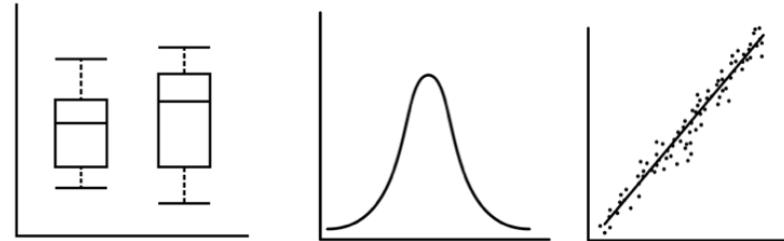


# PSYC 2300

## 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



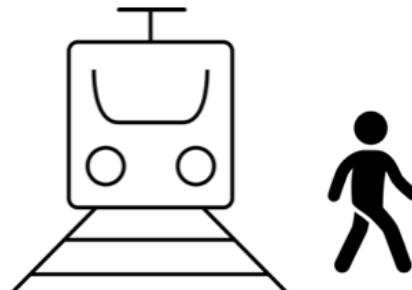
# Review

## One-way ANOVA

# Goal

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

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



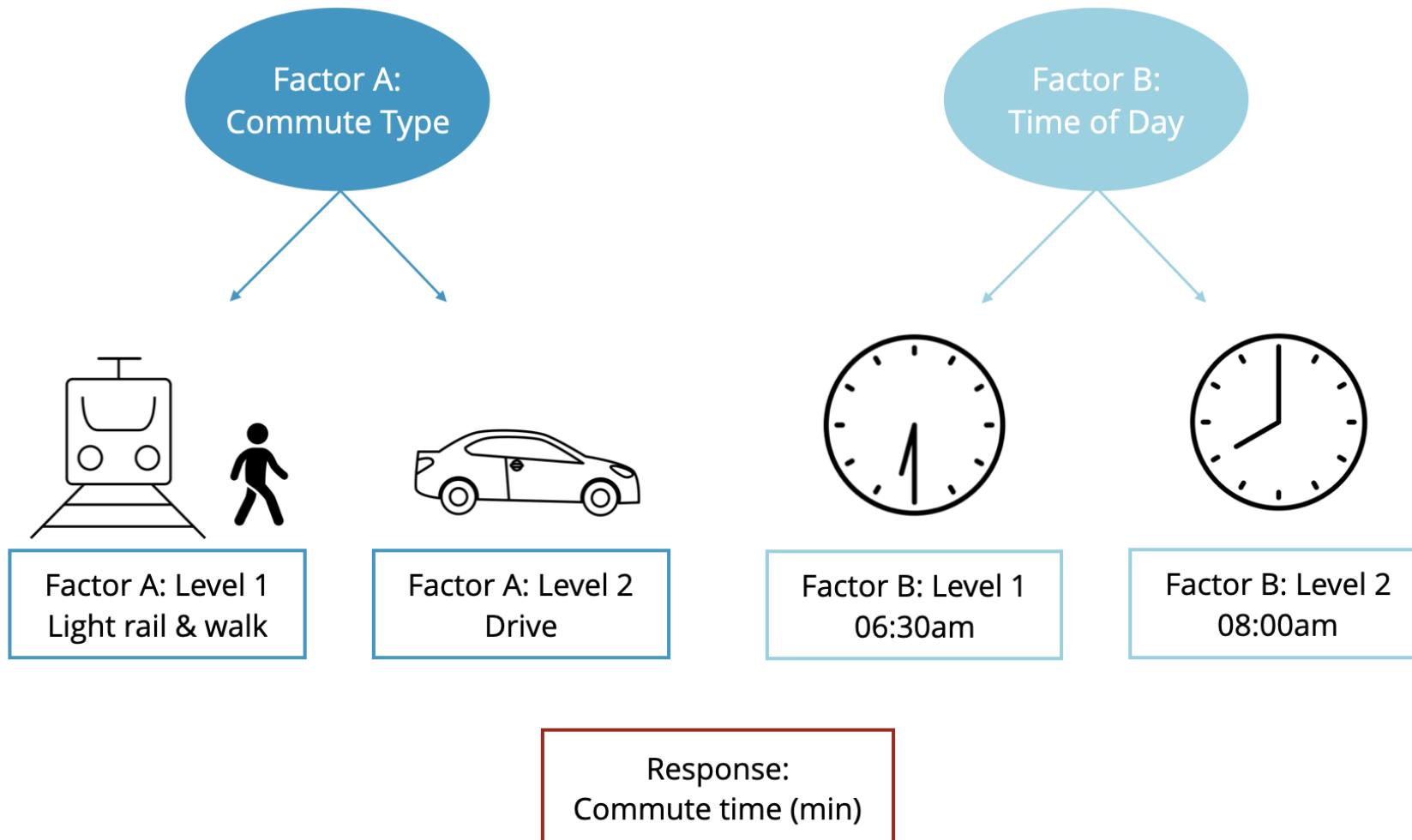
vs.



vs.



