Applied Statistical Analysis

EDUC 6050

Week 8

Finding clarity using data

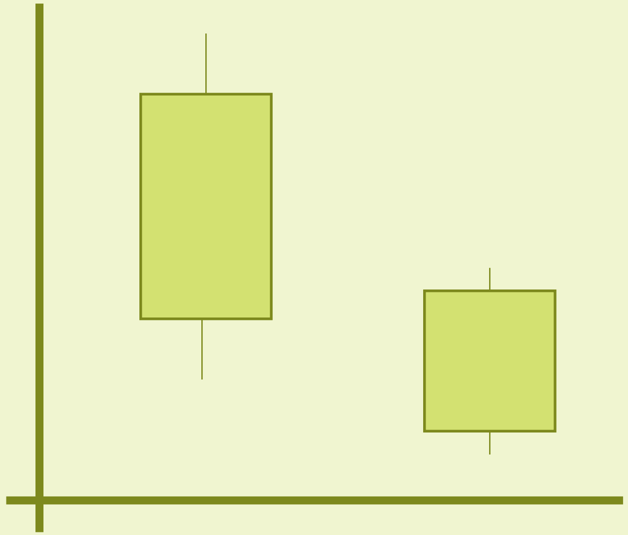
Today

1. Relationships!
2. Correlation and Intro to Regression
3. Chapter 13 in Book

Comparing Means

Is one group different than the other(s)?
- Z-tests
- T-tests
- ANOVA

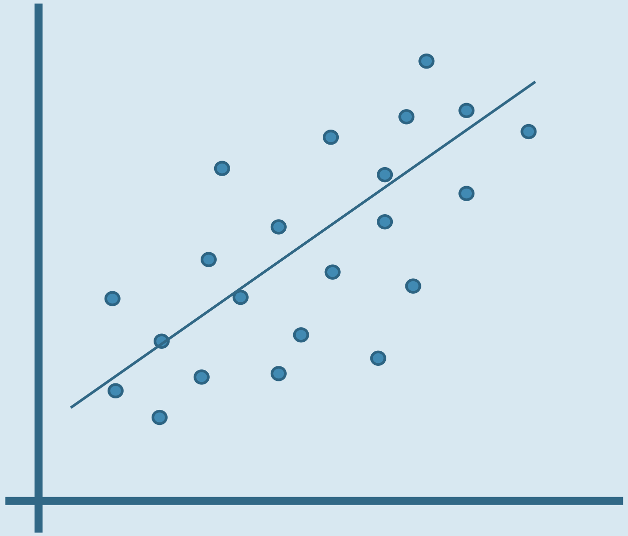
We compare the means and use the variability to decide if the difference is significant



Assessing Relationships

Is there a relationship between the two variables?
- Correlation
- Regression

We look at how much the variables “move together”



Correlation

- It is a whole class of methods
- Generally used with observational designs
- Has similar assumptions to t-test
- Is a measure of effect size
- Very related (and based on) z-scores
- Tells us direction and strength of a relationship between *two* variables

Correlation and Z-Scores

- Z-score is a univariate statistic (only uses info from ONE variable)
- Correlation is essentially the z-score between TWO variables

$$r = \frac{\sum Z_x Z_y}{N - 1}$$

Correlation and Z-Scores

- Z-score is a univariate statistic (only uses info from ONE variable)
- Correlation is essentially the z-score between TWO variables

z-score of variable x

$$r = \frac{\sum Z_x Z_y}{N-1}$$

z-score of variable y

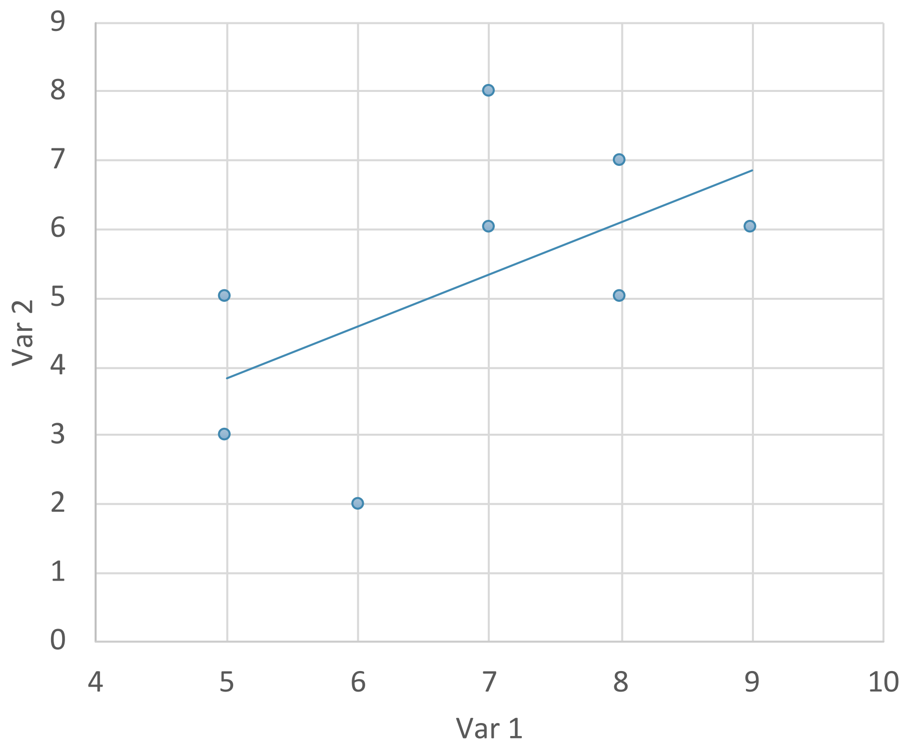
General Requirements

1. Two or more continuous variables,
2. Not necessarily directional (one causes the other)

ID	Var 1	Var 2
1	8	7
2	6	2
3	9	6
4	7	6
5	7	8
6	8	5
7	5	3
8	5	5

General Requirements

1. Two or more continuous variables,
2. Not necessarily directional (one causes the other)
3. Linear Relationship (or at least ordinal)



Hypothesis Testing with Correlation

The same 6 step approach!

