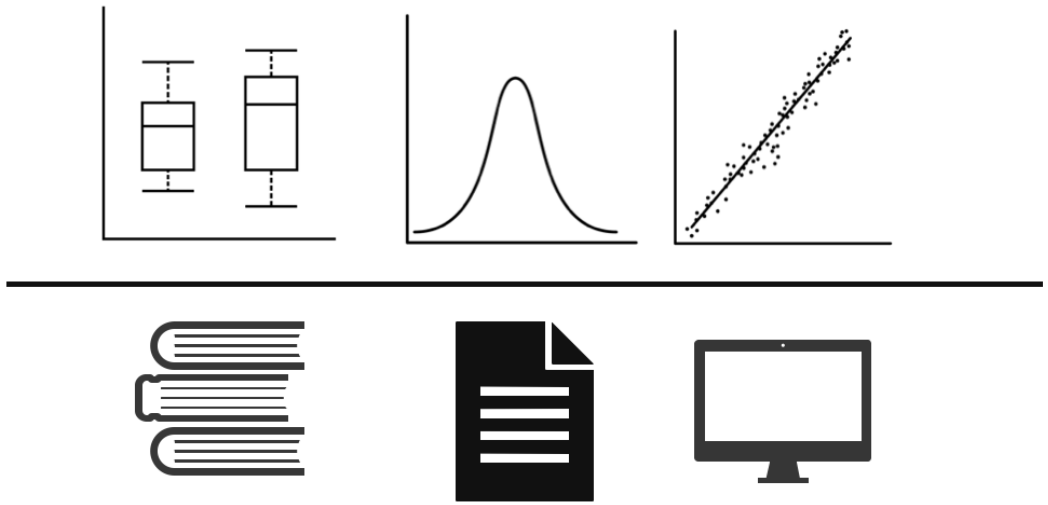
PSYC 2300

Introduction to Statistics



Lecture 16: Testing Relationships Using Correlations

Outline for today

- Review last parts of class
- Correlations Revisited
- Correlations in JASP
    - Download `Stats class 16 Dataset (Correlation T-Test).jasp`
- Continue one-way ANOVA practice activity

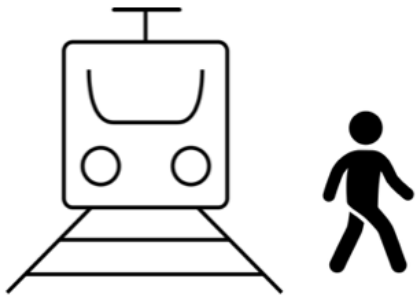
Review

One-way ANOVA

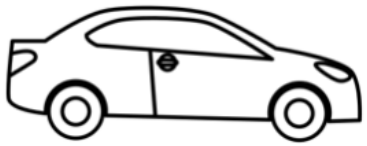
Goal

To be able to compare whether differences exist between multiple factors, as well as in interactions between the levels of different factors

Research question: What is the fastest way for me to get to campus?

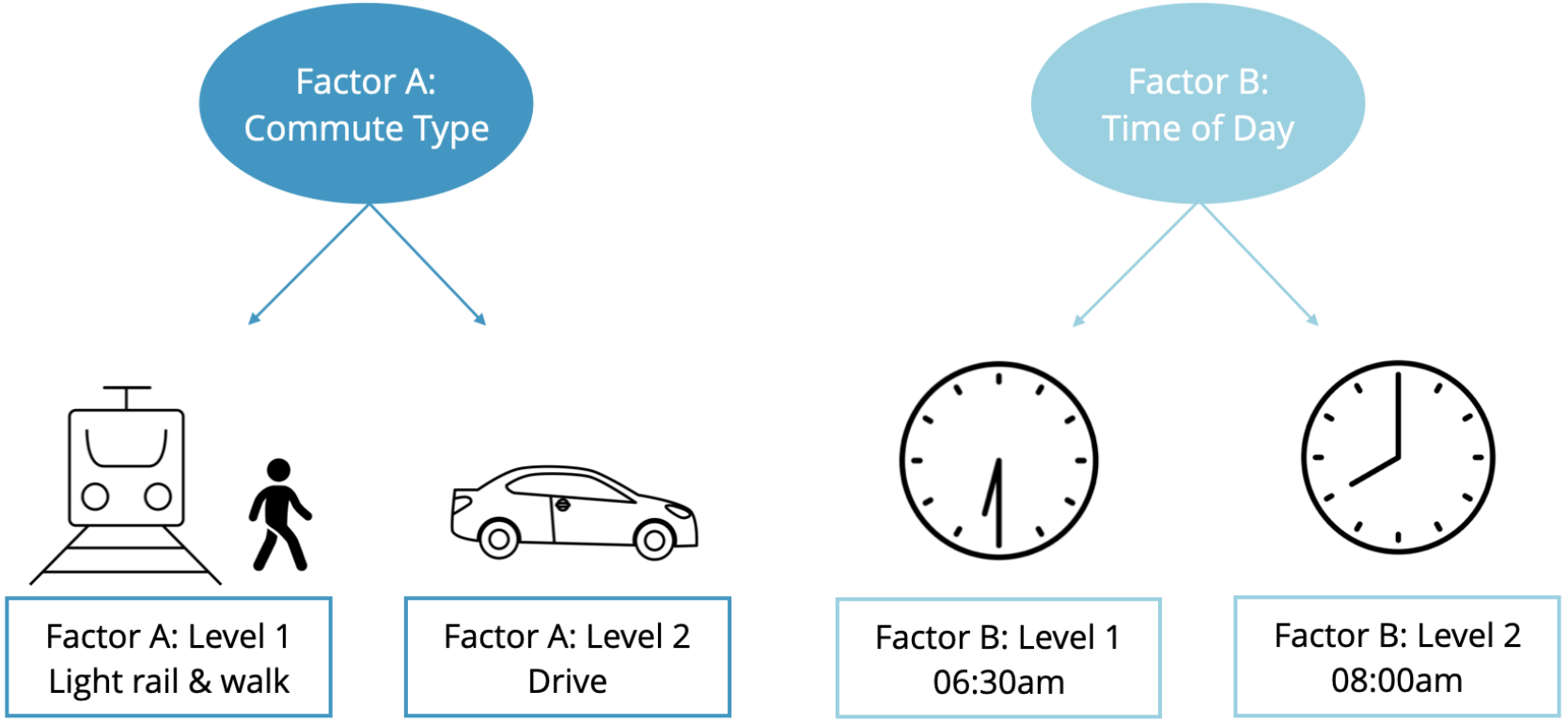


vs.



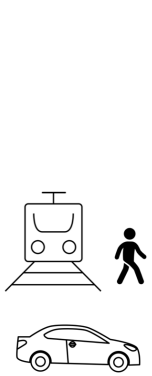
vs.

ANOVA: factors and levels



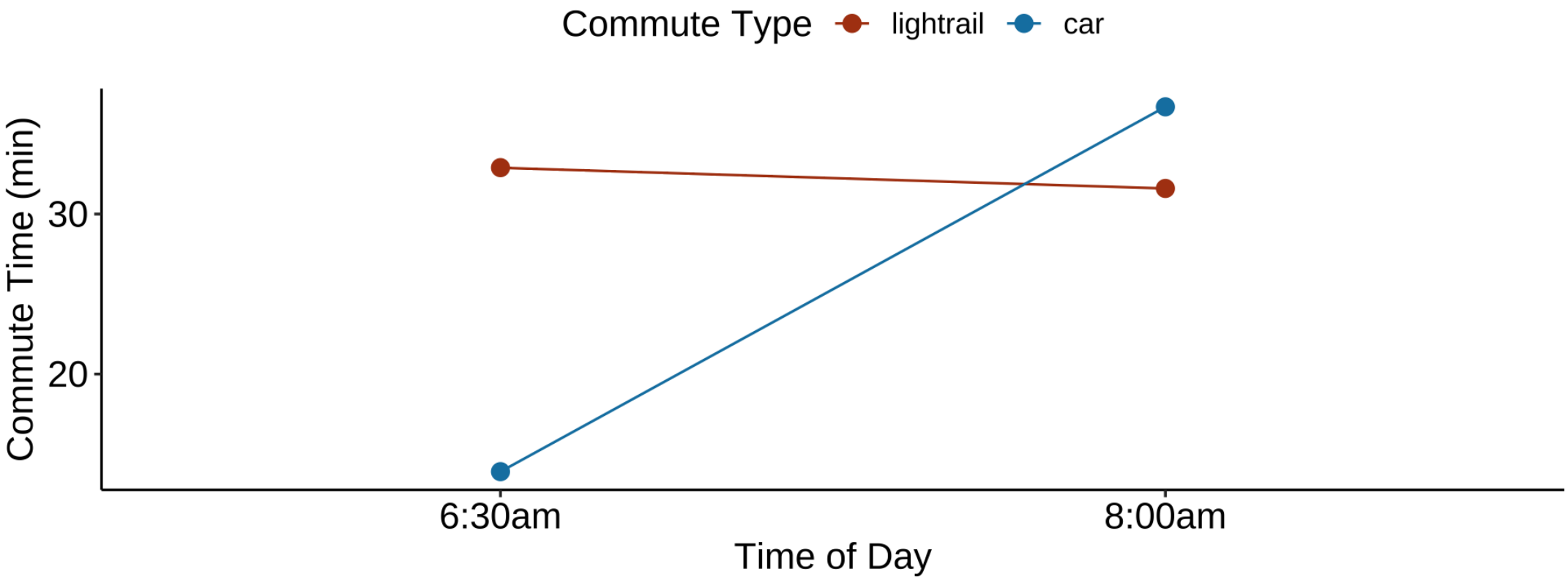
Factorial Designs

Factorial designs: Designs in which two or more factors are completely crossed (i.e., measurements are taken for every combination of factor levels)



