

2.5: Two Quantitative Variables: Scatterplots and Correlation

Dr. John Peter

An Example

Ex: (Election Margin Data from chart at beginning of Section 2.5)

An Example

Ex: (Election Margin Data from chart at beginning of Section 2.5)

Questions:

- ▶ What was the highest approval rating among the losing presidents?

An Example

Ex: (Election Margin Data from chart at beginning of Section 2.5)

Questions:

- ▶ What was the highest approval rating among the losing presidents?
- ▶ What was the lowest approval rating among the winning presidents?

An Example

Ex: (Election Margin Data from chart at beginning of Section 2.5)

Questions:

- ▶ What was the highest approval rating among the losing presidents?
- ▶ What was the lowest approval rating among the winning presidents?
- ▶ Make a conjecture about the approval rating needed to win an election.

Scatterplots

A **scatterplot** is a graph of the relationship between two quantitative variables.

Scatterplots

A **scatterplot** is a graph of the relationship between two quantitative variables.

- ▶ One axis for each variable

Scatterplots

A **scatterplot** is a graph of the relationship between two quantitative variables.

- ▶ One axis for each variable
- ▶ One point for each data pair

Scatterplots

A **scatterplot** is a graph of the relationship between two quantitative variables.

- ▶ One axis for each variable
- ▶ One point for each data pair
- ▶ Explanatory variable on horizontal axis (if possible).

Scatterplots

A **scatterplot** is a graph of the relationship between two quantitative variables.

- ▶ One axis for each variable
- ▶ One point for each data pair
- ▶ Explanatory variable on horizontal axis (if possible).

Look at scatterplot of Election Margin Data.

Questions We Ask About Scatterplots

- ▶ Is there an obvious pattern/trend?

Questions We Ask About Scatterplots

- ▶ Is there an obvious pattern/trend?
- ▶ Positive or Negative? Other pattern? Linear?

Questions We Ask About Scatterplots

- ▶ Is there an obvious pattern/trend?
- ▶ Positive or Negative? Other pattern? Linear?
- ▶ Are there any outliers?

Questions We Ask About Scatterplots

- ▶ Is there an obvious pattern/trend?
- ▶ Positive or Negative? Other pattern? Linear?
- ▶ Are there any outliers?

The Election Margin example data seem to have a strong, positive linear trend.

Correlation

Correlation (denoted by r or ρ) is a measure of the strength and direction of a linear association between two quantitative variables.

Correlation

Correlation (denoted by r or ρ) is a measure of the strength and direction of a linear association between two quantitative variables.

► $-1 \leq r \leq 1$

Correlation

Correlation (denoted by r or ρ) is a measure of the strength and direction of a linear association between two quantitative variables.

- ▶ $-1 \leq r \leq 1$
- ▶ r only measures *linear* relationships.

Correlation

Correlation (denoted by r or ρ) is a measure of the strength and direction of a linear association between two quantitative variables.

- ▶ $-1 \leq r \leq 1$
- ▶ r only measures *linear* relationships.
- ▶ A strong positive or negative correlation (close to 1 or -1 , respectively) does not imply a causal relationship! (Example 2.36 from text).

Correlation

Correlation (denoted by r or ρ) is a measure of the strength and direction of a linear association between two quantitative variables.

- ▶ $-1 \leq r \leq 1$
- ▶ r only measures *linear* relationships.
- ▶ A strong positive or negative correlation (close to 1 or -1 , respectively) does not imply a causal relationship! (Example 2.36 from text).
- ▶ A correlation near 0 does not mean the two variables are not associated!

Correlation

Correlation (denoted by r or ρ) is a measure of the strength and direction of a linear association between two quantitative variables.

- ▶ $-1 \leq r \leq 1$
- ▶ r only measures *linear* relationships.
- ▶ A strong positive or negative correlation (close to 1 or -1 , respectively) does not imply a causal relationship! (Example 2.36 from text).
- ▶ A correlation near 0 does not mean the two variables are not associated! This does mean that there is no linear relationship present. (Example 2.37 from text)

Correlation

Correlation (denoted by r or ρ) is a measure of the strength and direction of a linear association between two quantitative variables.

- ▶ $-1 \leq r \leq 1$
- ▶ r only measures *linear* relationships.
- ▶ A strong positive or negative correlation (close to 1 or -1 , respectively) does not imply a causal relationship! (Example 2.36 from text).
- ▶ A correlation near 0 does not mean the two variables are not associated! This does mean that there is no linear relationship present. (Example 2.37 from text)
- ▶ Correlation can be heavily influenced by outliers. (Example 2.38 from text)

The Formula for r

$$r = \frac{1}{n-1} \sum \left(\frac{x - \bar{x}}{s_x} \right) \left(\frac{y - \bar{y}}{s_y} \right)$$

The Formula for r

$$r = \frac{1}{n-1} \sum \left(\frac{x - \bar{x}}{s_x} \right) \left(\frac{y - \bar{y}}{s_y} \right)$$

Anything familiar?