1. Examine Variables to Assess Statistical Assumptions
2. State the Null and Research Hypotheses (symbolically and verbally)
3. Define Critical Regions
4. Compute the Test Statistic
5. Compute an Effect Size and Describe it
6. Interpreting the results

1 Examine Variables to Assess Statistical Assumptions

Basic Assumptions

1. Independence of data
2. Appropriate measurement of variables for the analysis
3. Normality of distributions
4. Homoscedastic

1 Examine Variables to Assess Statistical Assumptions

Basic Assumptions

1. Independence of data
2. Appropriat
   for the a
3. Normality
4. Homoscedastic

Individuals are independent of each other (one person’s scores does not affect another’s)

1 Examine Variables to Assess Statistical Assumptions

Basic Assumptions

1. Independence of data
2. **Appropriate measurement** of variables for the analysis
3. Normality of distributions
4. Homo

Here we need interval/ratio variables

1 Examine Variables to Assess Statistical Assumptions

Basic Assumptions

1. Independe
2. Appropria
   for the analysis
3. Normality of distributions
4. Homoscedastic

Multivariate normality (the two variables are jointly normal)

1 Examine Variables to Assess Statistical Assumptions

Basic Assumptions

1. Independence of data
2. Appropri
   for
3. Normalit
4. Homoscedastic

Variance around the line should be roughly equal across the whole line

1 Examine Variables to Assess Statistical Assumptions

Examining the Basic Assumptions

1. **Independence:** random sample
2. **Appropriate measurement:** know what your variables are
3. **Normality:** Histograms, **Q-Q**, skew and kurtosis
4. **Homoscedastic:** Scatterplots

2 State the Null and Research Hypotheses (symbolically and verbally)

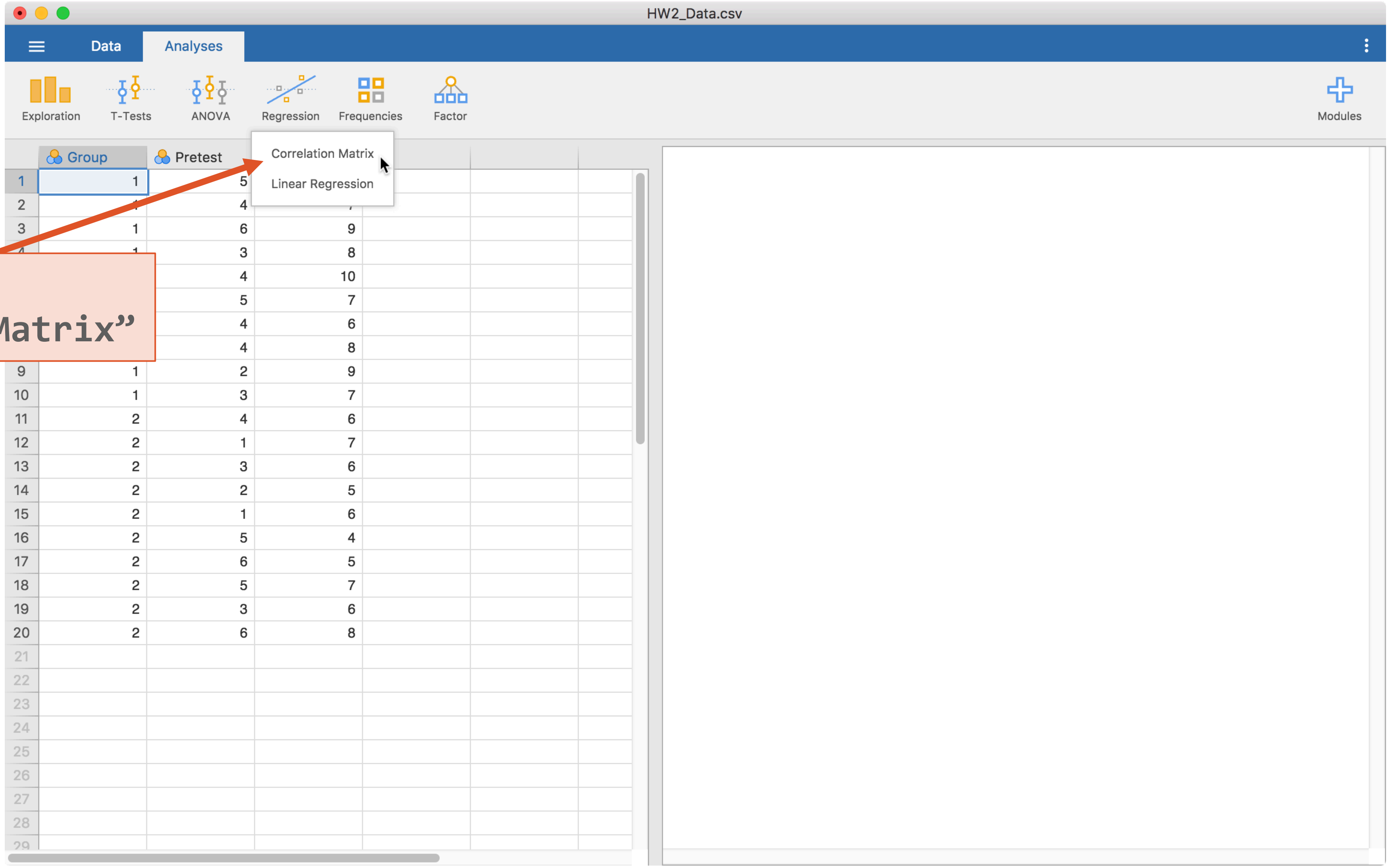
Hypothesis Type	Symbolic	Verbal	Difference between means created by:
Research Hypothesis	$\rho \neq 0$	There is a relationship between the variables	True relationship
Null Hypothesis	$\rho = 0$	There is no *real* relationship between the variables.	Random chance (sampling error)

3 Define Critical Regions

How much evidence is enough to believe the null is not true?