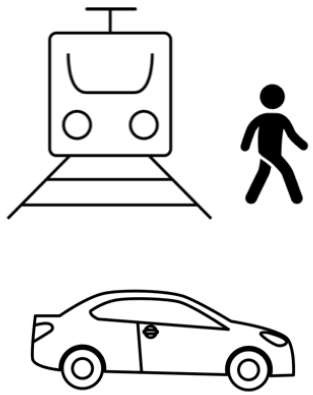
	6:30am	08:00am
Light rail & walk	32.9 min	31.6 min
Drive	13.9 min	36.7 min

Factorial designs: Plot



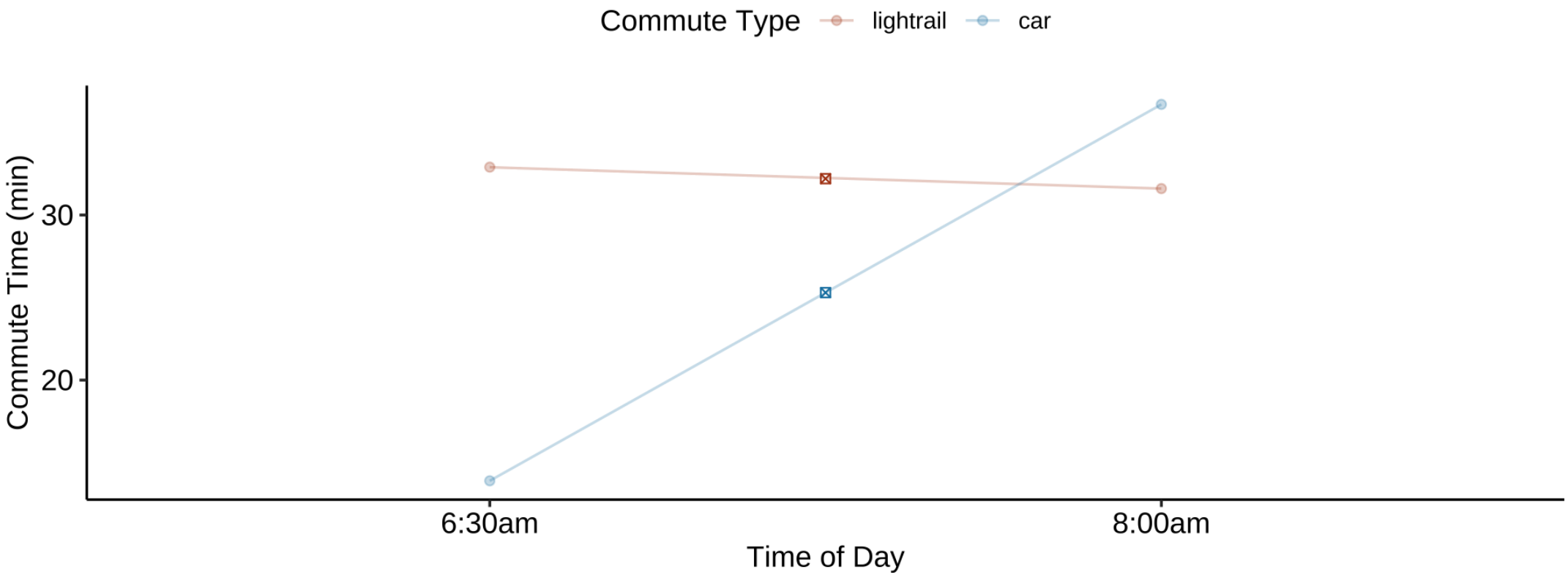
Main effect of Commute Type

Possible test #1: Effect of **Commute Type**



Light rail & walk	Average time for light rail and walking
Drive	Average time for driving

Commute Type: Plot



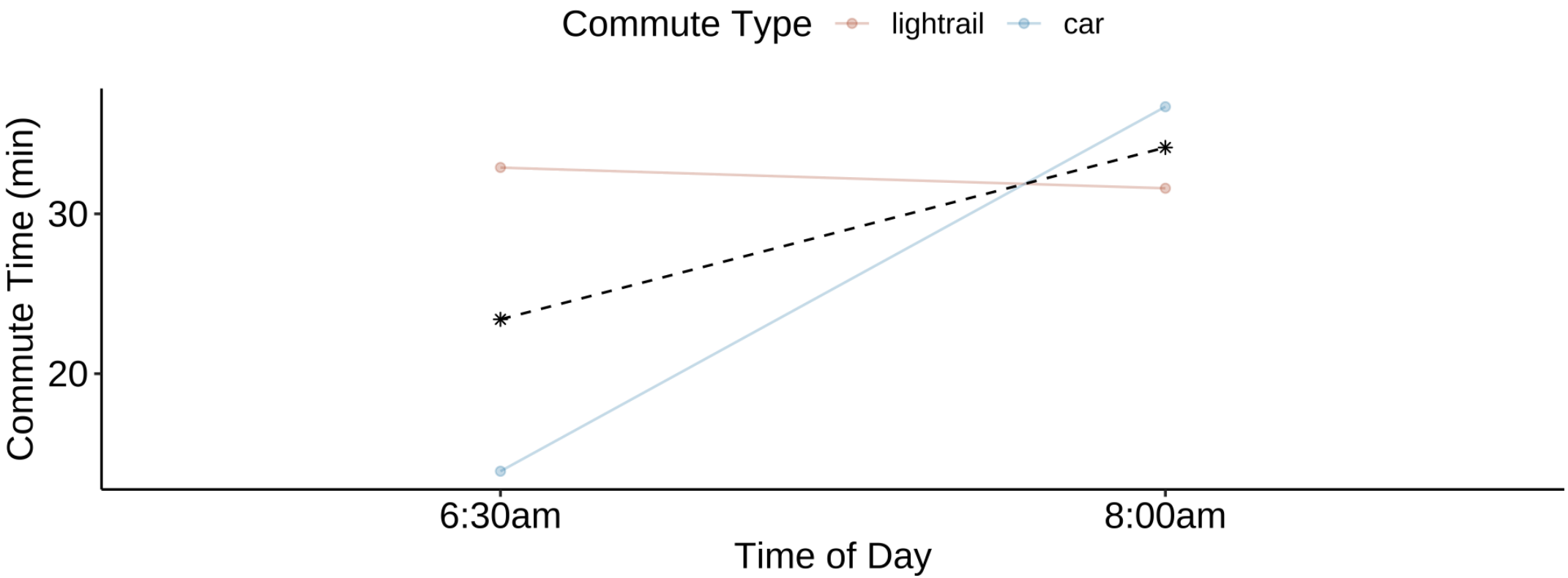
Taking the lightrail and walking takes longer than driving

Main effect of Time of Day

Possible test #2: Effect of **Time of Day**

6:30am	**08:00am**
Average commute at 6:30am	Average commute at 8:00am

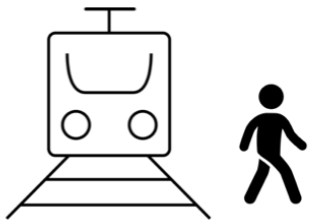
Time of Day: Plot

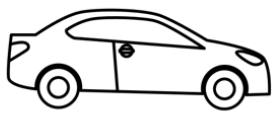


It takes longer to get to campus when leaving at 8:00am than at 6:30am

Interaction between Commute Type and Time

Possible test #3: Effect of specific **Commute Time** and **Time of Day** combinations

