

Chapter 3: Regression Potpourri

Skills

- Understand the dangers of extrapolation
 - Identify outliers and influential points
 - Interpret the slope and intercept of the Least Squares Regression Line
-
-
-
-
-
-
-
-
-
-

Extrapolation

The regression line is only valid for prediction within that range of observed x values

Using an x value outside of the original range is called **extrapolation**

Extrapolation assumes that the observed pattern continues

This assumption is often wrong

So: don't extrapolate!

Example

A regression was performed for resting heart rate versus age

The observed ages were 20 – 60

Should this regression model be used for a person who is 75 years old?

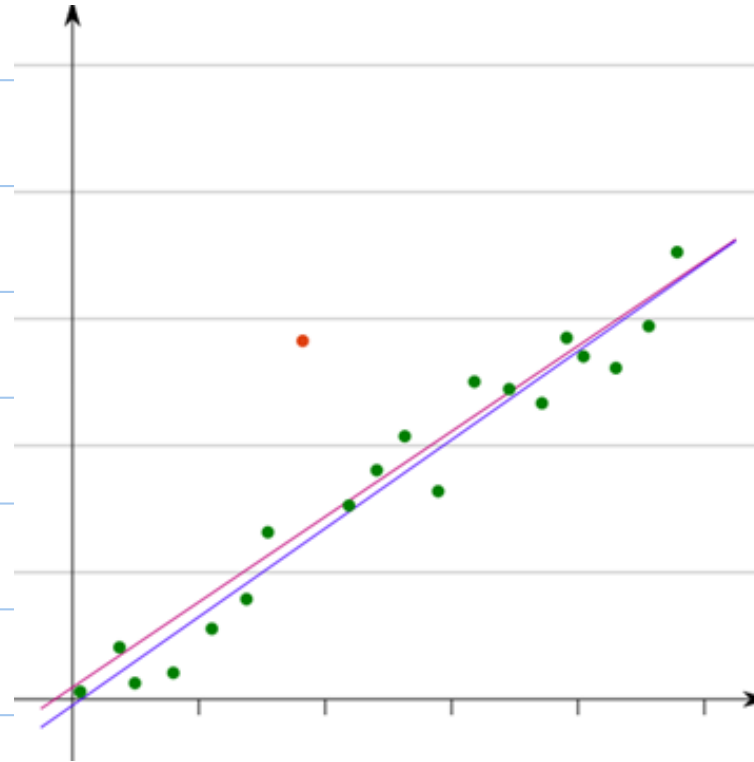
Outliers

For a single variable, Tukey's Rule helped to identify outliers

There is no such rule for two variables

Just look—are there points that don't follow the pattern?

Outlier Example



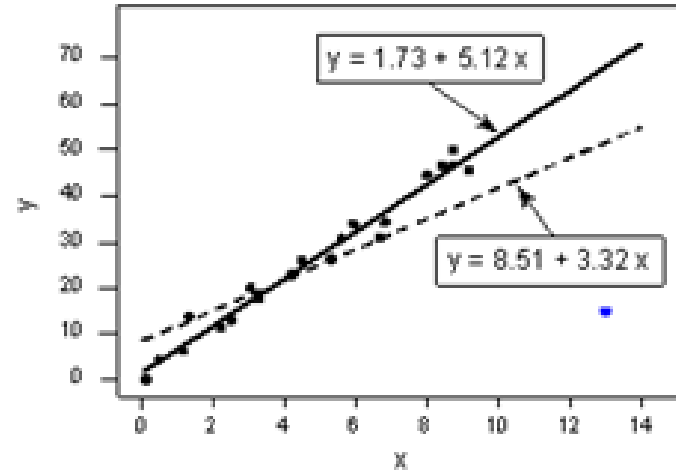
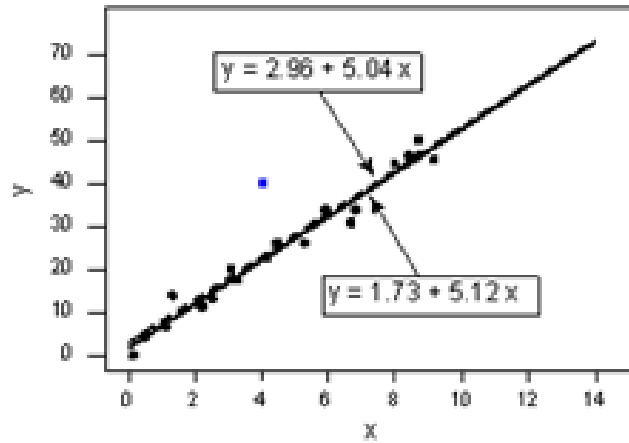
Influential Points

Some points exert more influence on the position/
slope of the regression line than others

If removing a point makes a big difference in the
position/slope of the line, then the point is
influential

Points far removed along the x axis tend to be
influential

Example



Example

A study investigated the relationship between the amount of fiber (grams) and potassium content (milligrams) in breakfast cereal...

$$\widehat{\text{Potassium}} = 38 + 27 \cdot (\text{Fiber})$$

Interpreting the Intercept

The intercept is the average y -value predicted by the model when the x -value is zero.

Example

A study of the relationship between fuel economy and horsepower in cars resulted in the following model:

$$\widehat{\text{mpg}} = 46.87 - 0.084 \cdot (\text{HP})$$