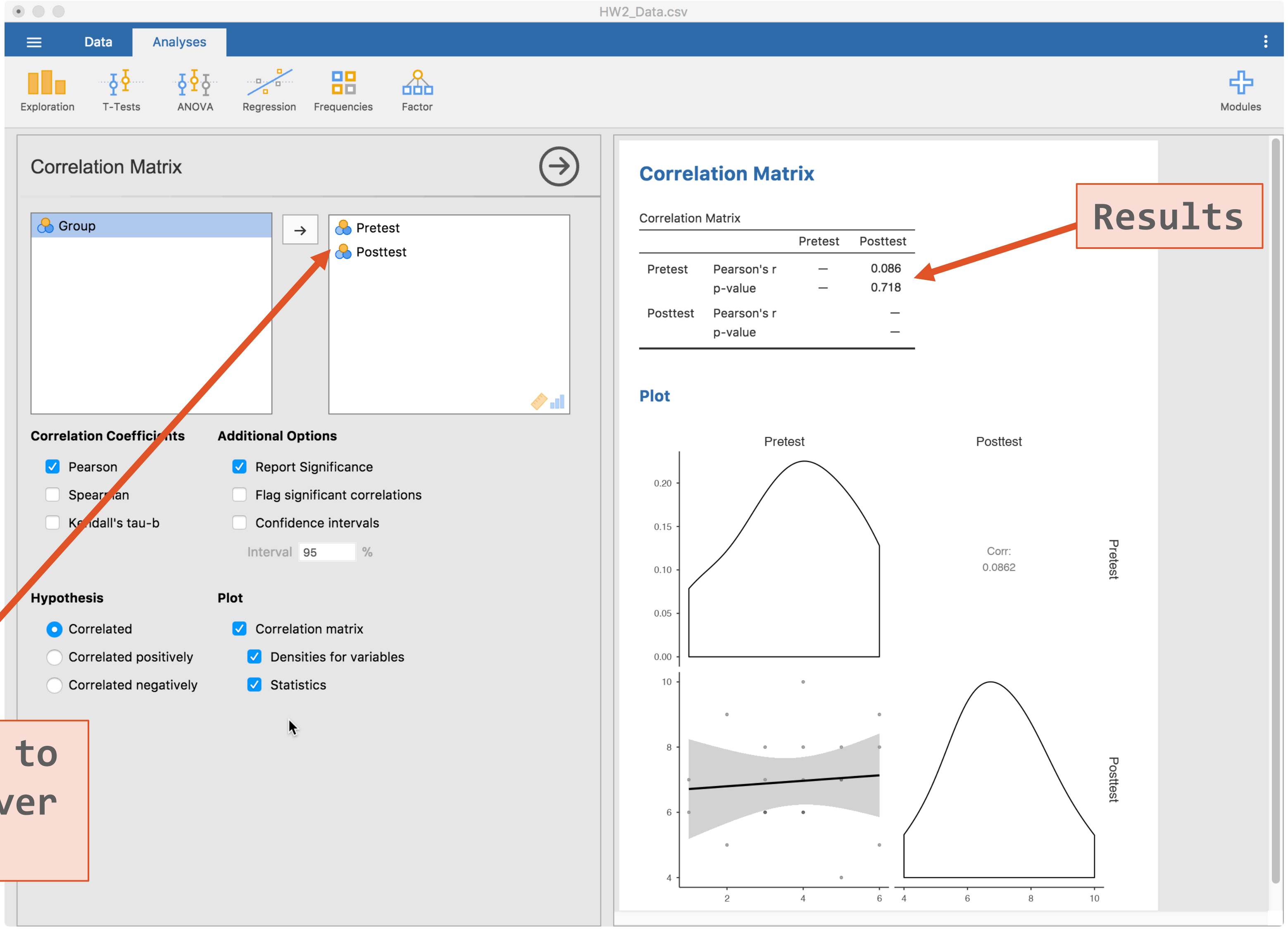
generally based on an alpha = .05

4 Compute the Test Statistic



Click on
“Correlation Matrix”