# ANOVA: factors and levels

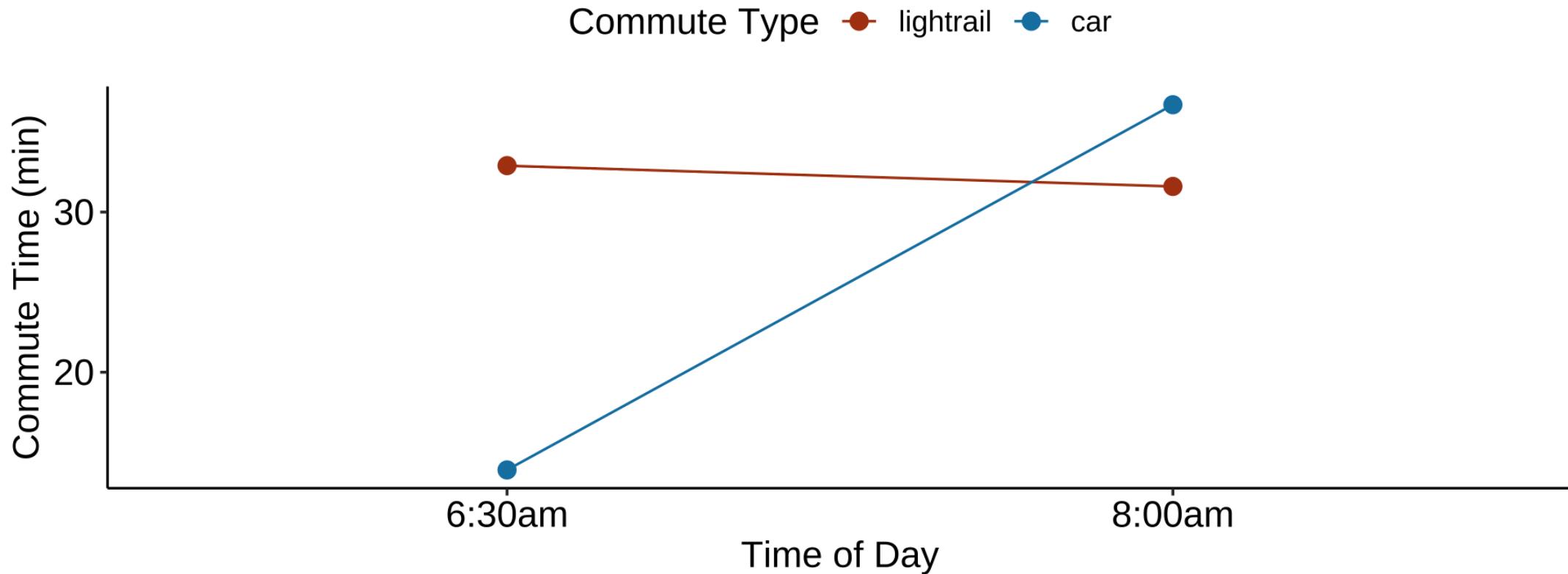


# Factorial Designs

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

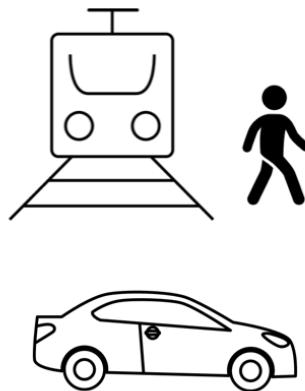


# Factorial designs: Plot



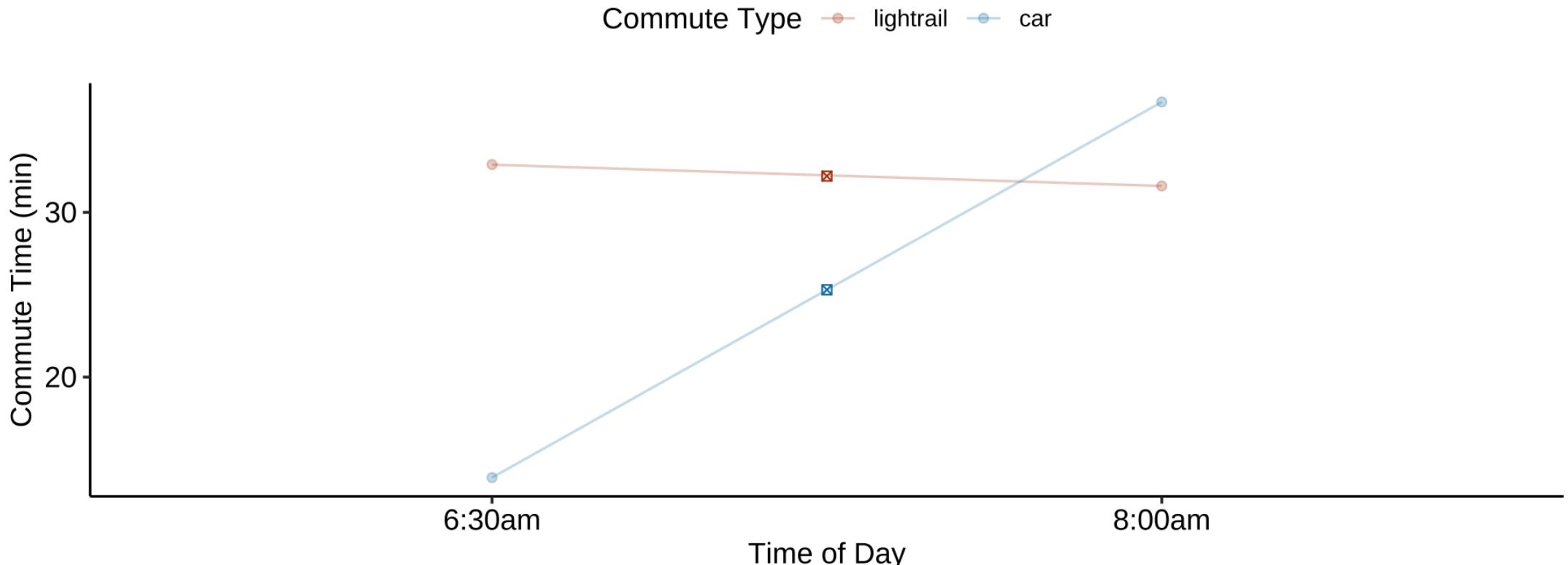
# Main effect of Commute Type

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



<b>Light rail &amp; walk</b>	Average time for light rail and walking
<b>Drive</b>	Average time for driving