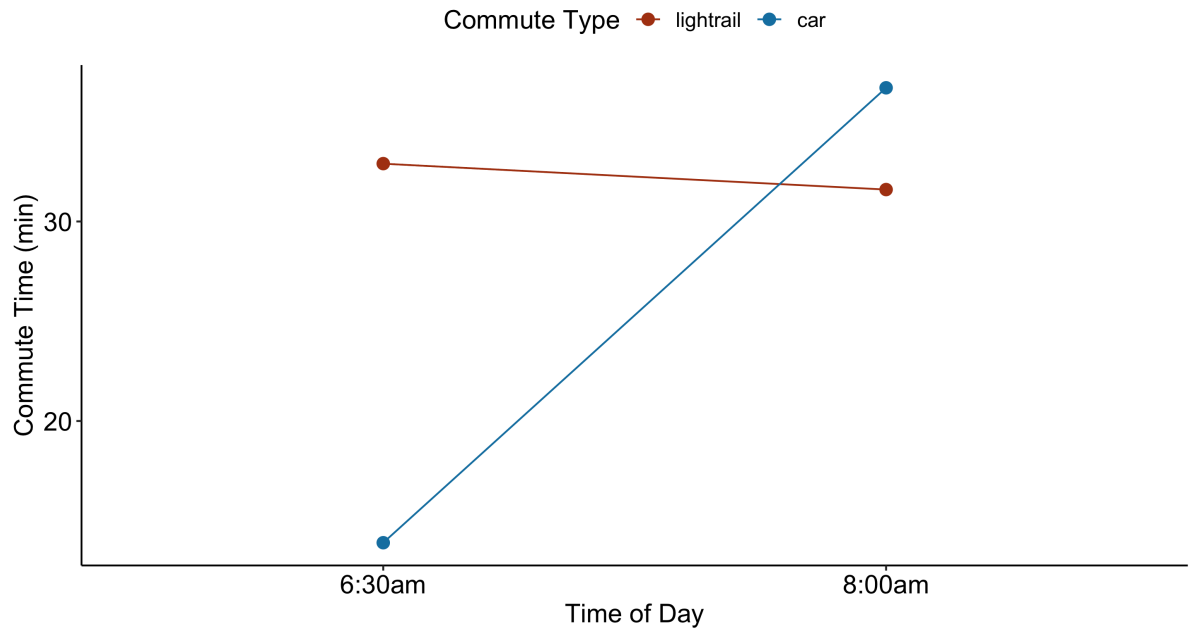




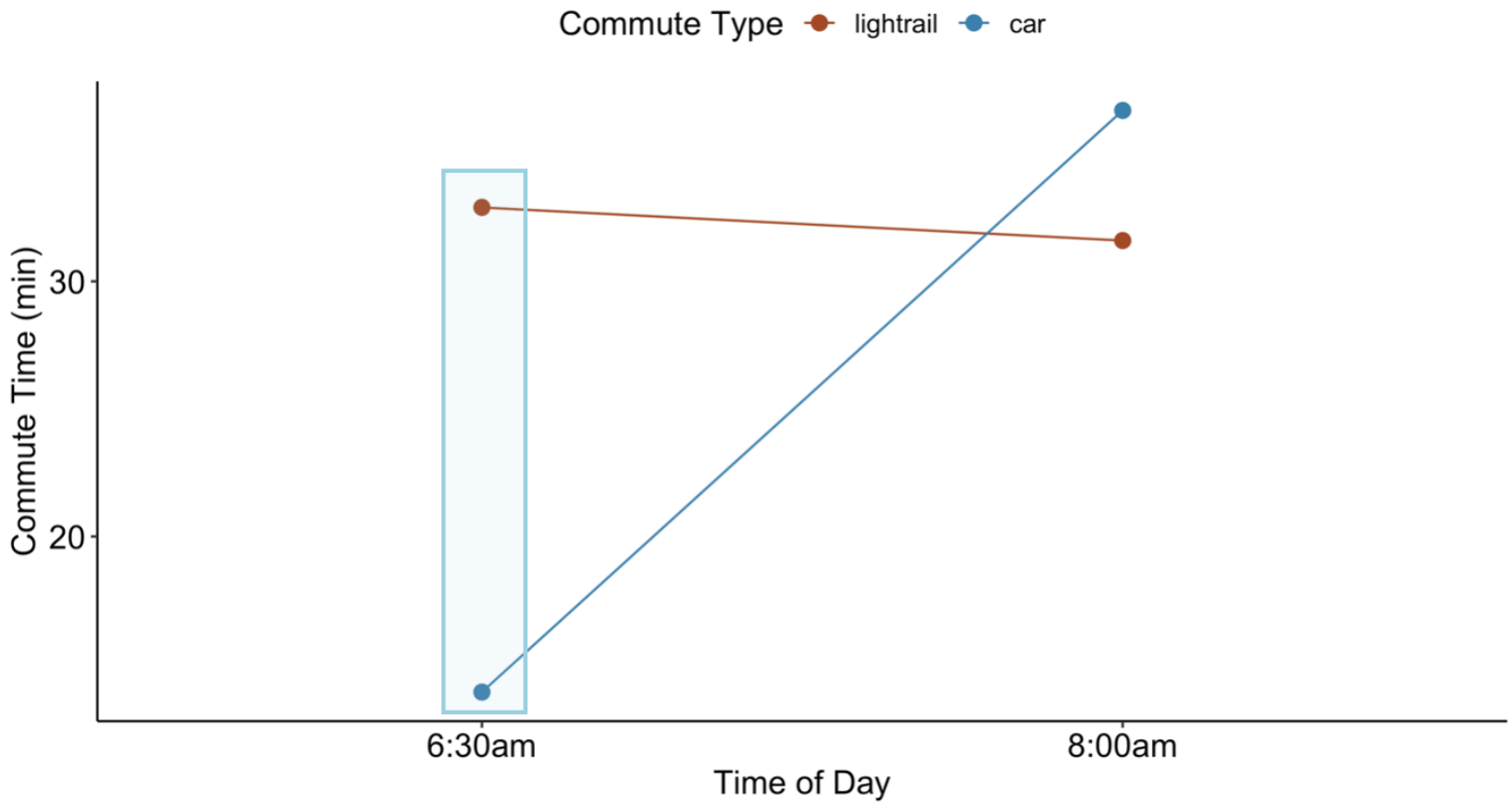
	6:30am	**08:00am**
Light rail & walk	Average time for light rail and walking at 6:30am	Average time for light rail and walking at 8:00am
Drive	Average time for driving at 6:30am	Average time for driving at 8:00am

Interaction

Interaction: describes the degree to which the effect of *one factor* depends on the *level* of the other factor

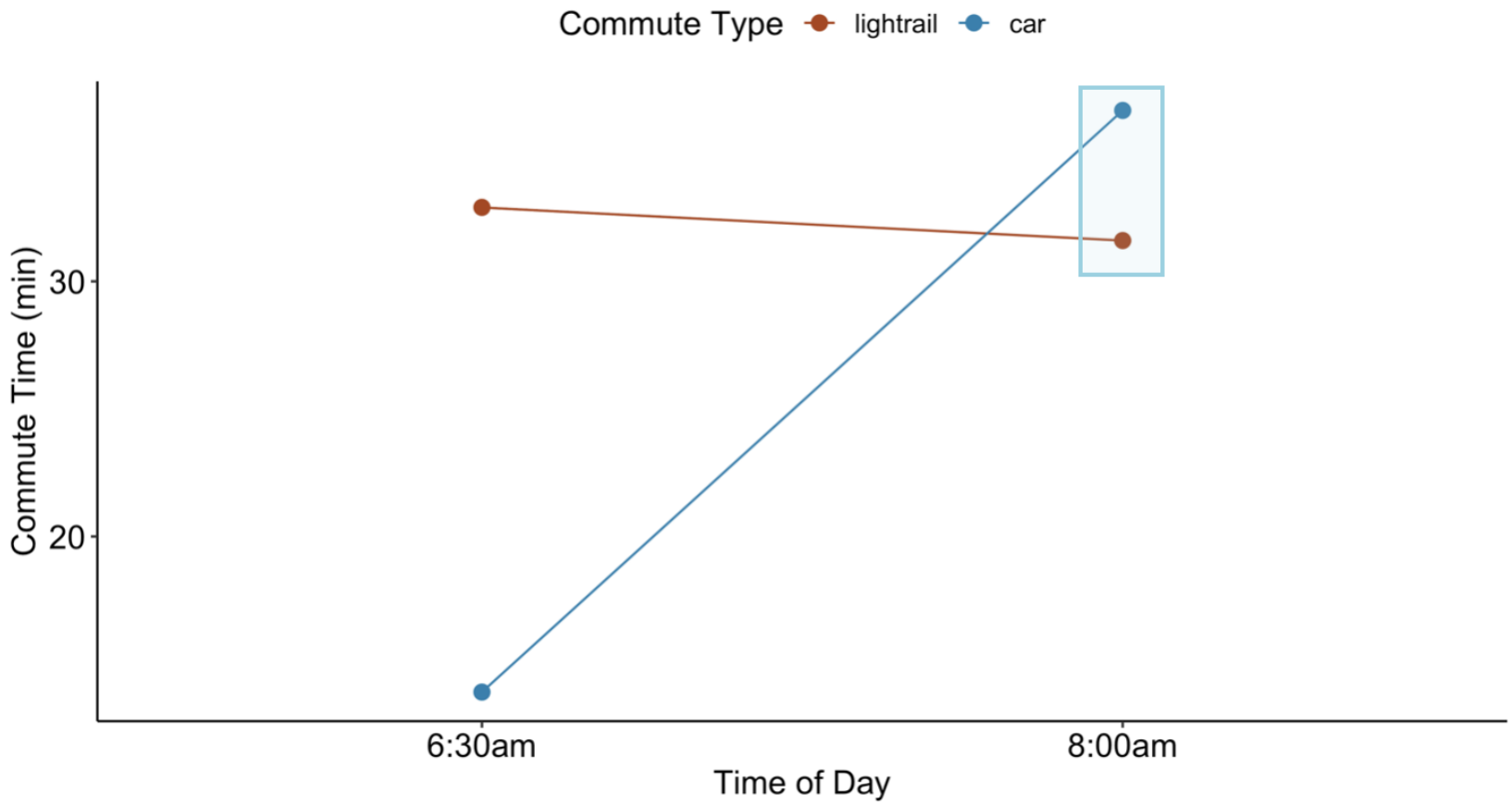


Interaction



At **6:30am**, **driving** is much *faster* than taking the **light rail and walking**

Interaction



But at **8:00am**, **light rail and walking** is much *faster* than **driving**

Correlations Revisited

Correlations

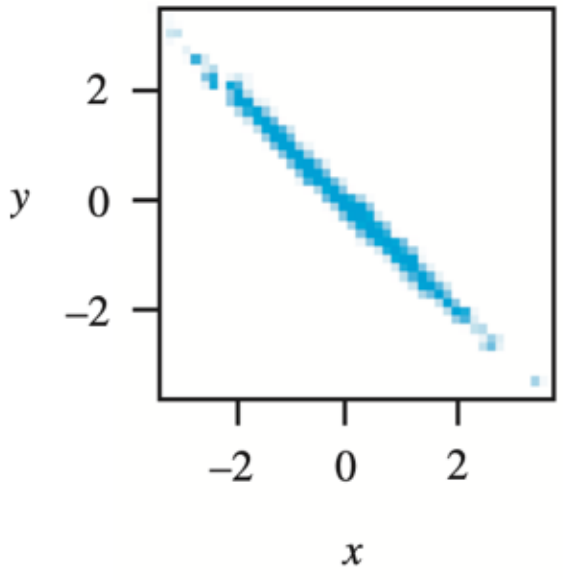
Correlational designs: examine the extent to which two variables are associated

Examples

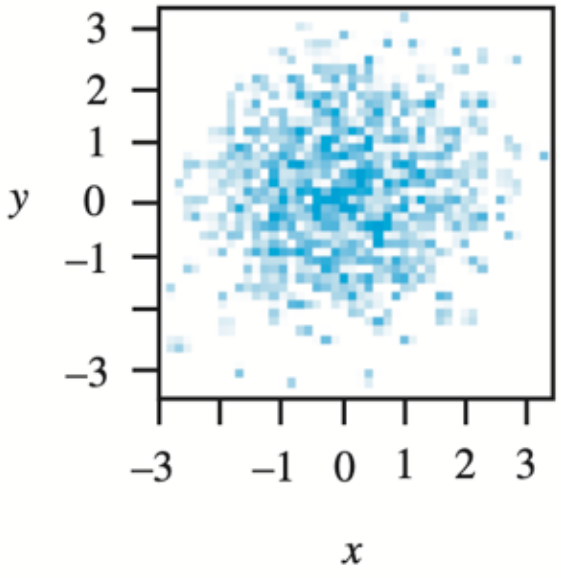
- Do people who are more creative do better in school?
- Is there a relationship between hours studied and exam scores? Class attendance and final grades?
- Are people who have higher marital satisfaction better parents?