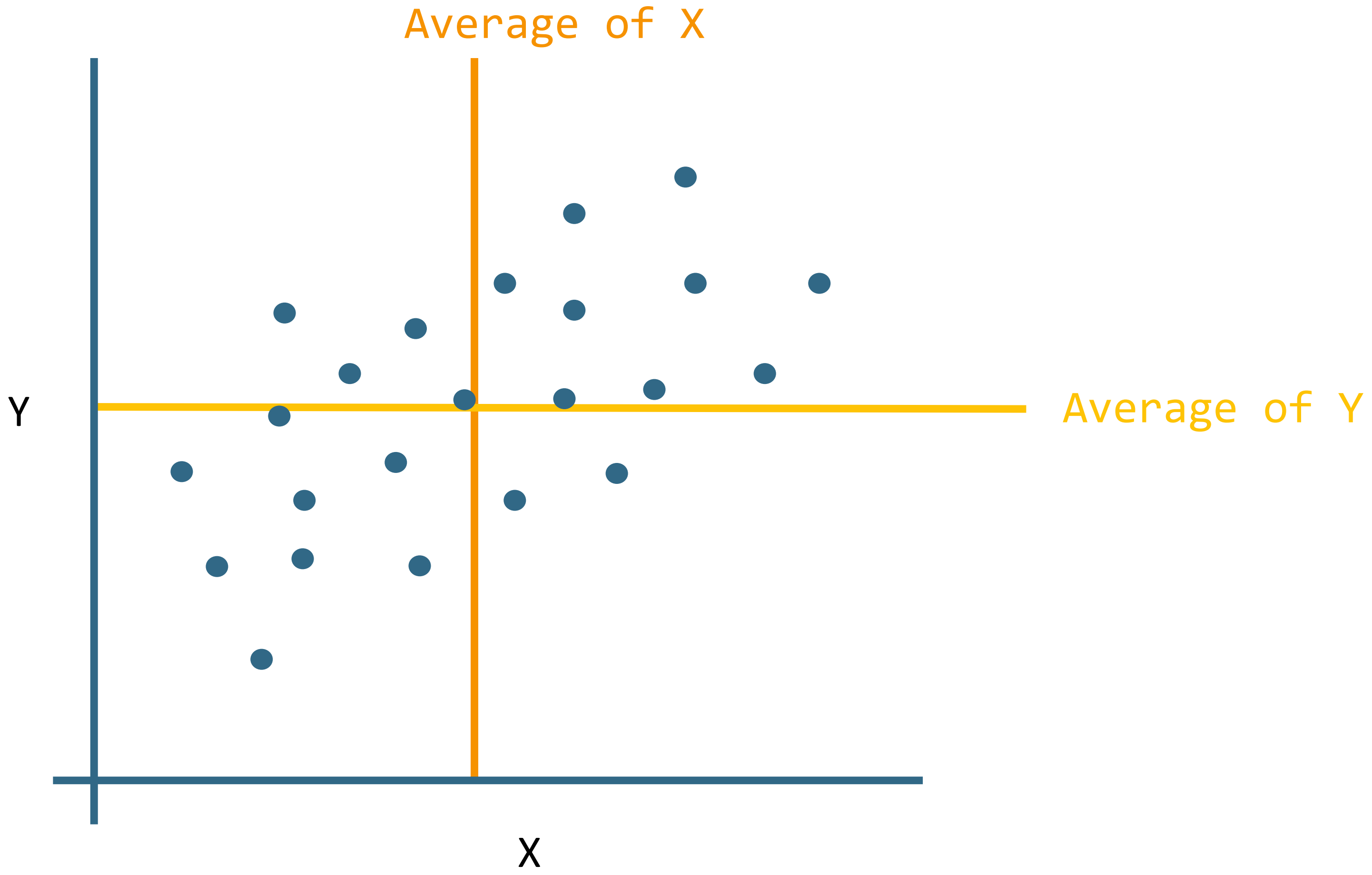
4 Compute the Test Statistic



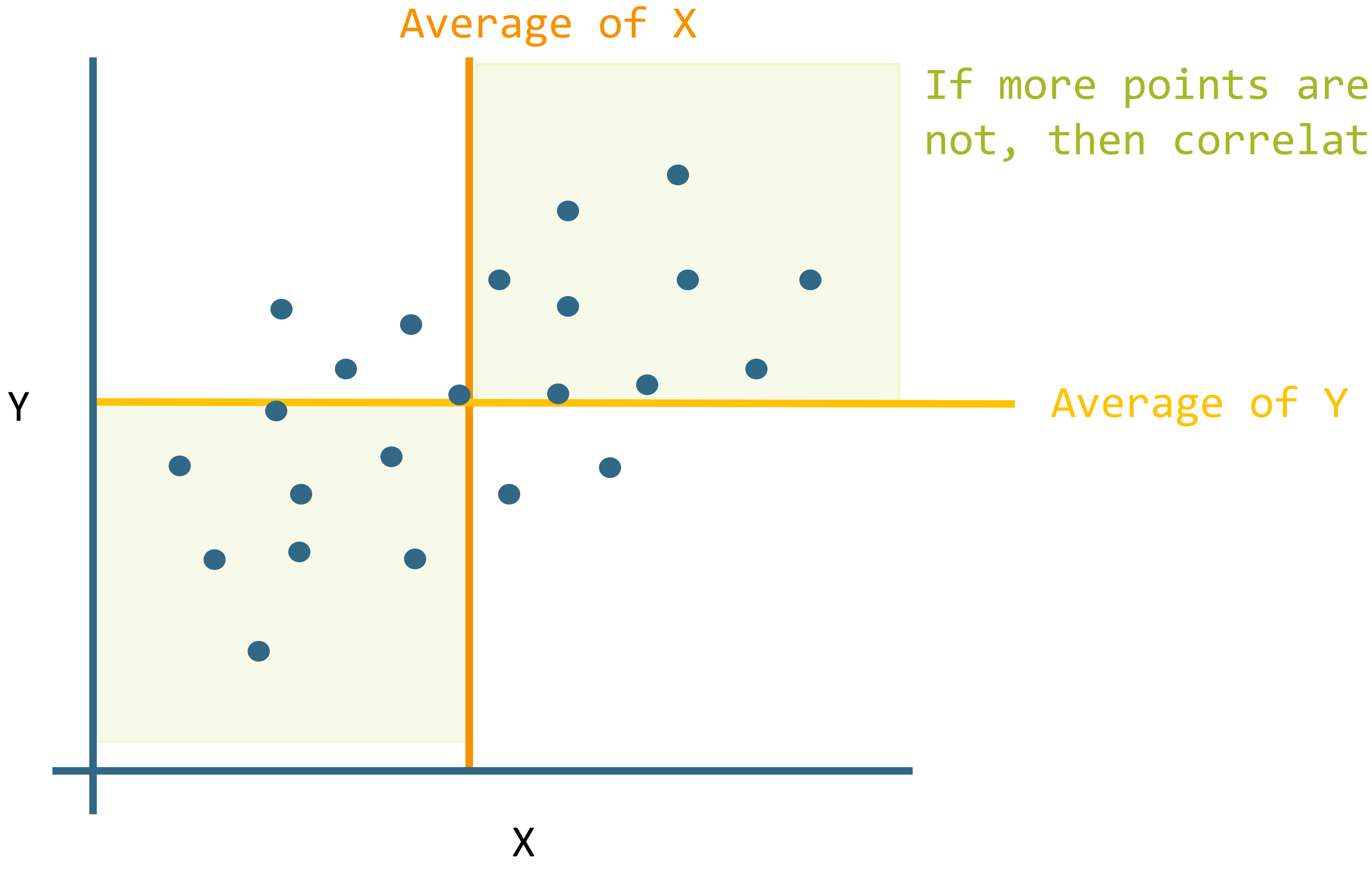
Results

Bring variables to be correlated over here

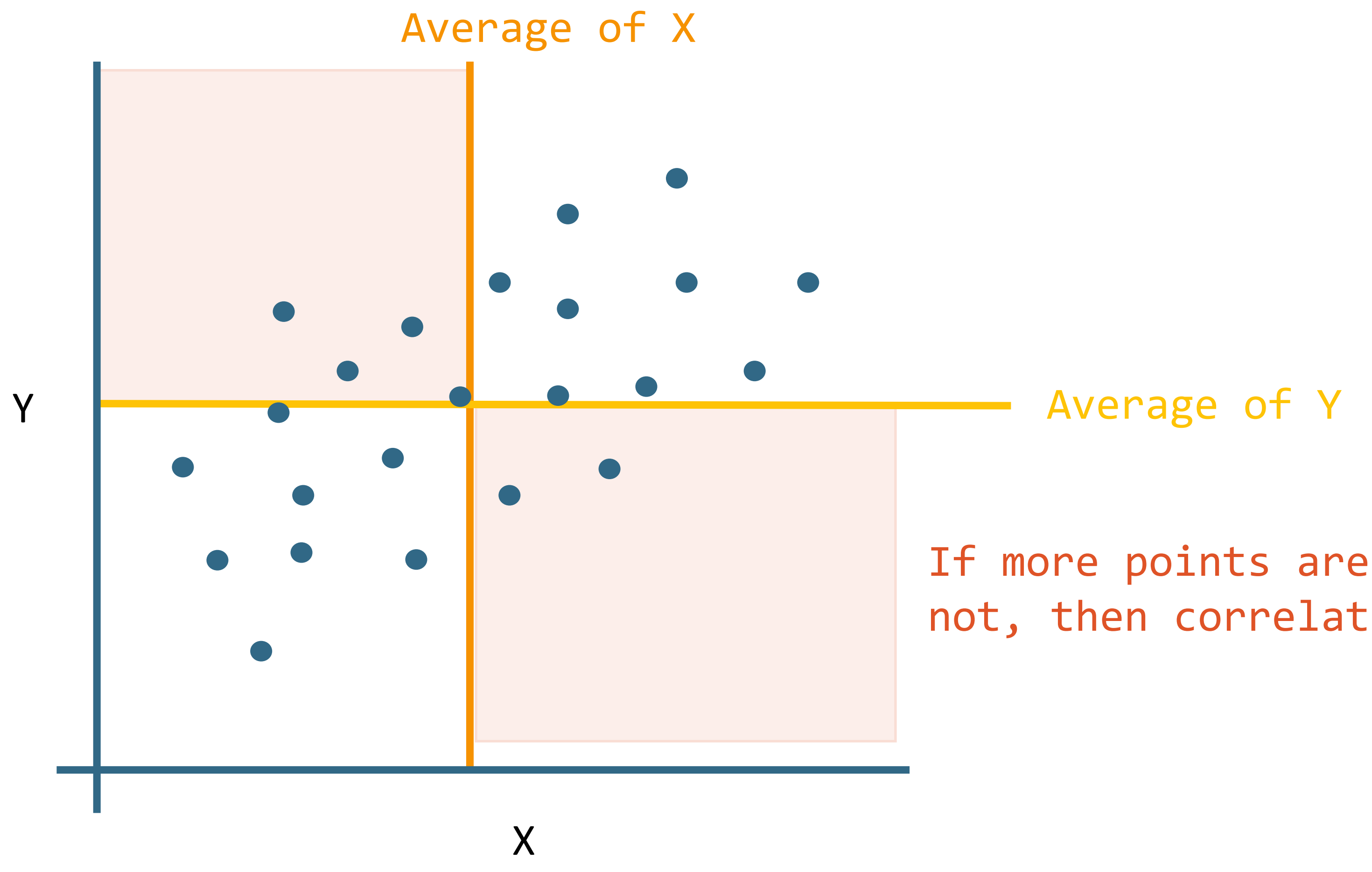
4 Compute the Test Statistic



4 Compute the Test Statistic



4 Compute the Test Statistic



If more points are in the red than not, then correlation is negative

5 Compute an Effect Size and Describe it

One of the main effect sizes for correlation is r^2

$$r^2 = (r)^2$$

r^2	Estimated Size of the Effect
Close to .01	Small
Close to .09	Moderate
Close to .25	Large

6 Interpreting the results

Put your results into words

Use the example around page 529 as a template

Let’s use Jamovi

Correlation

Intro to Regression

Intro to Regression

The foundation of almost everything we do in statistics

Comparing group means

Assess relationships

Compare means AND assess relationships at the same time

Can handle many types of outcome and predictor data types
Results are interpretable