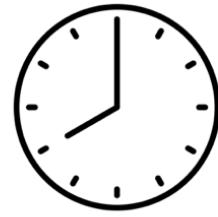
# Commute Type: Plot



Taking the lightrail and walking takes longer than driving

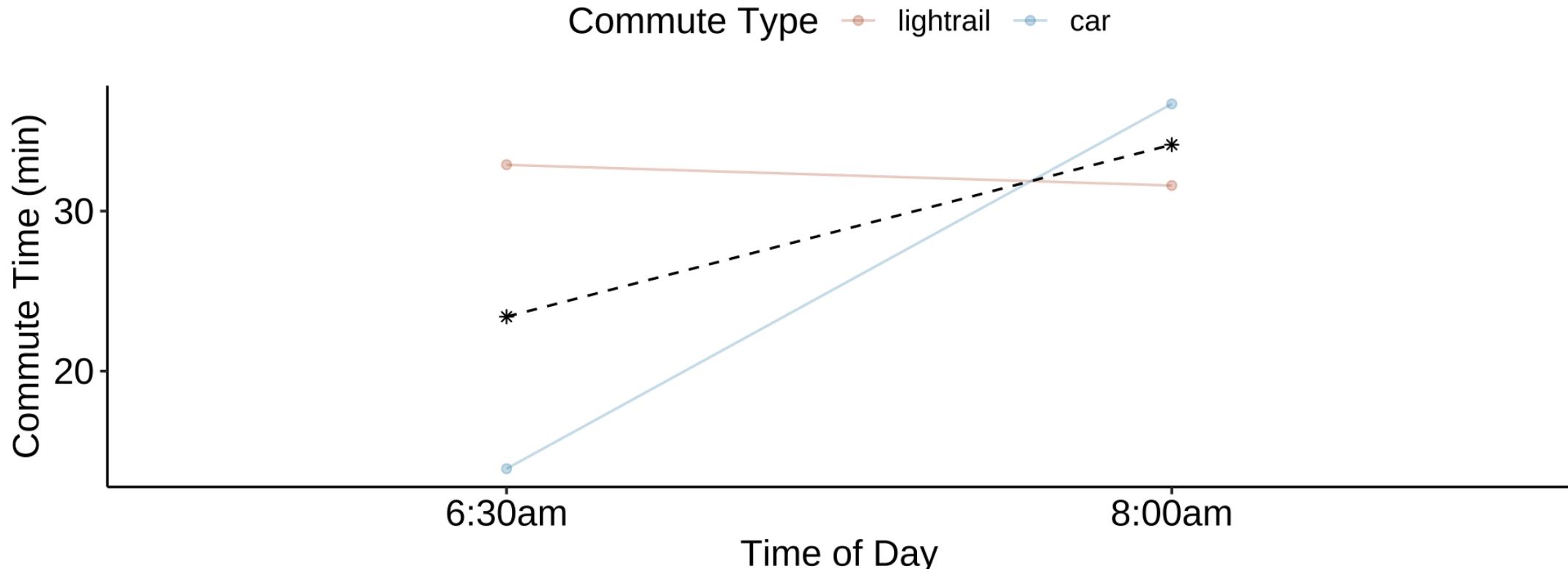
# Main effect of Time of Day

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



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

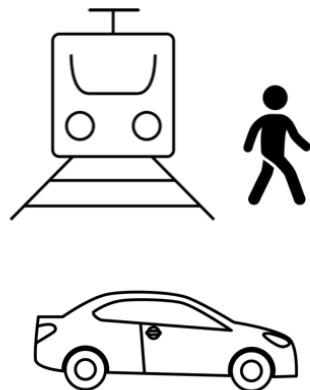
# Time of Day: Plot



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

# Interaction between Commute Type and Time

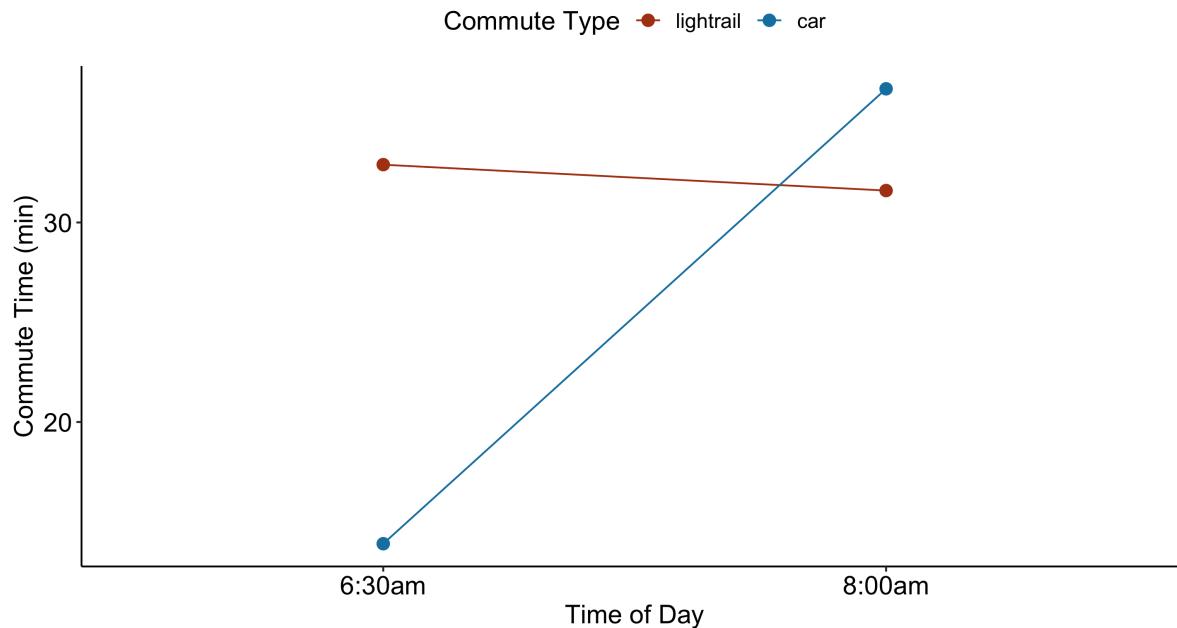
Possible test #3: Effect of specific **Commute Time** and **Time of Day** combinations



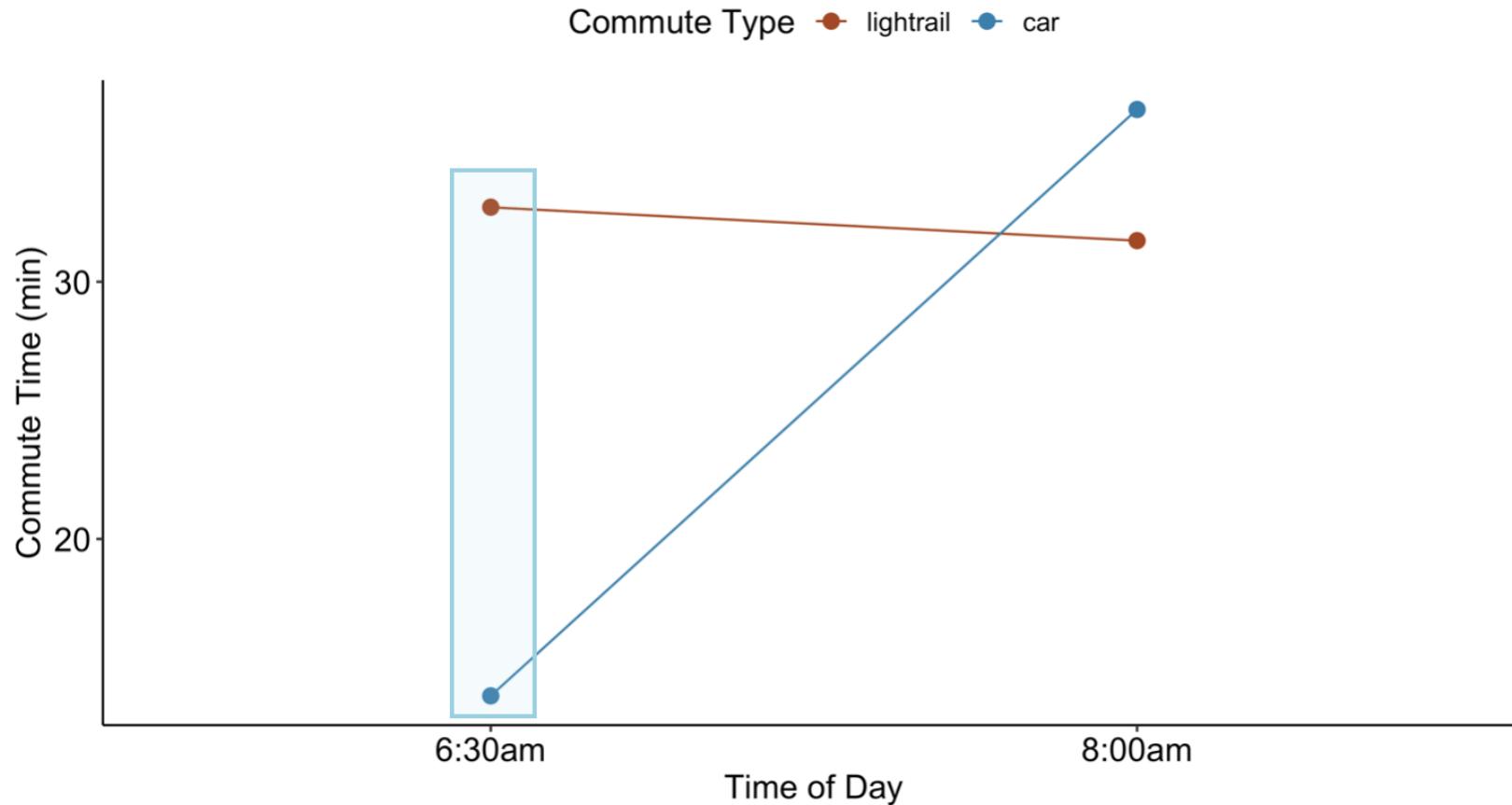
	6:30am	08:00am
Light rail & walk	Average time for light rail and walking at 6:30am	Average time for light rail and walking at 8:00am
Drive	Average time for driving at 6:30am	Average time for driving at 8:00am