Correlations

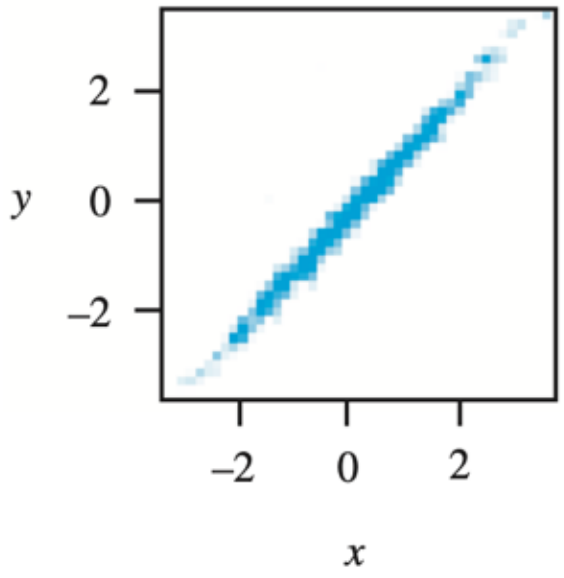
Correlations are either *positive*, *negative*, or *zero*



Negative



Zero



Positive

Correlations

Positive correlation: as one variable changes, the other variable changes in the *same direction*

Example

Relationship between
hours studied and exam grade

(more hours studied, higher exam grade)

Negative correlation: as one variable changes, the other variable changes in the *opposite direction*

Example

Relationship between
running speed and time to finish a race

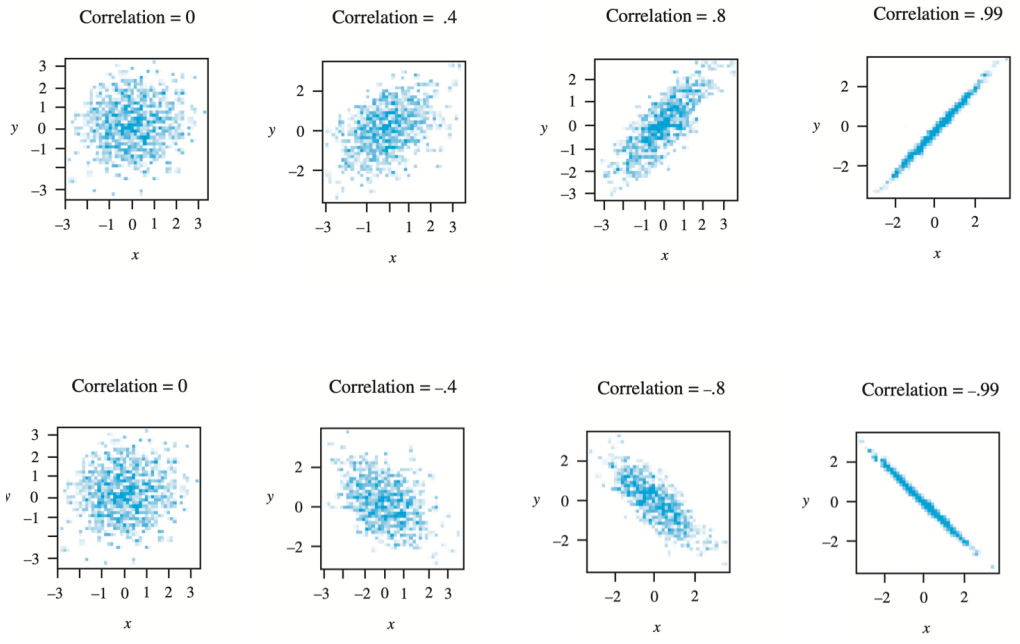
(faster running, lower time to finish race)

Zero correlation: there is *no* relationship between the two variables

Correlations

Correlations are *always* between -1.0 and +1.0

- This tells you the **magnitude** of relationship
  - Correlations closer to absolute value of 1 are *stronger*
  - Correlations closer to a value of 0 are *weaker*



Correlations: Practice

Which of the following correlation coefficients depicts the *strongest* relationship?

1. 0.43
2. -0.02
3. 0.67
4. -0.73

Answer: **-0.73**

Correlations: Practice

Which of the following correlation coefficients depicts the *weakest* relationship?

1. -0.13
2. 0.33
3. -0.03
4. 0.05

Answer: **0.05**

Correlations: Practice

Interpret the following statement in your own words:

> "We found a significant correlation of 0.37 between social media use and anxiety."

Correlations: Practice

Interpret the following statement in your own words:

> "We found a significant correlation of -0.26 between depressive symptoms and exercise."

Calculating Correlations

Pearson's *r*

$$r_{xy} = \frac{n\sum XY - \sum X \sum Y}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$$

We just need six values:

$\sum X$
$\sum Y$
$\sum XY$
$\sum X^2$
$\sum Y^2$
n

Calculating Correlations

Set up our table with the x and y values

i	x	y	x^2	y^2	xy
1	5	6			
2	9	11			
3	10	6			
4	3	4			
5	5	6			
6	7	9			

Calculating Correlations

Square our x values

i	x	y	x^2	y^2	xy
1	5	6	25		
2	9	11	81		
3	10	6	100		
4	3	4	9		
5	5	6	25		
6	7	9	49		

Calculating Correlations

Square our y values

i	x	y	x^2	y^2	xy
1	5	6	25	36	
2	9	11	81	121	
3	10	6	100	36	
4	3	4	9	16	
5	5	6	25	36	
6	7	9	49	81	

Correlation: Calculation

Multiply x and y together row-wise

i	x	y	x^2	y^2	xy
1	5	6	25	36	30
2	9	11	81	121	99
3	10	6	100	36	60
4	3	4	9	16	12
5	5	6	25	36	30
6	7	9	49	81	63

Correlation: Calculation

i	x	y	x^2	y^2	xy
1	5	6	25	36	30
2	9	11	81	121	99
3	10	6	100	36	60
4	3	4	9	16	12
5	5	6	25	36	30
6	7	9	49	81	63

Sum each of the columns

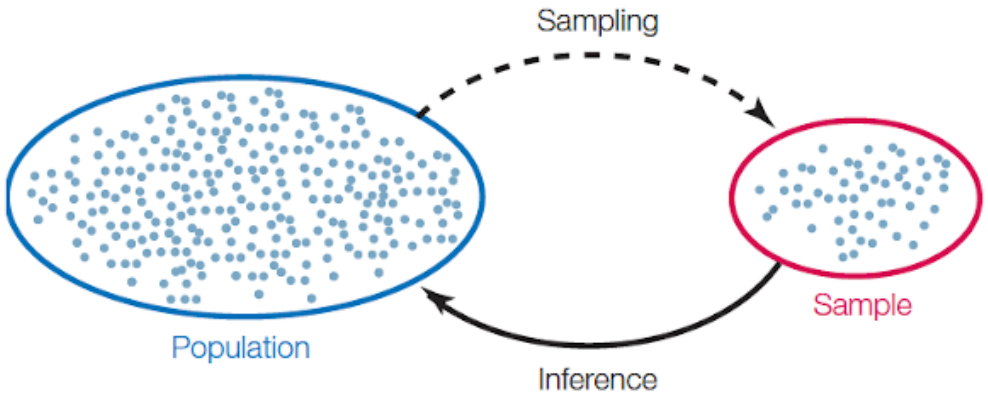
$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$
39	42	289	326	294

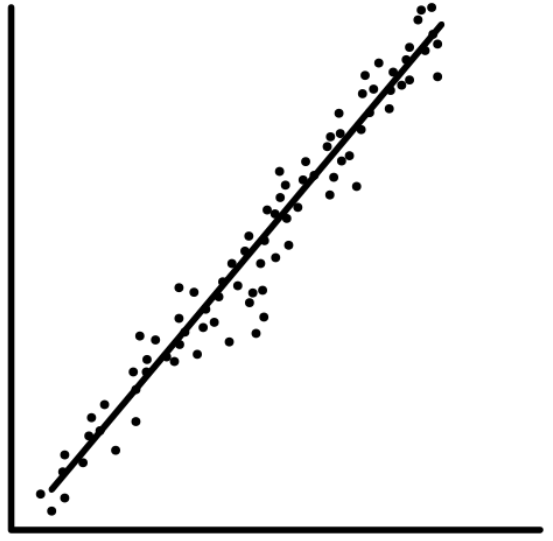
$$r_{xy} = \frac{126}{\sqrt{(213)(192)}} = 0.62306 = 0.62$$

The Problem

So far, we've only used correlations as descriptive statistics

What if we want to make inferences about relationships in the population?