Two Main Types of Regression

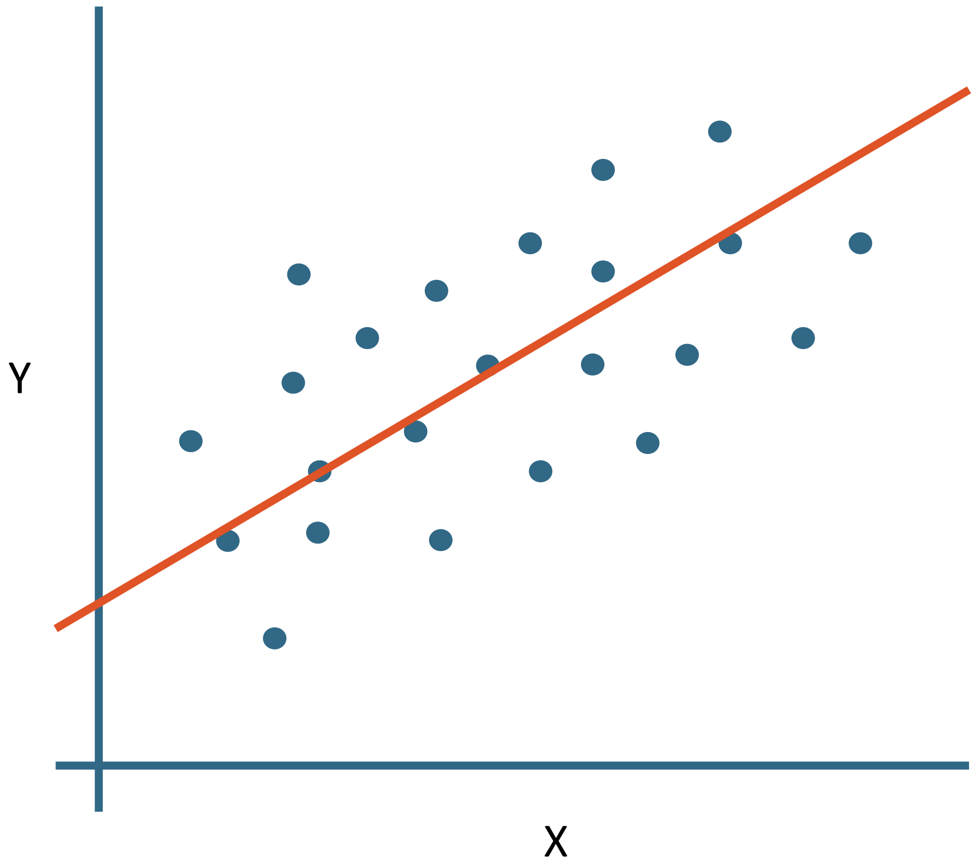
Simple

- Only one predictor in the model
- When variables are standardized, gives same results as correlation
- When using a grouping variable, same results as t-test or ANOVA

Multiple

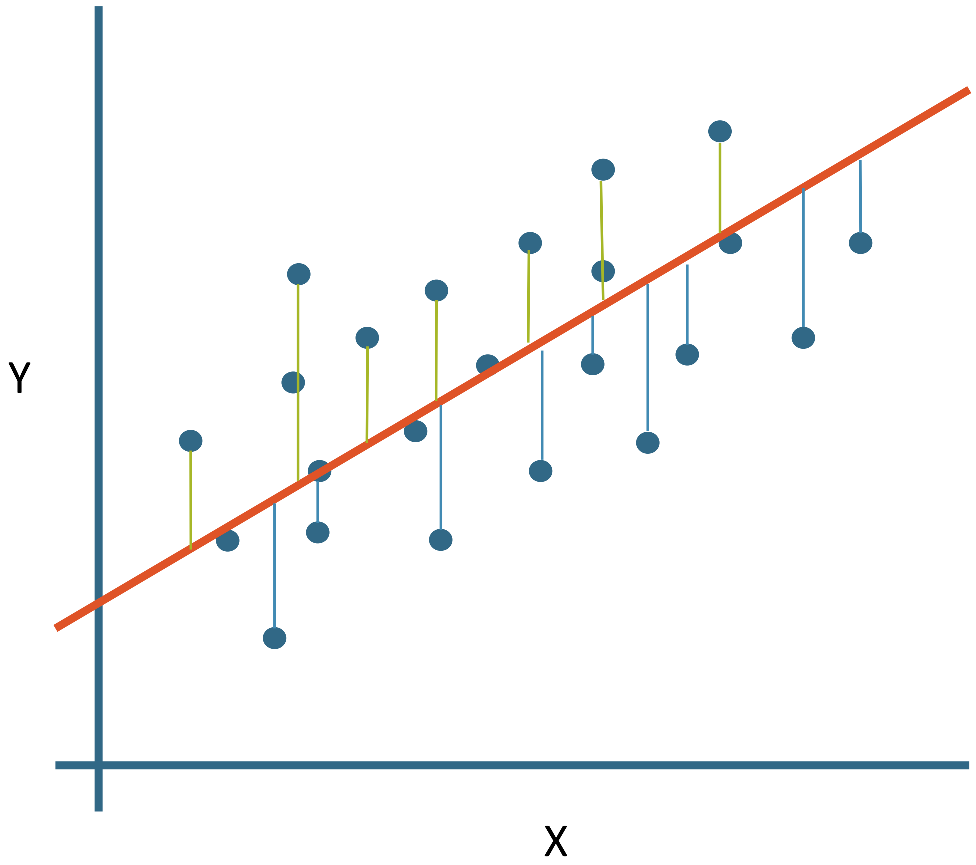
- More than one variable in the model
- When variables are standardized, gives "partial" correlation
- Predictors can be any combination of categorical and continuous