# Interaction

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

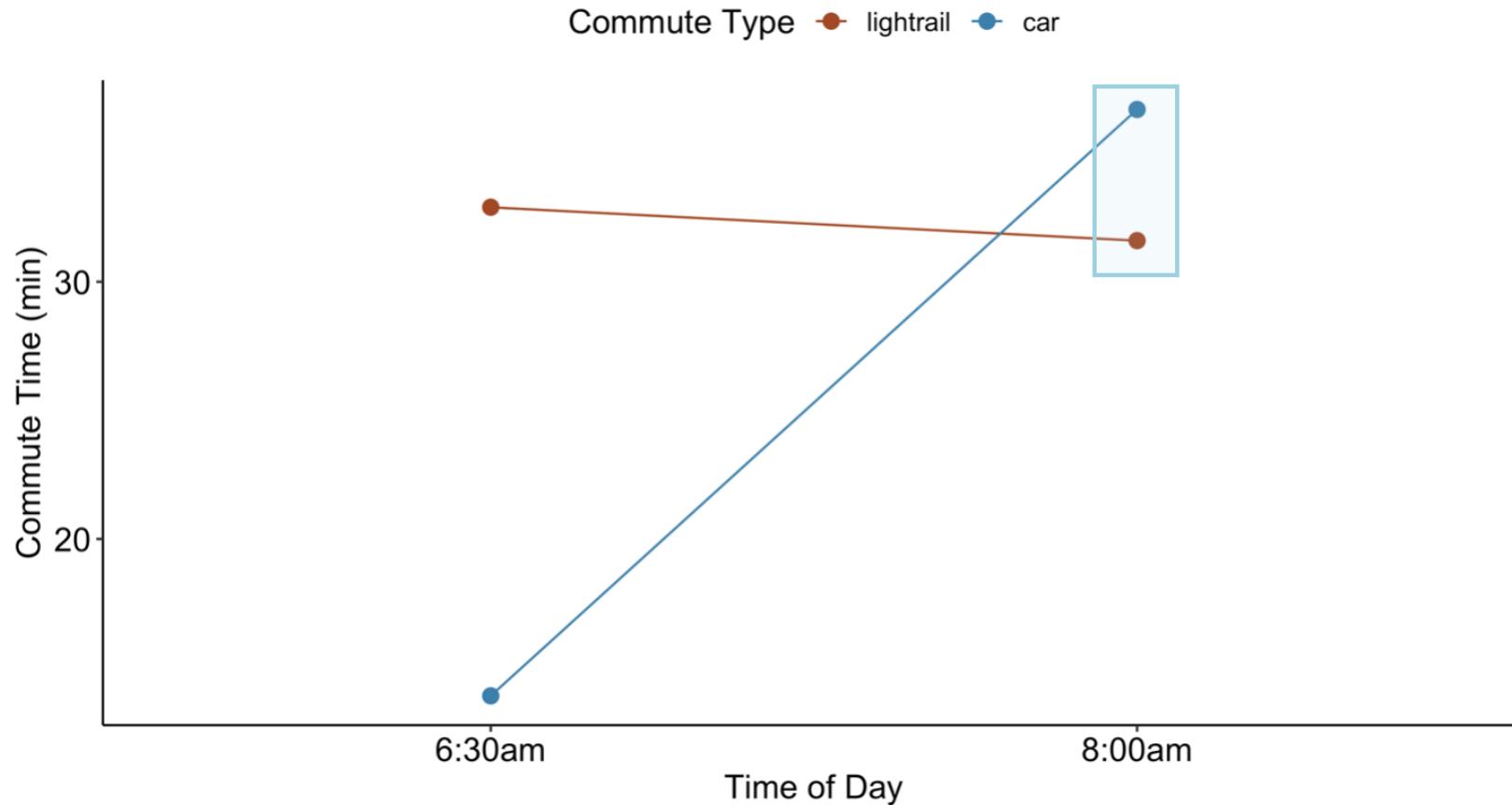


# Interaction



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

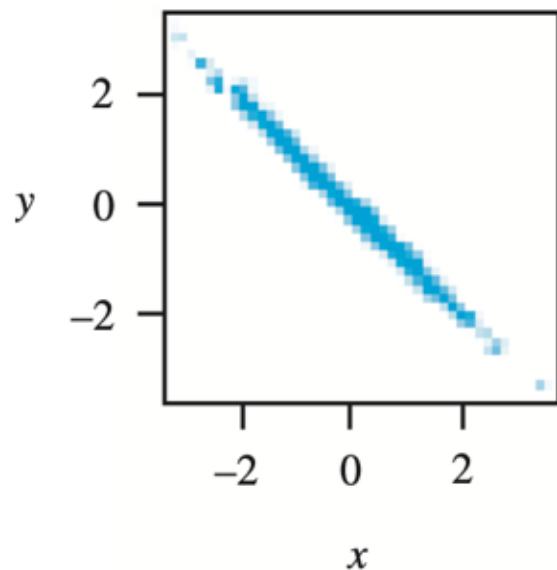
**Correlational designs:** examine the extent to which two variables are associated

## *Examples*

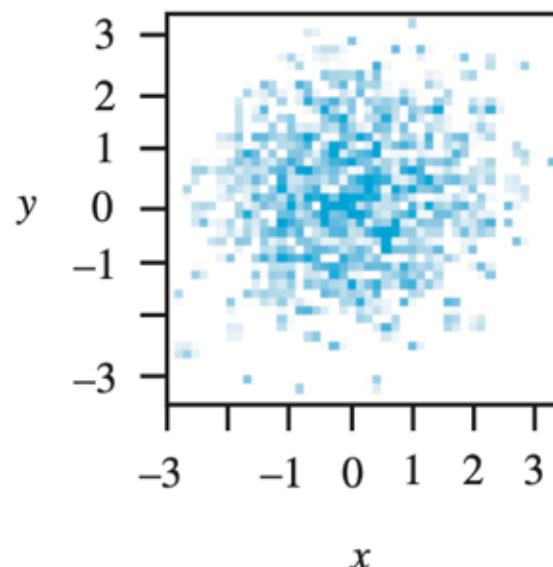
- Do people who are more creative do better in school?
- Is there a relationship between hours studied and exam scores? Class attendance and final grades?
- Are people who have higher marital satisfaction better parents?