Goal

To be able to determine whether an observed correlation coefficient is significantly different from 0 or not

Null Hypothesis H_0

There is no relationship

$$H_0: \rho = 0$$

Alternative Hypothesis H_a

There is a relationship

$$H_a: \rho \neq 0$$

Note: ρ is "rho", the population parameter version of the sample statistic, r

t-test for correlation coefficient

t-test for correlation coefficient

$$t_{r_{xy}} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}}$$

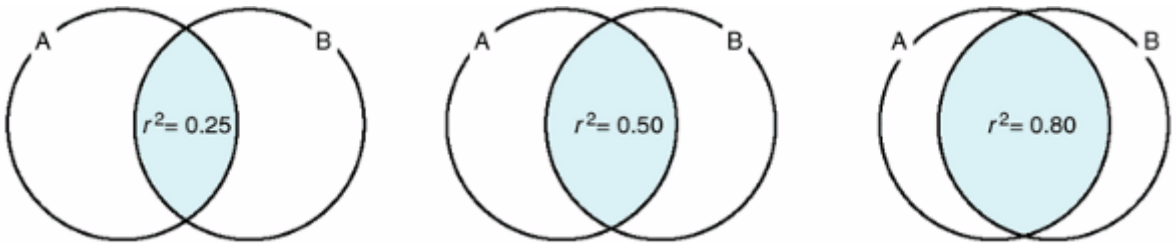
$r_{xy} =$ the correlation between the two variables

$r_{xy}^2 =$ the coefficient of determination

$n =$ the sample size (number of paired observations)

Coefficient of Determination

Coefficient of Determination: Percentage of variances in one variable that is accounted for by the variance in the other variable



Calculating Correlations

i	x	y	x^2	y^2	xy
1	5	6	25	36	30
2	9	11	81	121	99
3	10	6	100	36	60
4	3	4	9	16	12
5	5	6	25	36	30
6	7	9	49	81	63

Sum each of the columns

$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$
39	42	289	326	294

$$r_{xy} = \frac{126}{\sqrt{(213)(192)}} = 0.62306 = 0.62$$

Calculating coefficient of determination

$$r_{xy} = \frac{126}{\sqrt{(213)(192)}} = 0.62306 = 0.62$$

$$r_{xy}^2 = 0.62^2 = 0.38$$

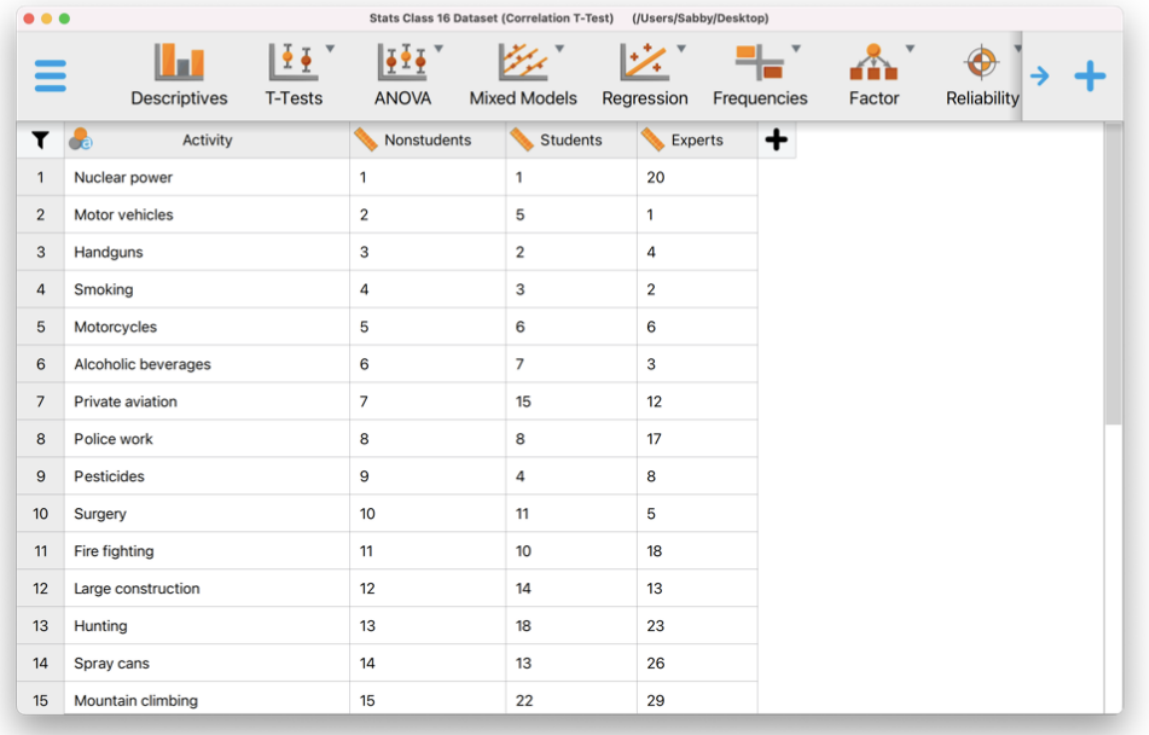
Interpretation

- 38% of the variance in x is accounted for by the variance in y
- 62% of the variance is not accounted for

Correlations in JASP

Correlations in JASP

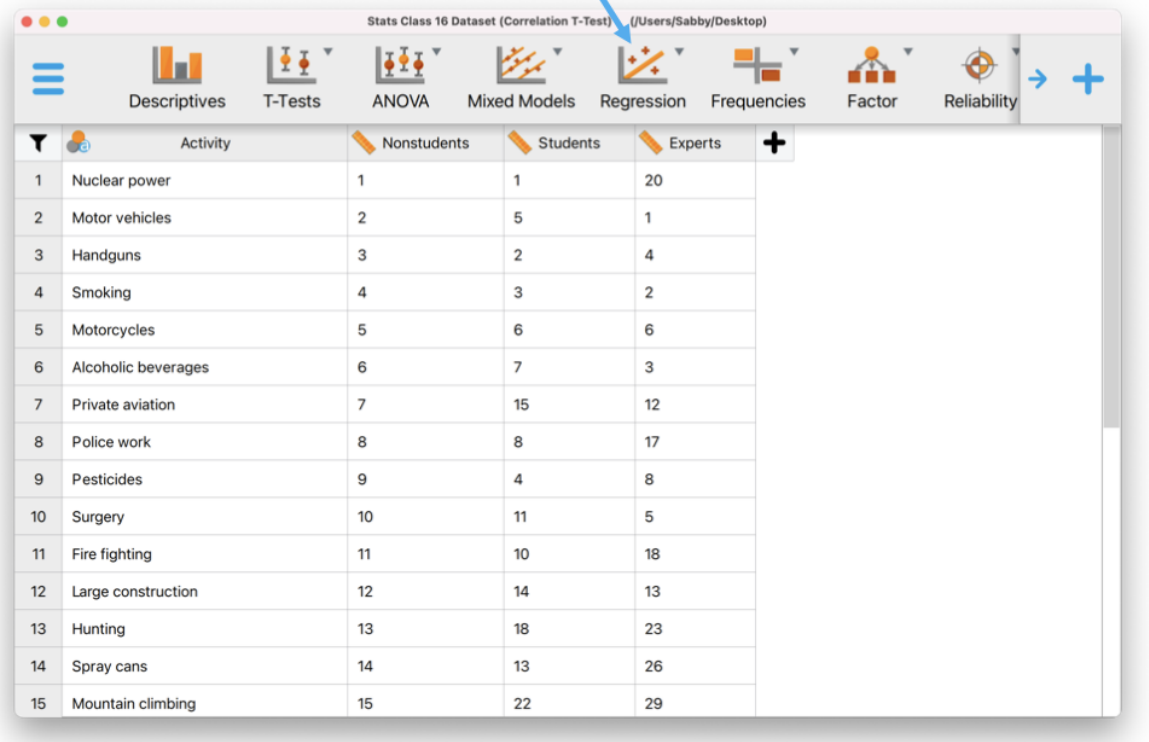
This dataset shows average ratings of the danger of various activities from nonstudents, students, and experts.



	Activity	Nonstudents	Students	Experts
1	Nuclear power	1	1	20
2	Motor vehicles	2	5	1
3	Handguns	3	2	4
4	Smoking	4	3	2
5	Motorcycles	5	6	6
6	Alcoholic beverages	6	7	3
7	Private aviation	7	15	12
8	Police work	8	8	17
9	Pesticides	9	4	8
10	Surgery	10	11	5
11	Fire fighting	11	10	18
12	Large construction	12	14	13
13	Hunting	13	18	23
14	Spray cans	14	13	26
15	Mountain climbing	15	22	29

