native language only.

But, wait? I want to consider all these things together. There are so many examples. This is so confusing. There are multiple ANOVAs? When I do do a t-test? **WHAT?!** What do I do?



How many IVs?
1
2 or more
How many levels?
2
3 or more
Dif. peeps for both IVs?
no
yes
Will cover this after the Midterm ☺
Factorial ANOVA
Dif. peeps?
yes
no
Independent t test
Paired samples (Dependent) t test
Dif. peeps?
yes
no
One way ANOVA
Will cover this after the Midterm ☺

Good luck!