# Correlations

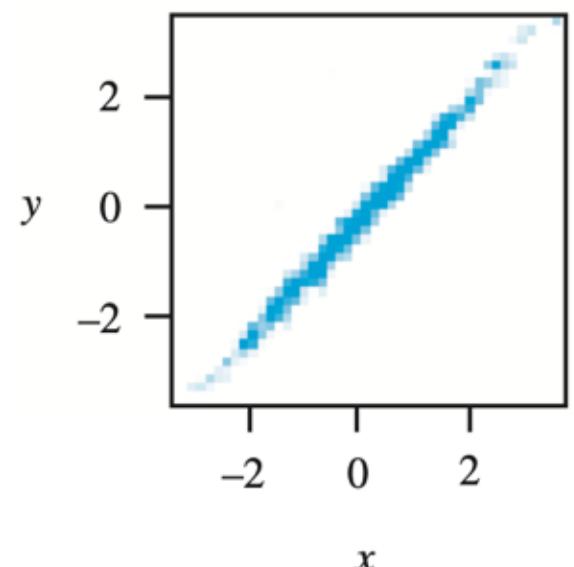
Correlations are either *positive*, *negative*, or *zero*



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*

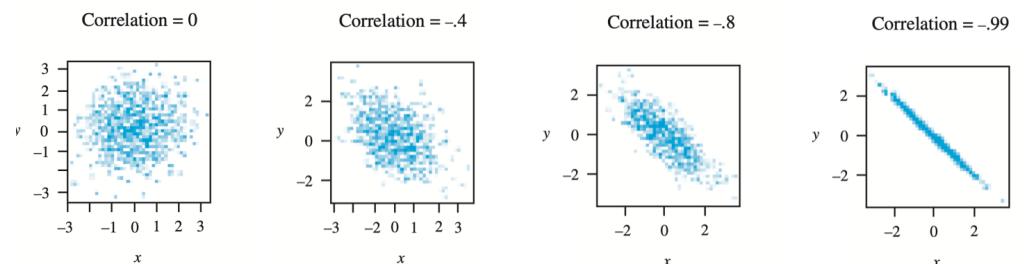
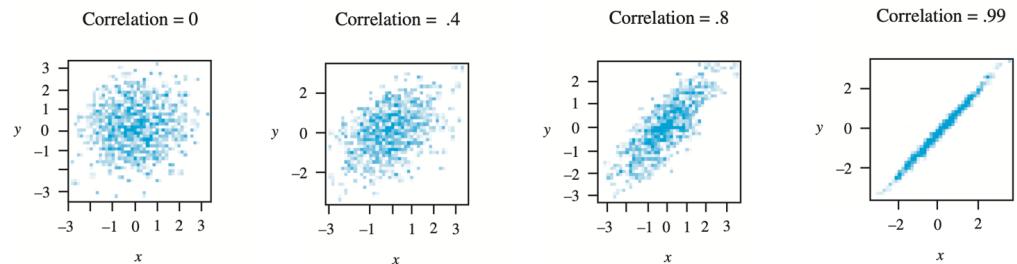
Relationship between running speed and time to finish a race  
(faster running, lower time to finish race)

**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:

$$\begin{array}{ll}\sum X & \sum X^2 \\ \sum Y & \sum Y^2 \\ \sum XY & n\end{array}$$

# 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

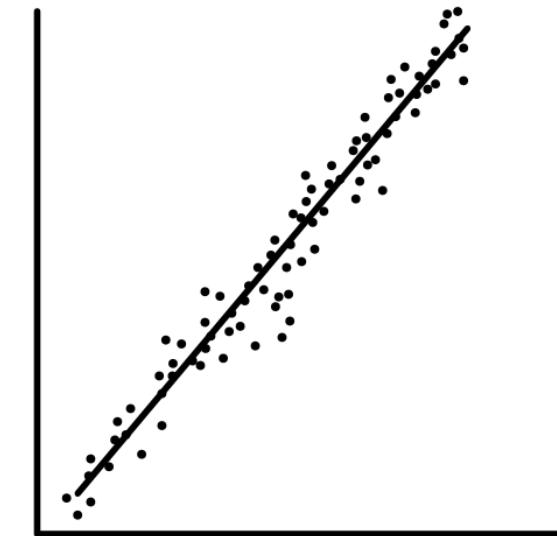
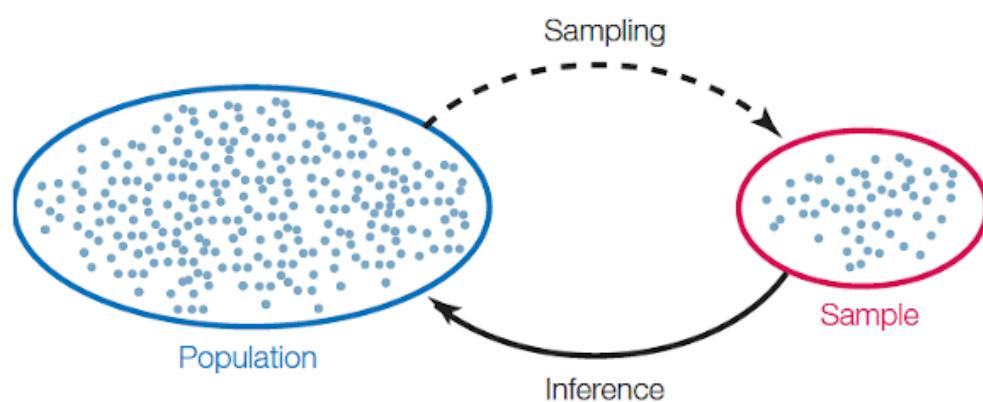
$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$
39	42	289	326	294

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

# 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:  $\rho$  is "rho", the population parameter version of the sample statistic,  $r$