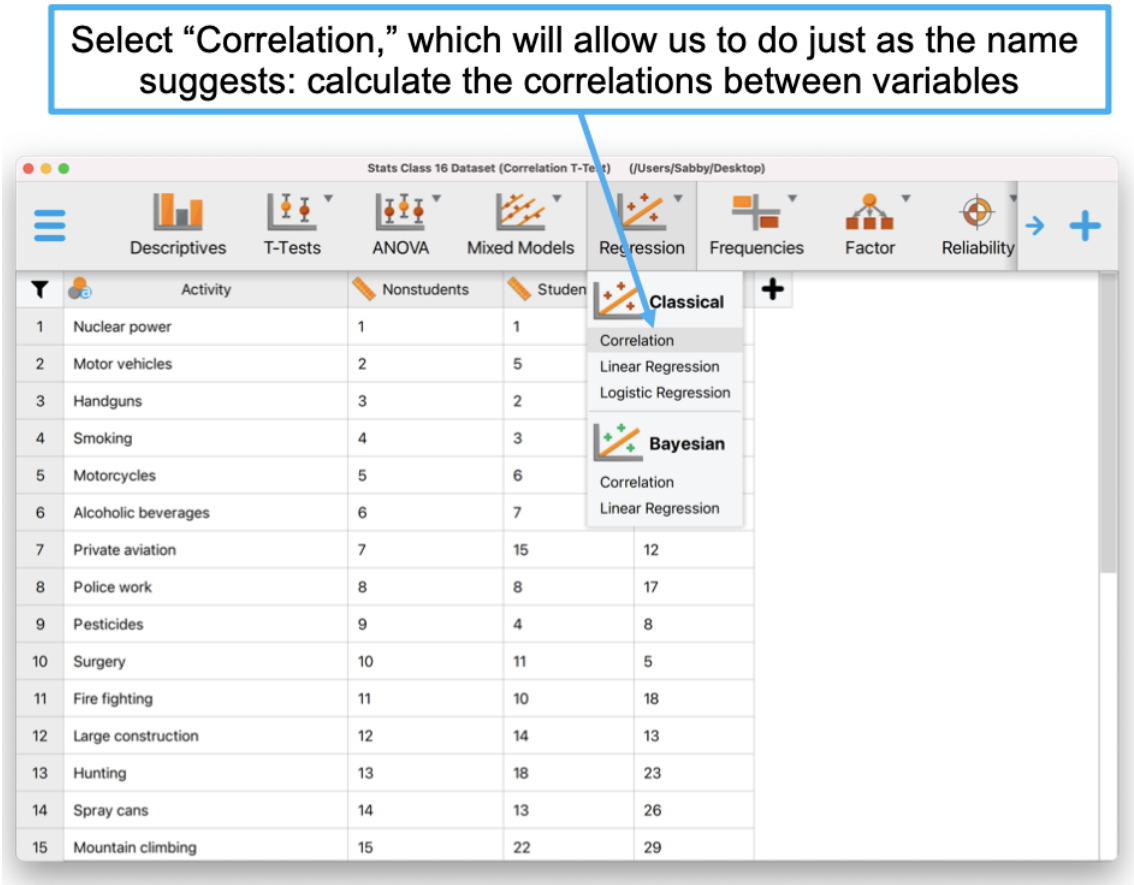
Correlations in JASP

Let's calculate the correlations between these groups' ratings by selecting the "Regression" menu, which contains correlations.



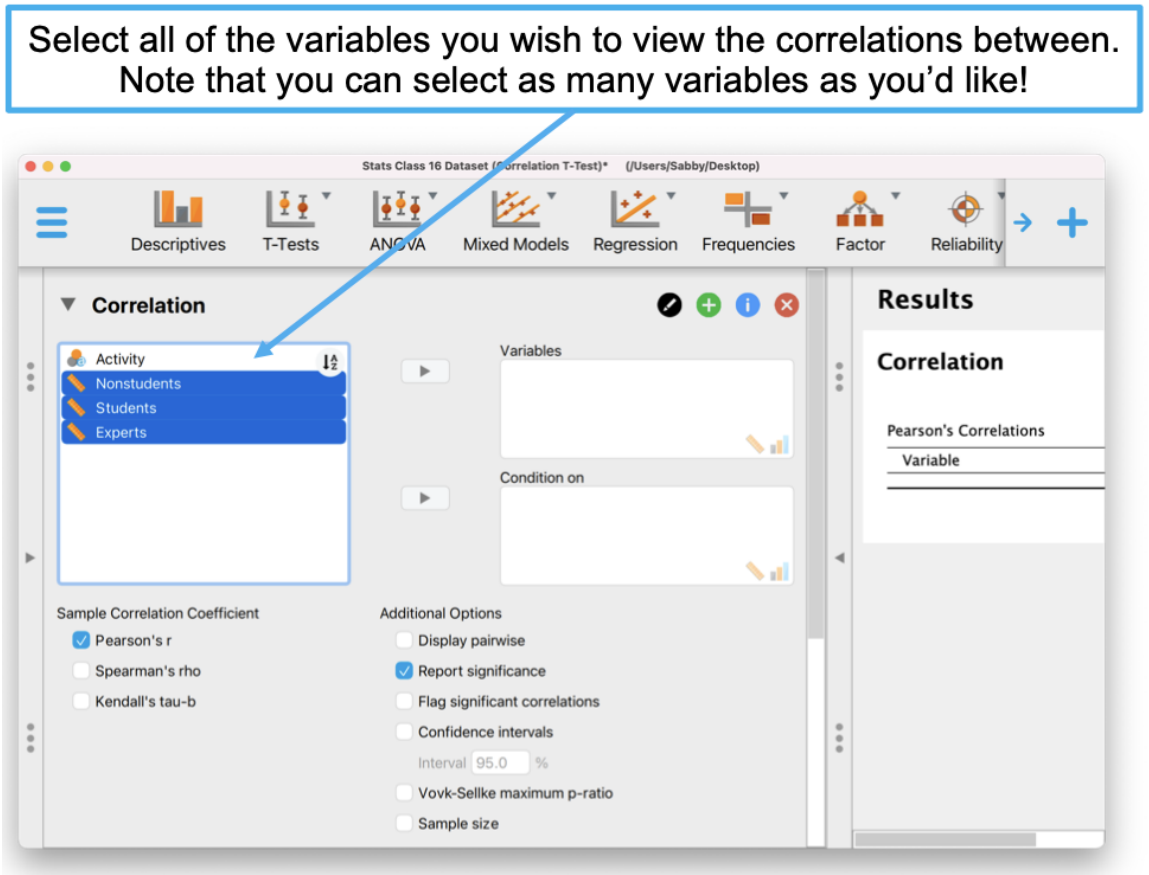
	Activity	Nonstudents	Students	Experts
1	Nuclear power	1	1	20
2	Motor vehicles	2	5	1
3	Handguns	3	2	4
4	Smoking	4	3	2
5	Motorcycles	5	6	6
6	Alcoholic beverages	6	7	3
7	Private aviation	7	15	12
8	Police work	8	8	17
9	Pesticides	9	4	8
10	Surgery	10	11	5
11	Fire fighting	11	10	18
12	Large construction	12	14	13
13	Hunting	13	18	23
14	Spray cans	14	13	26
15	Mountain climbing	15	22	29

Correlations in JASP



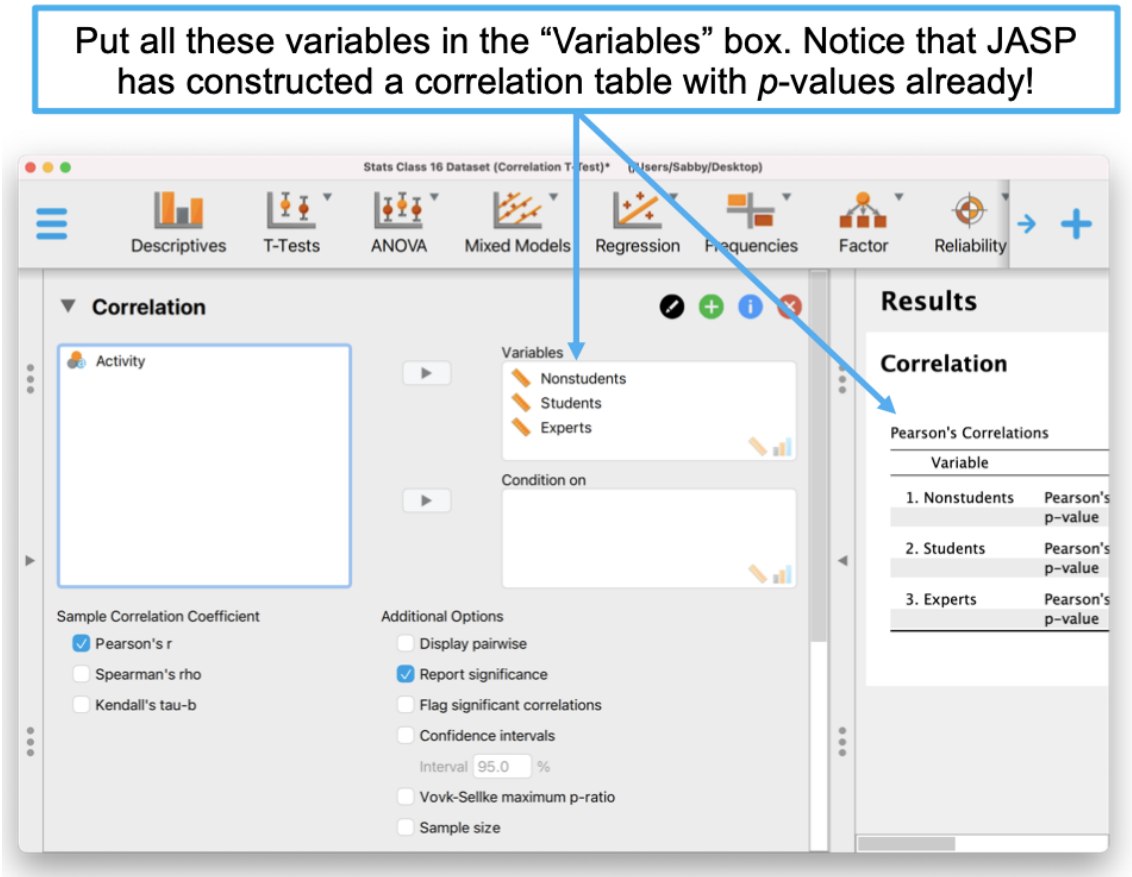
	Activity	Nonstudents	Studen	
1	Nuclear power	1	1	
2	Motor vehicles	2	5	
3	Handguns	3	2	
4	Smoking	4	3	
5	Motorcycles	5	6	
6	Alcoholic beverages	6	7	
7	Private aviation	7	15	12
8	Police work	8	8	17
9	Pesticides	9	4	8
10	Surgery	10	11	5
11	Fire fighting	11	10	18
12	Large construction	12	14	13
13	Hunting	13	18	23
14	Spray cans	14	13	26
15	Mountain climbing	15	22	29