Logic of Regression



We are trying to find the best fitting line

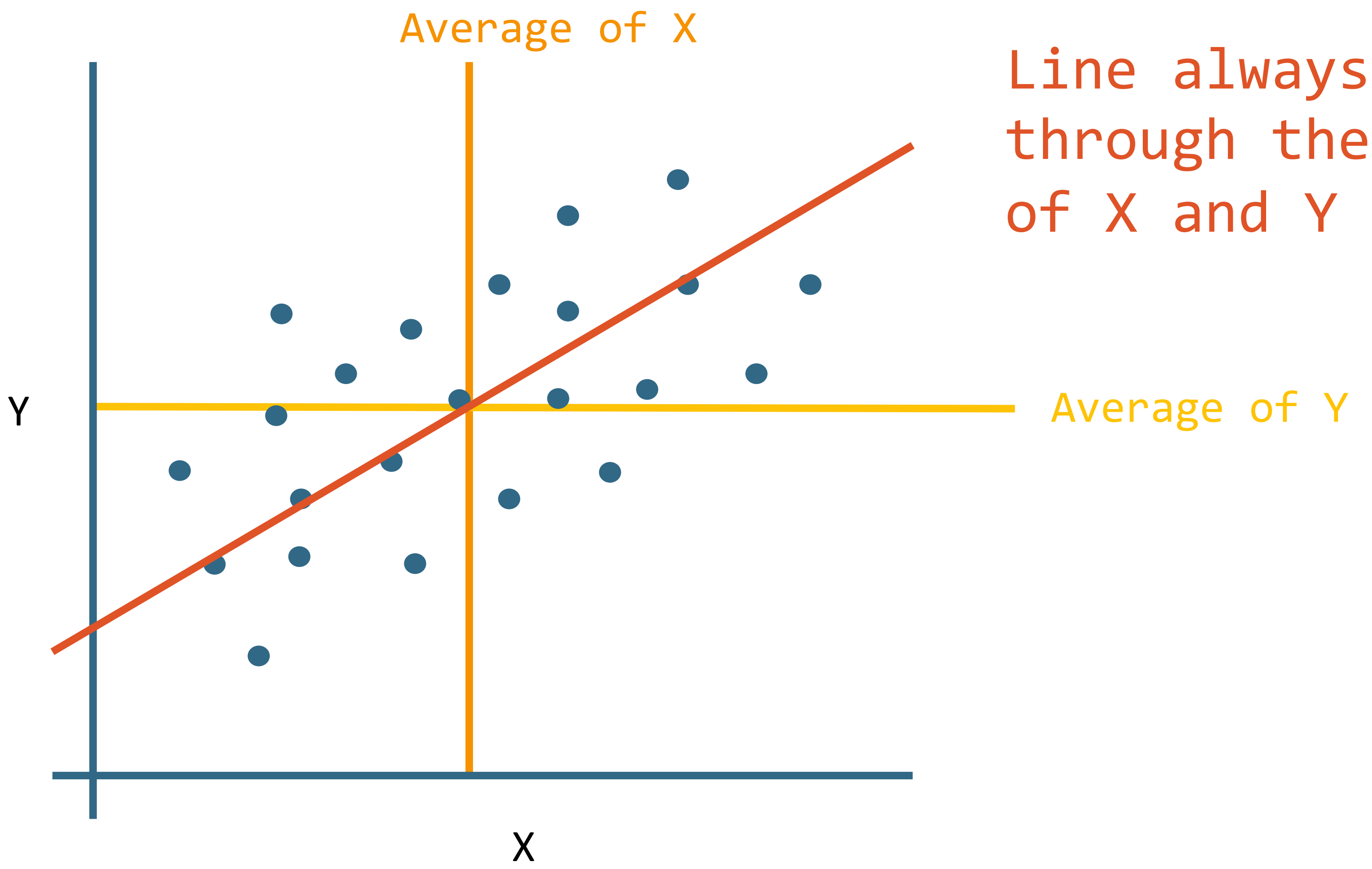
Logic of Regression



We are trying to find the best fitting line

We do this by minimizing the difference between the points and the line (called the residuals)

Logic of Regression



Line always goes through the averages of X and Y

Questions?