# *t*-test for correlation coefficient

## *t*-test for correlation coefficient

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

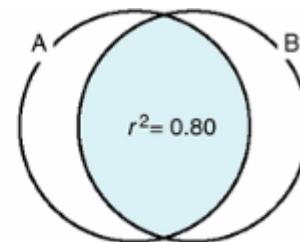
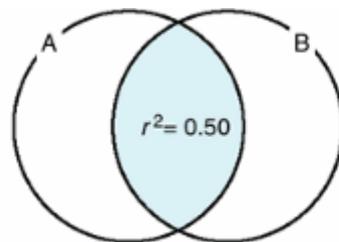
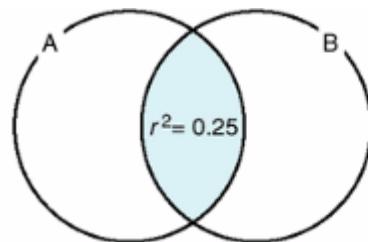
$r_{xy}$  = the correlation between the two variables

$r_{xy}^2$  = 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

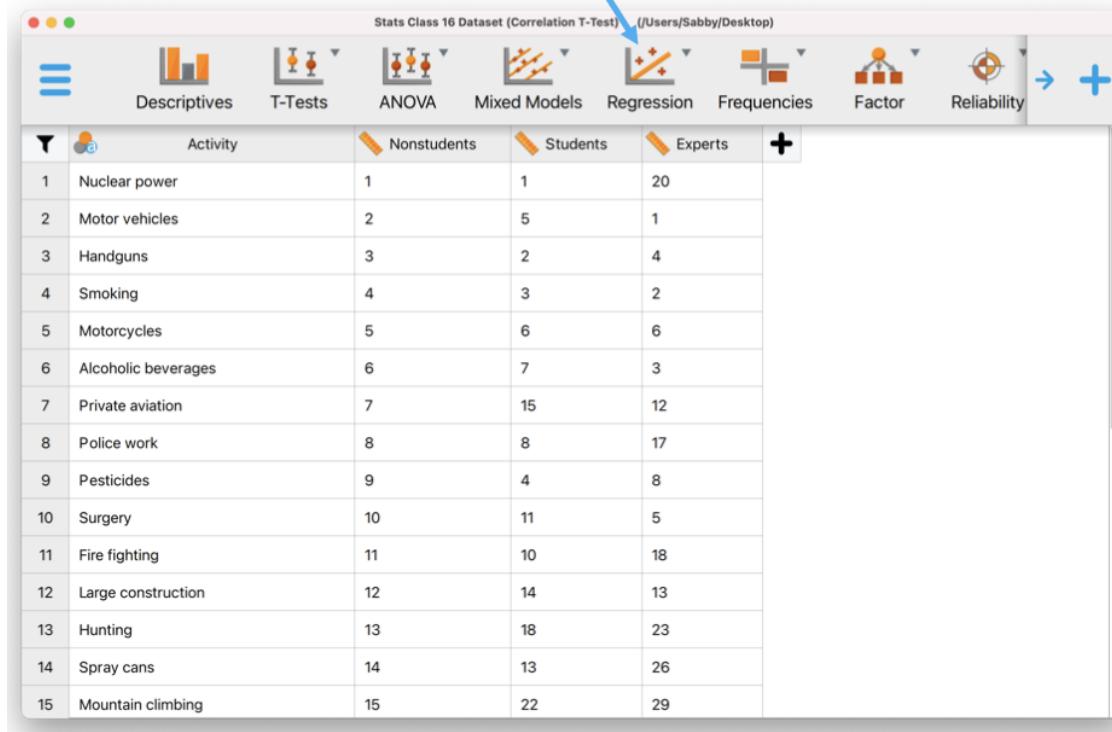
This dataset shows average ratings of the danger of various activities from nonstudents, students, and experts.

The screenshot shows the JASP software interface with a menu bar at the top. Below the menu bar is a toolbar with various statistical analysis icons: Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, and Reliability. To the right of the toolbar are two buttons: a blue arrow pointing right and a blue plus sign. The main area of the window displays a table with 15 rows and 4 columns. The columns are labeled 'Activity', 'Nonstudents', 'Students', and 'Experts'. The rows are numbered 1 through 15 and list various activities: Nuclear power, Motor vehicles, Handguns, Smoking, Motorcycles, Alcoholic beverages, Private aviation, Police work, Pesticides, Surgery, Fire fighting, Large construction, Hunting, Spray cans, and Mountain climbing. The 'Nonstudents' column contains values 1 through 15. The 'Students' column contains values 1 through 22. The 'Experts' column contains values 20, 1, 4, 2, 6, 3, 12, 17, 8, 5, 18, 13, 23, 26, and 29.

	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.



The screenshot shows the JASP software interface. At the top, there is a toolbar with various statistical analysis icons: Descriptives, T-Tests, ANOVA, Mixed Models, Regression (which is highlighted with a blue arrow), Frequencies, Factor, and Reliability. Below the toolbar is a table titled "Stats Class 16 Dataset (Correlation T-Test) (/Users/Sabby/Desktop)". The table has four columns labeled "Activity", "Nonstudents", "Students", and "Experts". The data rows are numbered 1 to 15, listing activities such as Nuclear power, Motor vehicles, Handguns, Smoking, Motorcycles, Alcoholic beverages, Private aviation, Police work, Pesticides, Surgery, Fire fighting, Large construction, Hunting, Spray cans, and Mountain climbing, along with their corresponding ratings for each group.

	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

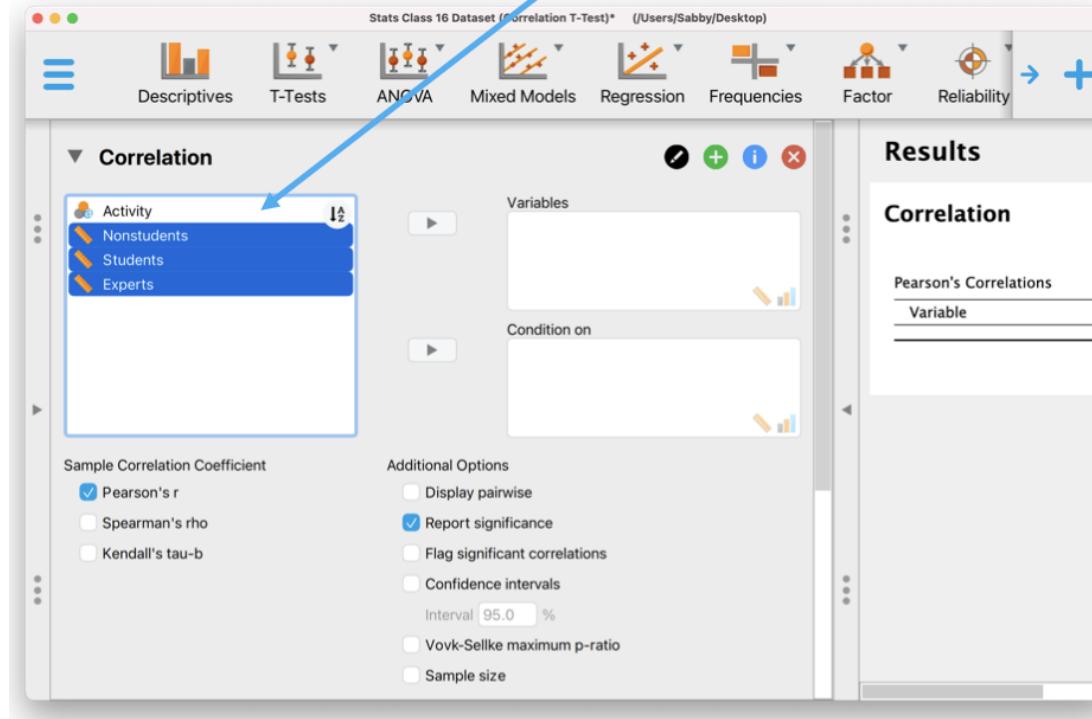
Select “Correlation,” which will allow us to do just as the name suggests: calculate the correlations between variables