Correlations in JASP



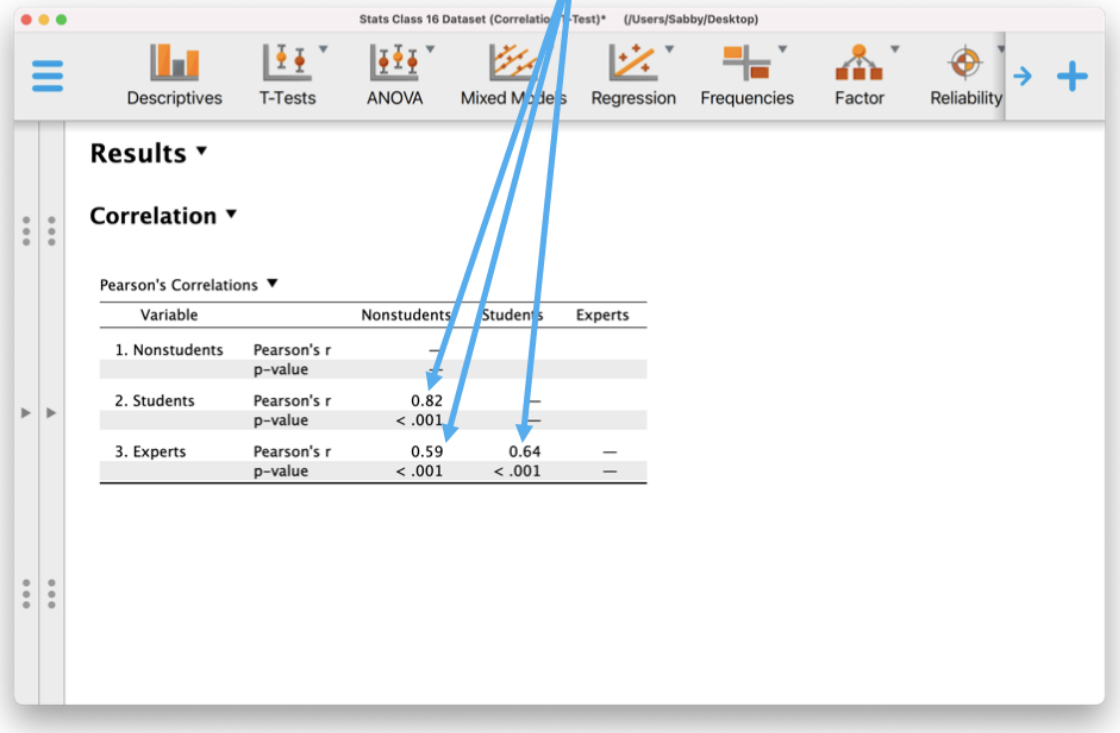
4

Correlations in JASP

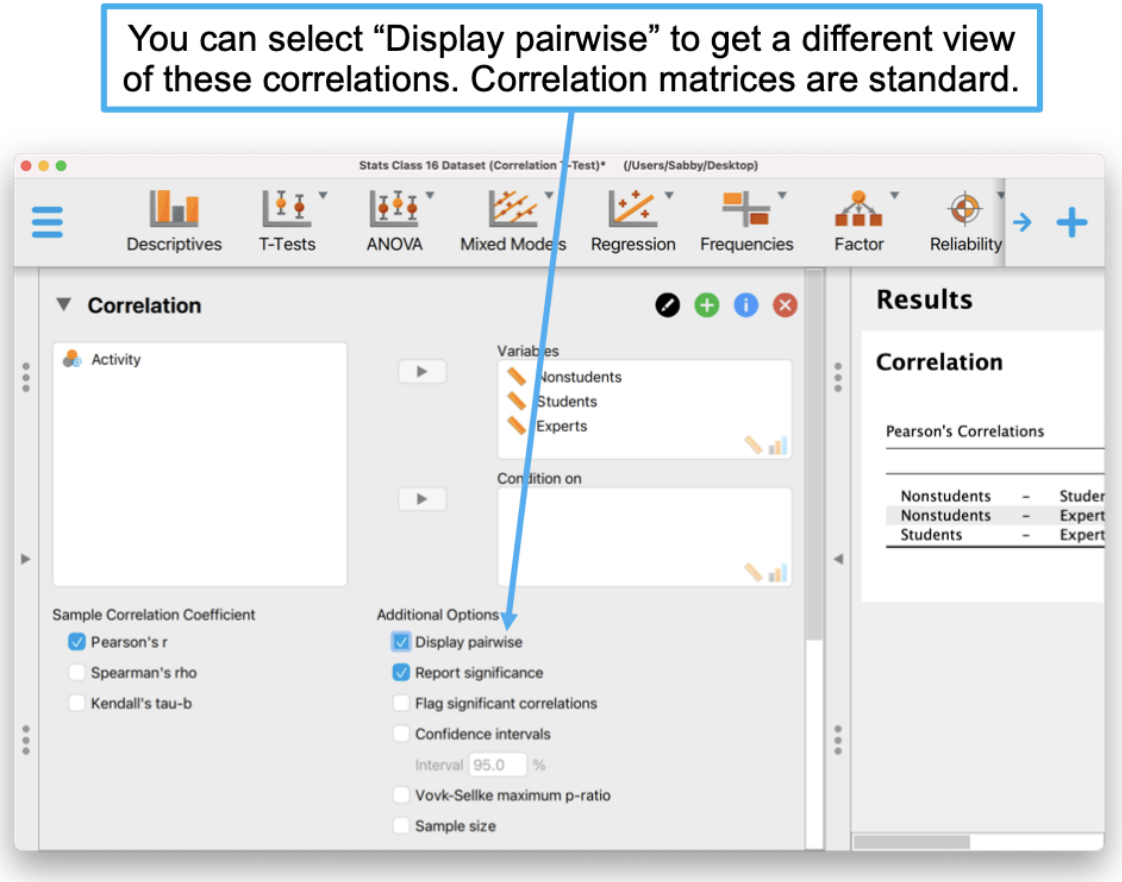


Correlations in JASP

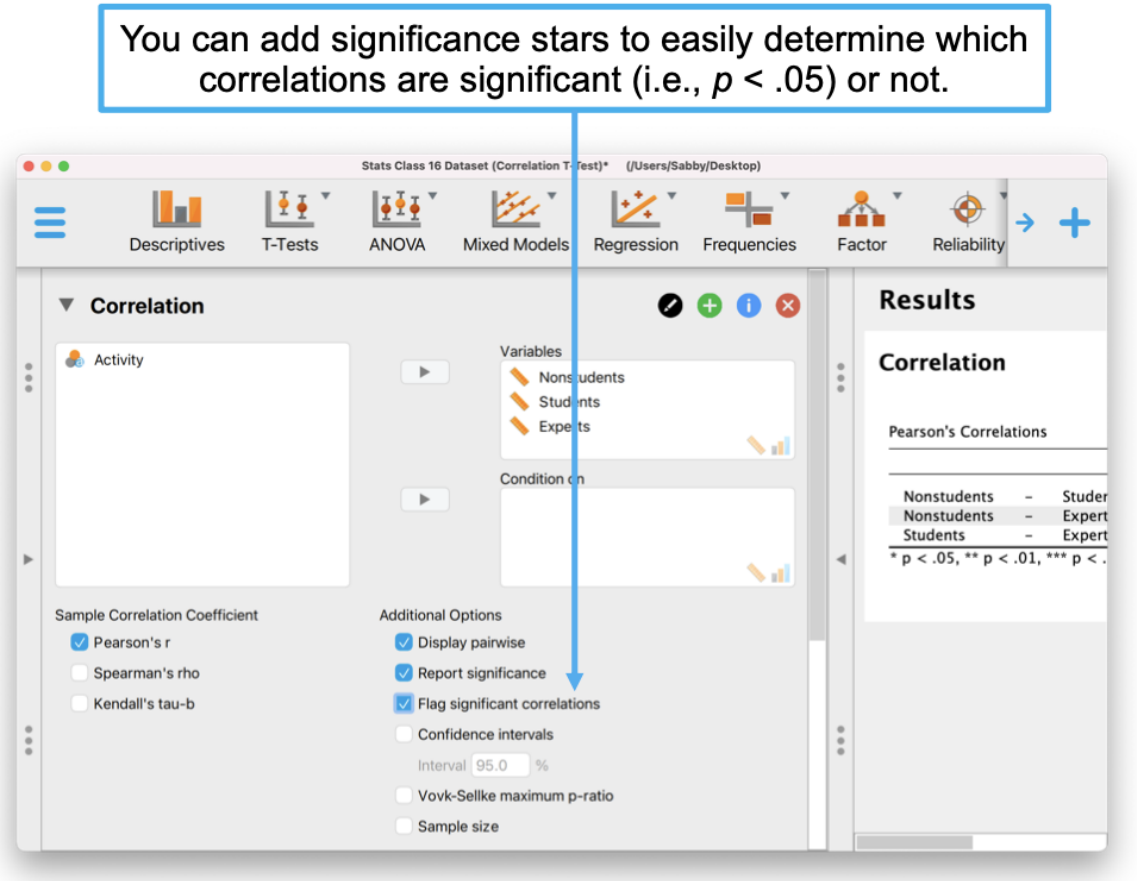
All these *p*-values are less than our alpha of .05, which means there are significant relationships between all our variables.



