Stat 101 – Lecture 15

Getting the “Bends”

- A fundamental assumption is that the relationship is a straight line.
- What looks straight on a scatter plot may show a curve when one looks at the plot of residuals versus the explanatory variable.

1

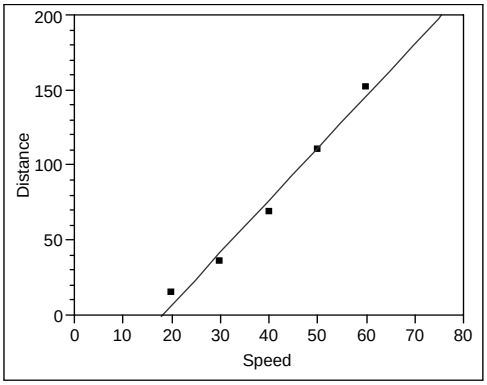
Example

- *y*, Stopping distance (feet)
- *x*, Speed (miles per hour)

$$R^2 = 0.984$$

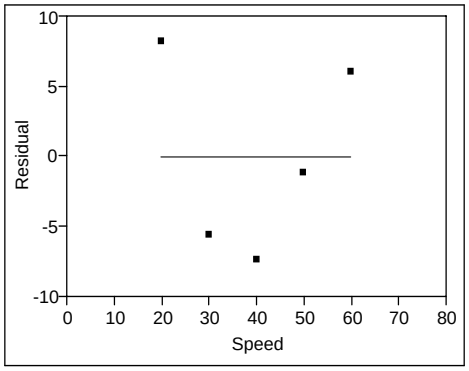
$$\hat{y} = -62.8 + 3.48\,x$$

2



3

Stat 101 – Lecture 15



4

Interpretation

- There is a curved pattern in the residuals.
  - under predicts at 20 mph
  - over predicts at 30, 40 and 50 mph
  - under predicts at 60 mph

5

Interpretation

- Although the straight line does a very good job explaining the variation in stopping distance, a curved relationship model would do even better.

6

Stat 101 – Lecture 15

Regression Wisdom

- Sifting Residuals for Groups
  - Display residuals versus the explanatory variable.
  - Look at the distribution of residuals.

7

Example

- *y*, Life Expectancy (years)
- *x*, Wealth Index

$$r = 0.874$$

$$\hat{y} = 2.41 + 7.71\,x$$

8

9

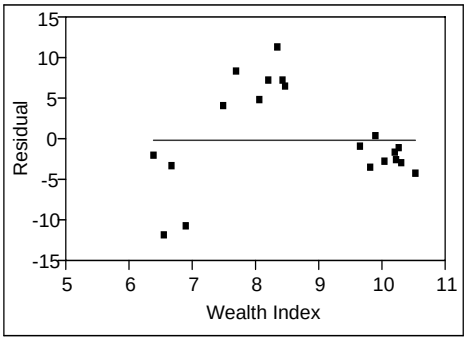
Stat 101 – Lecture 15

Interpretation

- There appear to be several groups of residuals.
  - two large negative residuals.
  - a large group between –5 and 0.
  - another group between 5 and 10.

10

11



12

Stat 101 – Lecture 15

Dangers of Extrapolation

- Suppose we use the least squares equation relating speed to stopping distance for a vehicle traveling at 5 mph?
- The predicted stopping distance is – 45.4 feet.

Special Points

- Outlier – In regression, this is a point with a large residual.
- Leverage – In regression a point has high leverage if it is an extreme value for the explanatory variable.

Influence

- Outliers and high leverage points can greatly influence what the intercept and the slope of the least squares line will be.

http://netfiles.uiuc.edu/jimarden/www/cuwu/datalist.html