The screenshot shows the JASP software interface. At the top, there is a toolbar with various statistical analysis icons: Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, and Reliability. Below the toolbar is a data table titled "Stats Class 16 Dataset (Correlation T-Test)" located at "/Users/Sabby/Desktop". The table has columns labeled "Activity", "Nonstudents", "Students", and two unnamed columns. The data consists of 15 rows, each containing an activity name and three numerical values. A blue arrow points from the text above to the "Correlation" icon in the toolbar. A tooltip for the "Correlation" icon is displayed, showing "Classical" with "Correlation" and "Linear Regression" listed under it. Another tooltip for the "Bayesian" icon is also visible, showing "Correlation" and "Linear Regression".

	Activity	Nonstudents	Students		
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

Select all of the variables you wish to view the correlations between.  
Note that you can select as many variables as you'd like!



# Correlations in JASP

Put all these variables in the “Variables” box. Notice that JASP has constructed a correlation table with  $p$ -values already!

The screenshot shows the JASP software interface with the following details:

- Toolbar:** Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, Reliability.
- Left Panel (Correlation):**
  - Variables: Nonstudents, Students, Experts.
  - Condition on: (empty)
- Bottom Left (Sample Correlation Coefficient):**
  - Pearson's r
  - Spearman's rho
  - Kendall's tau-b
- Bottom Left (Additional Options):**
  - Display pairwise
  - Report significance
  - Flag significant correlations
  - Confidence intervals  
Interval 95.0 %
  - Vovk-Sellke maximum p-ratio
  - Sample size
- Right Panel (Results):**

**Correlation**

Variable	Pearson's Correlations
1. Nonstudents	Pearson's p-value
2. Students	Pearson's p-value
3. Experts	Pearson's p-value

# Correlations in JASP

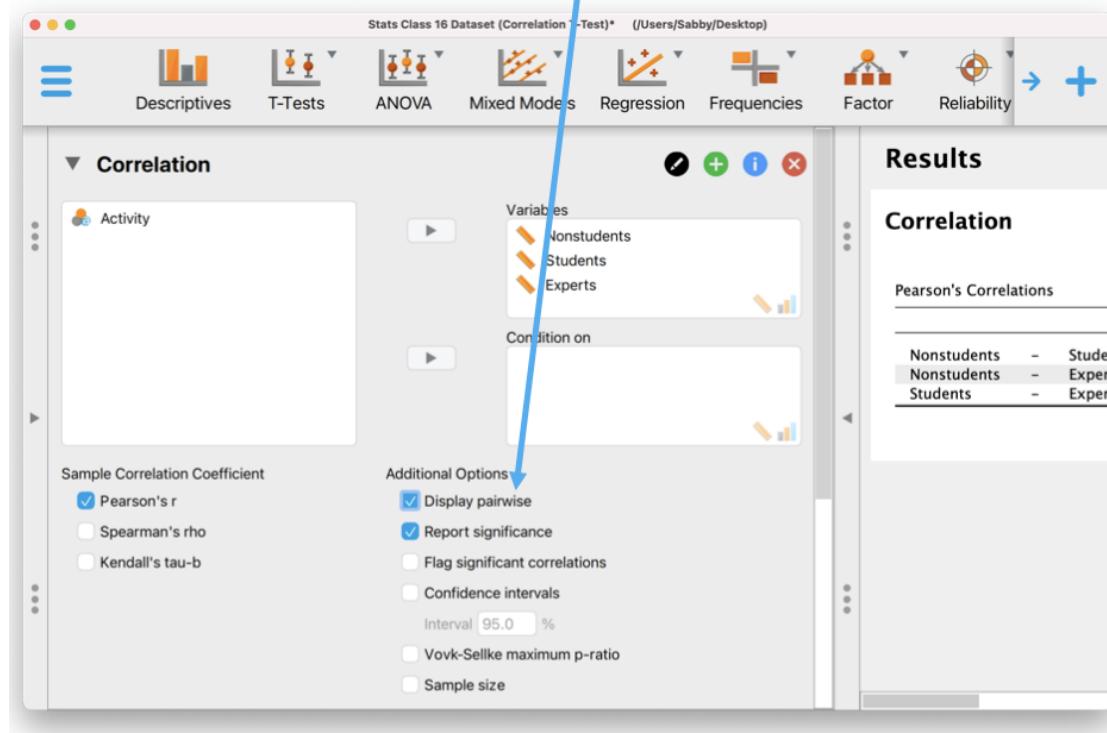
All these  $p$ -values are less than our alpha of .05, which means there are significant relationships between all our variables.

The screenshot shows the JASP software interface with a blue header bar. Below the header, there is a toolbar with various statistical analysis icons: Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, and Reliability. The 'Results' section is expanded, showing the 'Correlation' section. Under 'Correlation', the 'Pearson's Correlations' table is displayed. The table has 'Variable' in the first column and three other variables ('Nonstudents', 'Students', 'Experts') in the subsequent columns. The table includes Pearson's r and p-value for each correlation. Blue arrows point from the text above to the p-values in the 'Students' and 'Experts' columns of the table.

Variable	Nonstudents	Students	Experts
1. Nonstudents	Pearson's r p-value	— —	—
2. Students	Pearson's r p-value	0.82 < .001	—
3. Experts	Pearson's r p-value	0.59 < .001	0.64 < .001

# Correlations in JASP

You can select “Display pairwise” to get a different view of these correlations. Correlation matrices are standard.



# Correlations in JASP

You can add significance stars to easily determine which correlations are significant (i.e.,  $p < .05$ ) or not.