Correlations in JASP

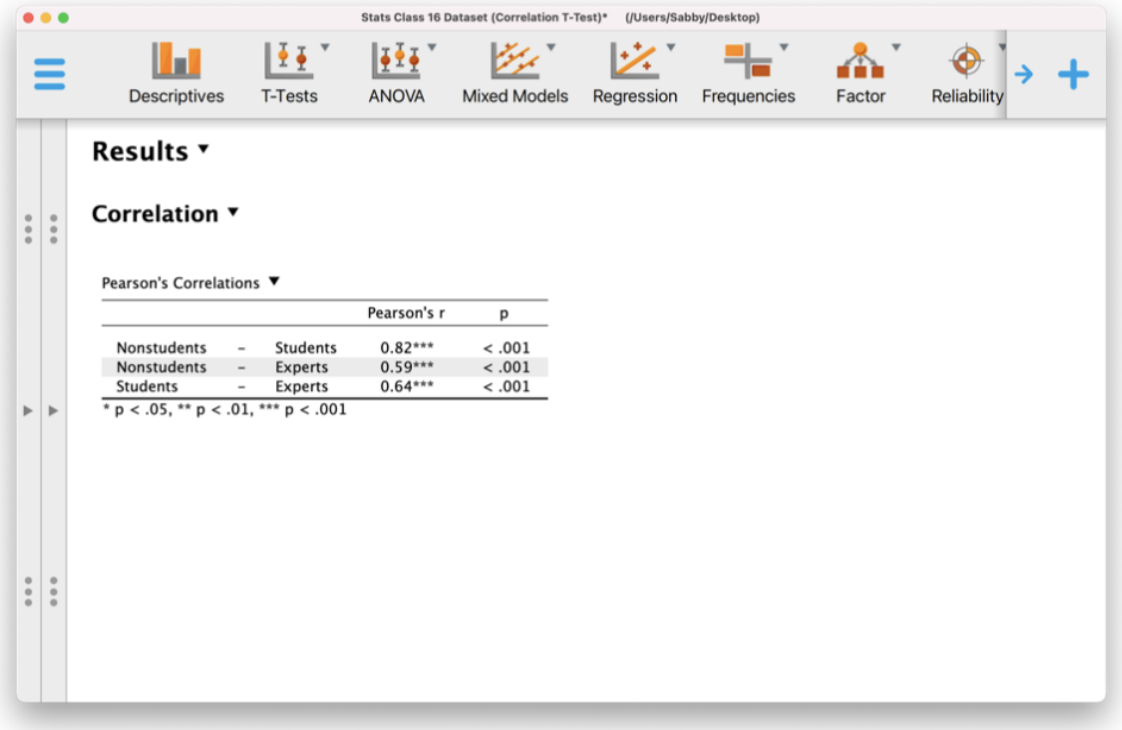


Correlations in JASP

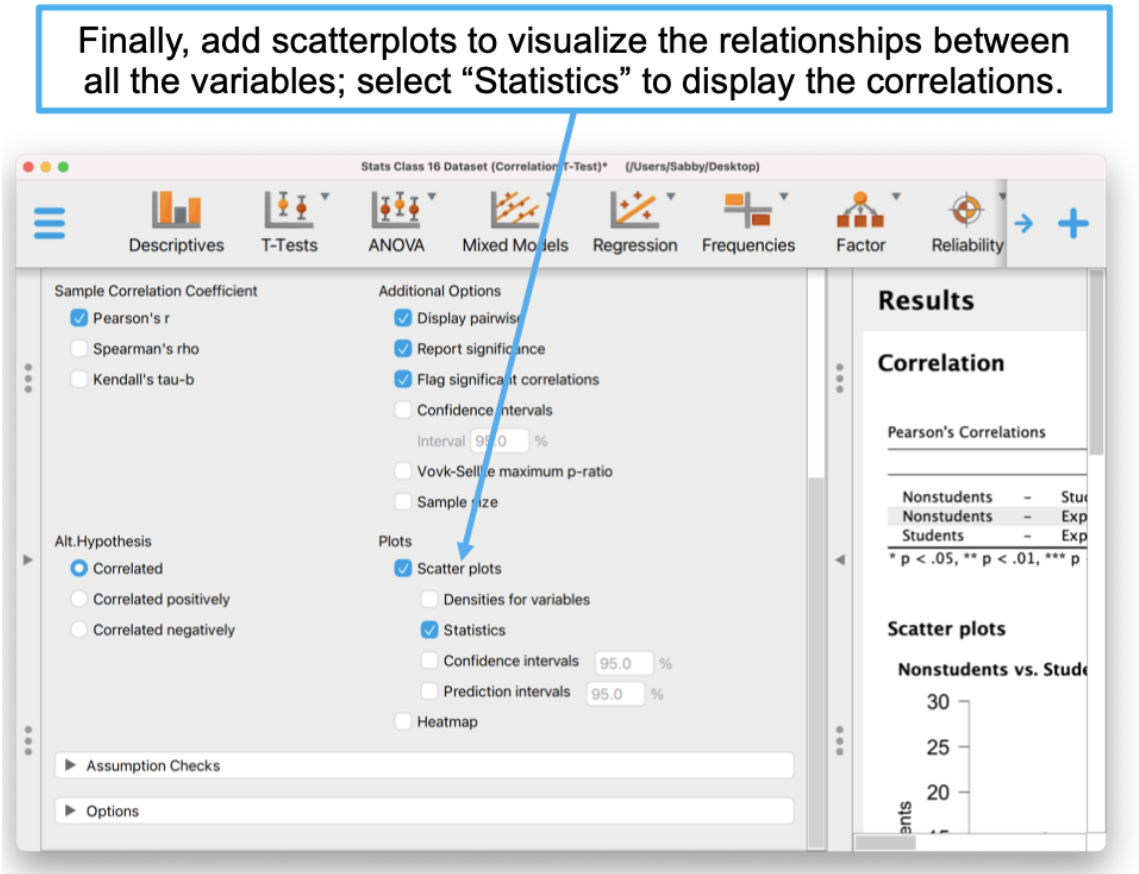


Correlations in JASP

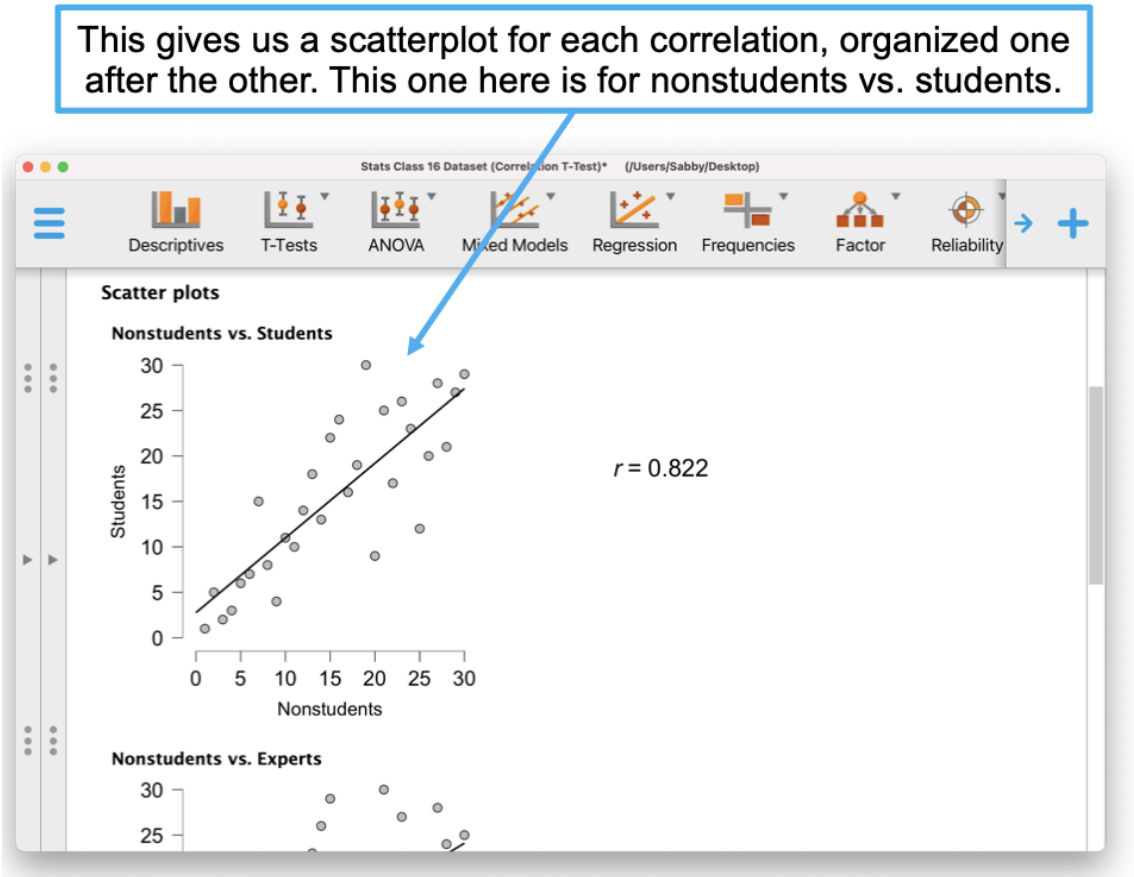
Some people prefer the pairwise view—what's your preference?



Correlations in JASP



Correlations in JASP



Continue one-way ANOVA

Practice Activity

Next time

Lecture

- Making predictions using regression

Reading

- Chapter 15

Quiz 5

- Due tonight 02/23/2022 at 11:59pm MT
    - Ch.13-14