Chapter 7 Lecture Notes

Scatterplots, Association and Correlation

We will now focus on the relationship between two variables. We distinguish two types of variables:

Response Variable

Explanatory Variable

Examples:

Example:
A real estate agent is interested in relating the price of a home to the number of bedrooms. In this situation, which variable would be considered the response variable?

Response Variable:

Explanatory Variable:

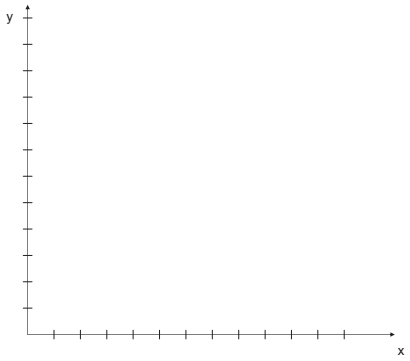
Graphing two quantitative variables:
When both the explanatory and response variables are quantitative, we can display them using a **scatterplot**.

Explanatory variable:

Response Variable:

EXAMPLE: height (in.) and weight (lbs.) of 5 Stat 101 students

Height (x)	63	64	70	72	78
Weight (y)	115	130	150	170	190

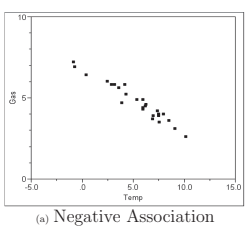


Describing Scatterplots: The four things to mention when describing a scatterplot.

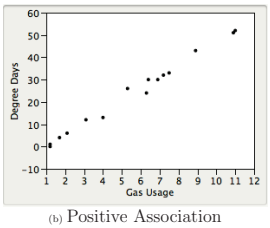
(1.) Direction

Positive Association

Negative Association

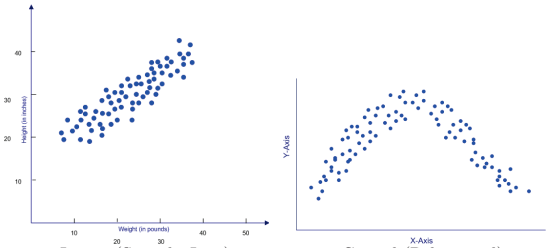


(a) Negative Association



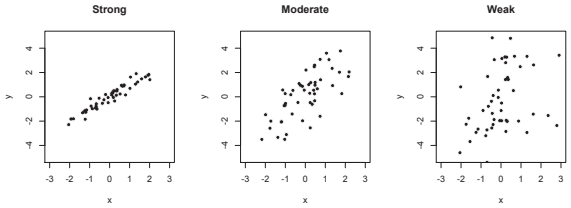
(b) Positive Association

(2.) Form



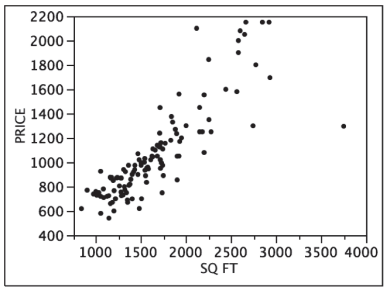
(c) Linear (Straight Line) (d) Curved (Polynomial)

• Strength



• Outliers

Example Describing a Scatterplot:
House Price ($) and Size of house (sq. ft.)



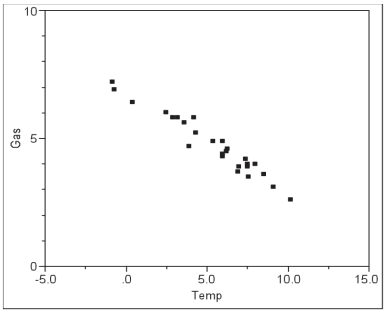
Form:

Strength:

Direction:

Outliers:

Example Describing a Scatterplot:
Outside Temp (C) and Natural Gas Used (1000ft3)



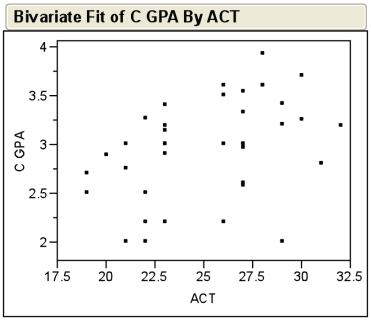
Form:

Strength:

Direction:

Outliers:

Example Describing a Scatterplot:



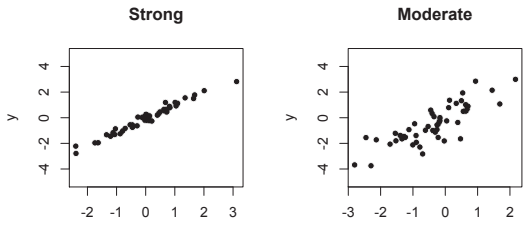
Form:

Strength:

Direction:

Outliers:

Correlation



How can we quantify the strength of a linear relationship?

Correlation:

Properties

- denoted by r
- The sign of r indicates the directions of the association.

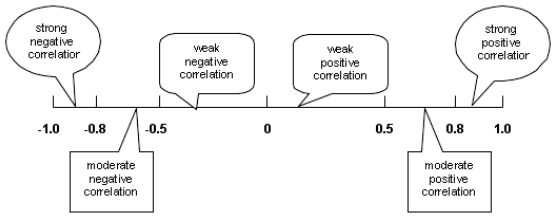
  $r > 0$ _____

  $r < 0$ _____
- r ranges in value from -1 to 1

  _____ indicates a straight, increasing line

  _____ indicates a straight, decreasing line

  _____ indicates no **linear** relationship



How to Calculate r

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

Where

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} \qquad s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}}$$

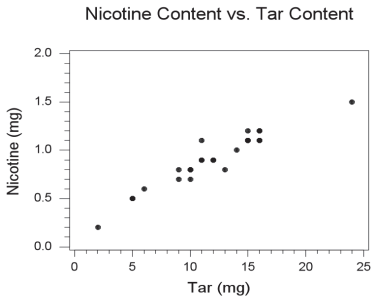
EXAMPLE: Calculating r with height (in.) and weight (lbs.) of $n = 5$ Stat 101 students.

Height (x)	63	64	70	72	78
Weight (y)	115	130	150	170	190

Some other useful information:

$\Sigma(x - \bar{x})(y - \bar{y}) = 728 \qquad s_x = 6.148 \qquad s_y = 30.08$

Example Describing a Scatterplot:



r=0.956

Lurking Variables

- Association does not imply causation.
- Variables other than the two in the scatterplot can affect the relationship between the two variables.
- There is a strong positive correlation between the number of crimes committed in communities and the number of 2^{nd} graders in those communities.
- Beware of lurking variables.