The screenshot shows the JASP software interface with the following details:

- Top Bar:** Stats Class 16 Dataset (Correlation T-Test)\*/ (/Users/Sabby/Desktop)
- Toolbars:** Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, Reliability.
- Left Panel:** A tree view under "Correlation" with "Activity" selected. Below it, "Sample Correlation Coefficient" has "Pearson's r" checked, while "Spearman's rho" and "Kendall's tau-b" are unchecked.
- Middle Panel:** "Variables" list includes "Nonstudents", "Students", and "Experts". "Condition on" is empty.
- Bottom Panel:** "Additional Options" section has "Display pairwise", "Report significance" (checked), "Flag significant correlations" (checked), "Confidence intervals" (unchecked), "Vovk-Sellke maximum p-ratio" (unchecked), and "Sample size" (unchecked). The "Interval" field is set to 95.0%.
- Right Panel:** "Results" section titled "Correlation" shows Pearson's Correlations for Nonstudents, Students, and Experts. The correlation matrix is:

	Nonstudents	Students	Experts
Nonstudents	-		
Students	-		
Experts	-		

\*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

# Correlations in JASP

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

The screenshot shows the JASP software interface with the following details:

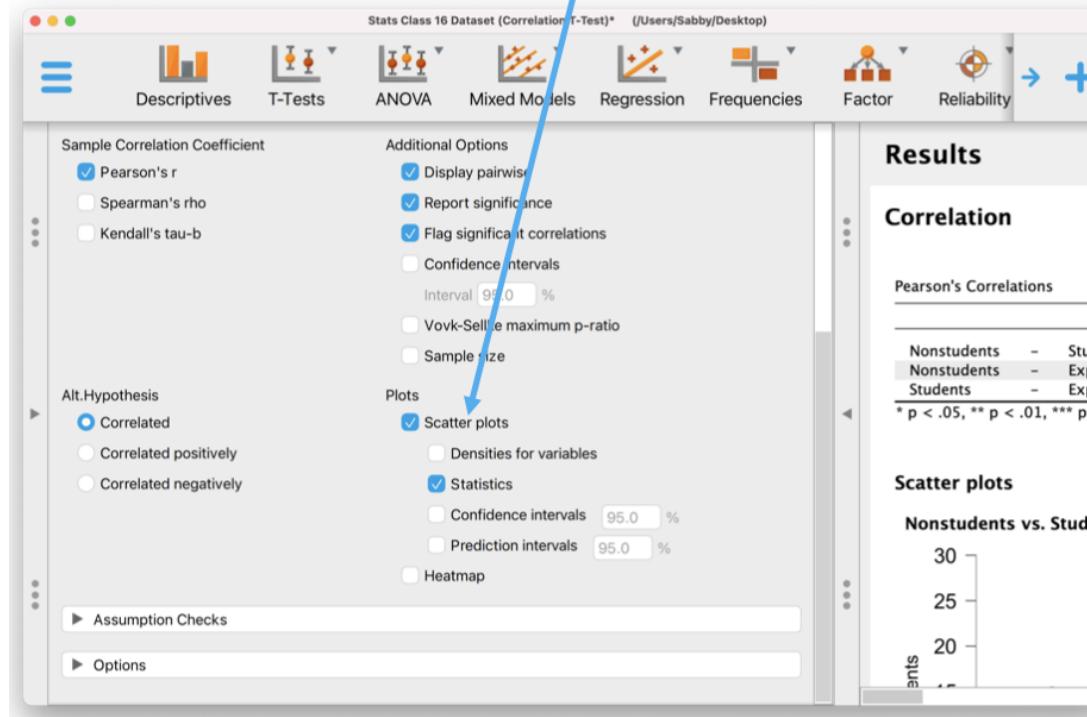
- Toolbar:** Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, Factor, Reliability.
- Title Bar:** Stats Class 16 Dataset (Correlation T-Test)\* (/Users/Sabby/Desktop)
- Results Panel:** Correlation
- Data Table:** Pearson's Correlations

		Pearson's r	p
Nonstudents	- Students	0.82***	< .001
Nonstudents	- Experts	0.59***	< .001
Students	- Experts	0.64***	< .001

- Note:** \* p < .05, \*\* p < .01, \*\*\* p < .001

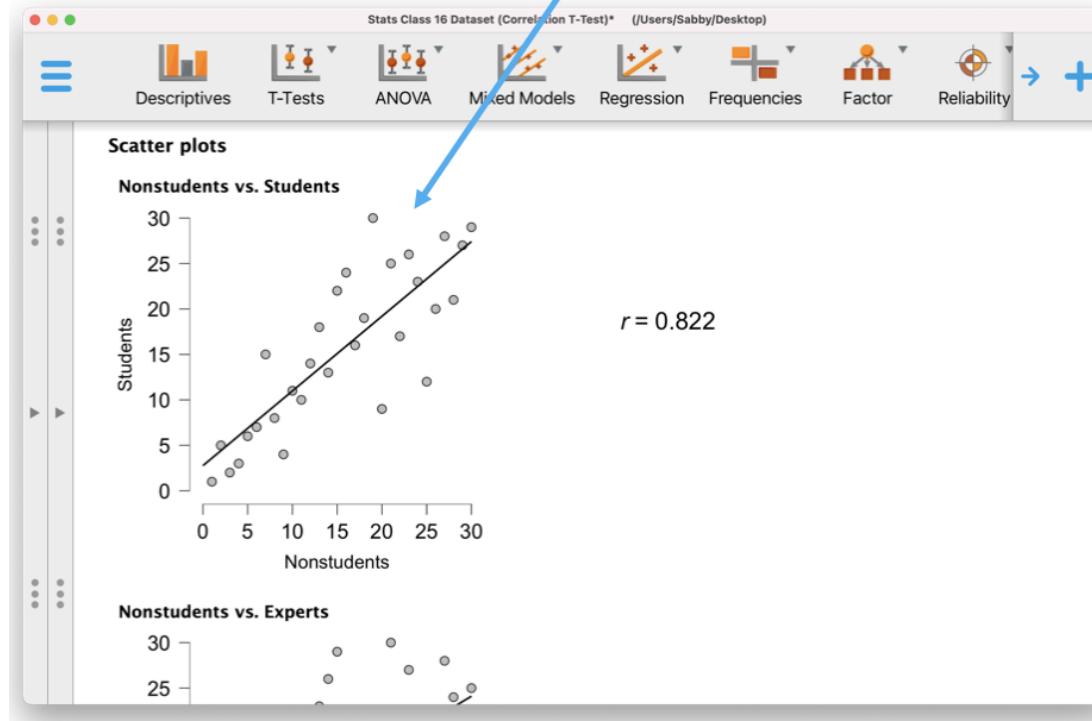
# Correlations in JASP

Finally, add scatterplots to visualize the relationships between all the variables; select “Statistics” to display the correlations.



# Correlations in JASP

This gives us a scatterplot for each correlation, organized one after the other. This one here is for nonstudents vs. students.



# 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

