Main Ideas

- Learn about Bivariate data
- Learn how to interpret data from a scatterplot
- Learn about the difference between Linear and Non-linear relationships.
- Explain the different types of correlation
- Learn how to calculate the correlation coefficient, r
Section 5.1 Scatterplots and Correlation

Bivariate data is data in which two variables are recorded or measured on an entity.

The explanatory (independent) variable, \( x \), explains the variation in the data

- \( x \)-axis, horizontal axis

The response (dependent) variable, \( y \), changes as a response to changes in the explanatory variable

- \( y \)-axis, vertical axis

A scatterplot shows all of the points of data with the explanatory variable along the \( x \)-axis and the response variable along the \( y \)-axis

Each point in a scatterplot: \((x, y)\)

Linear vs Nonlinear

- positive linear association
- negative linear association
- nonlinear association
- no association
Linear Vs. Non-Linear

Linear

The scatterplot seems to increase or decrease at a constant rate.

Non-linear

The scatterplot seems to increase and then decrease or vice versa, making what looks like a curve.

2/24/2020
Positive vs Negative (Inverse)

Positive

Perfect Positive Correlation

$r = 1$

High Positive Correlation

$r = 0.9$

Low Positive Correlation

$r = 0.5$

Negative (Inverse)

As $x$ increases, so does $y$

$r = -1$

$r = -0.9$

$r = -0.5$

No Correlation

$r = 0$

Correlation Coefficient, $r$: Measures the degree of a linear relationship; i.e., how well the data clusters around a line. $r$ is always between $-1$ and $1$. The correlation is stronger closer to $-1$ and $1$, and the relationship is weaker closer to $0$. 

Notes by Caxton Petri page 4
If the scatterplot indicates a nonlinear relationship, you cannot use \( r \) to measure the strength of the relationship.

Finding the correlation coefficient in Excel: \( =\text{CORREL(array1, array2)} \)

Note: The arrays need to be the same size since the data is paired

Facts about \( r \):
- Between -1 and +1.
- If \( r > 0 \), the relationship is positive (the line has a positive slope).
- If \( r < 0 \), the relationship is negative (the line has a negative slope).
- A value of \( r \) near -1 to +1 means the data is tightly bundled around the line.
- A value of \( r \) near zero means there is no linear relationship between \( x \) and \( y \).
- \( r \) depends on the mean of the explanatory and response variables. Therefore, it is very sensitive to outliers in the data.

Open Past Course Data In Excel:
Find the correlation coefficient and create a scatterplot for the following sets of data.
1. Homework Average vs Course Grade
2. Attendance vs Course Grade
3. No. Absences vs Course Grade