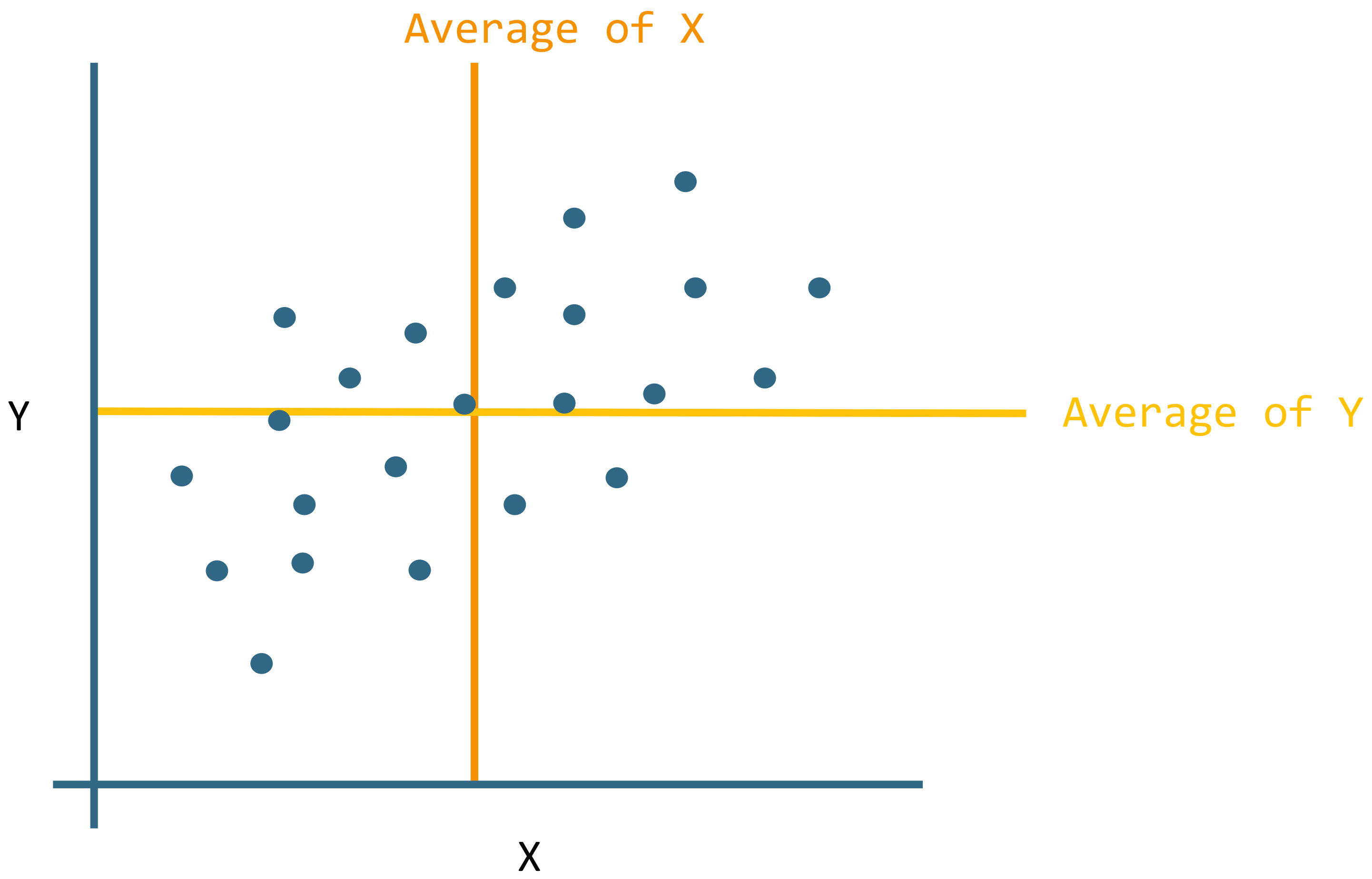
Please post them to the discussion board before class starts

End of Pre-Recorded Lecture Slides

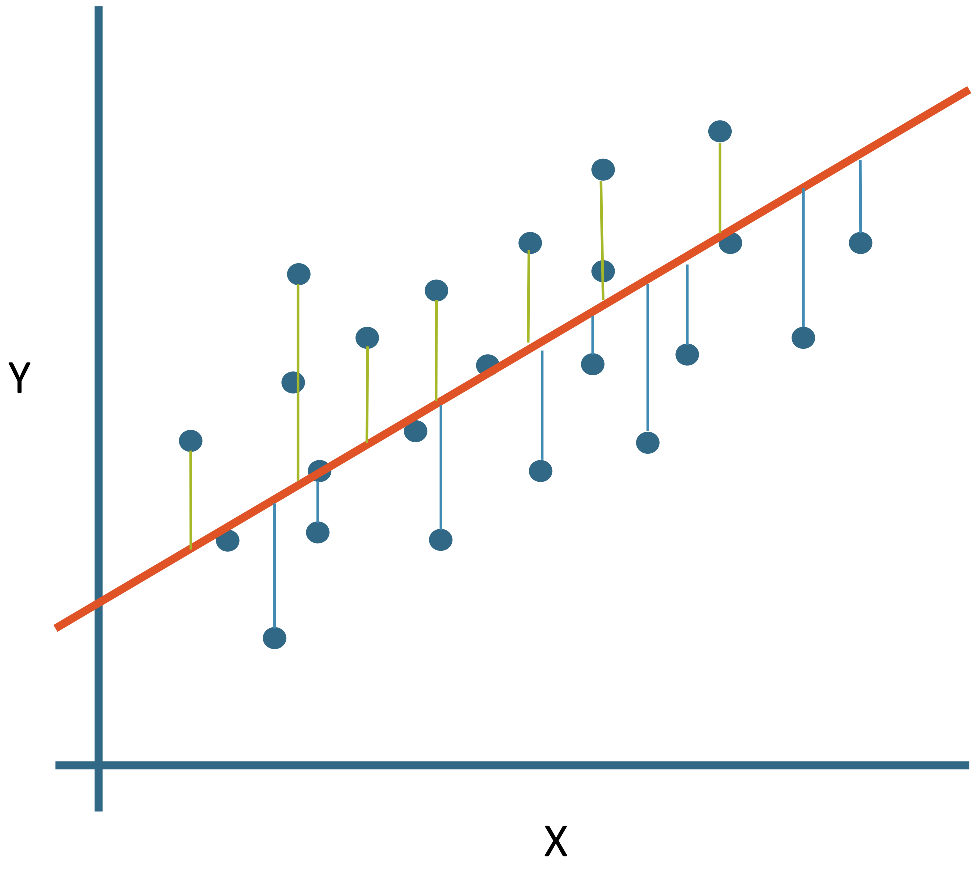
In-class discussion slides

https://www.youtube.com/watch?v=sxYrzzy3cq8

How Correlation Works



How Regression Works



We are trying to find the best fitting line

We do this by minimizing the difference between the points and the line (called the residuals)

Application

Example Using
The Office/Parks and Rec Data Set

Hypothesis Test with Correlation