Regression Wisdom

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

Example

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

\[ r = 0.874 \]
\[ \hat{y} = 2.41 + 7.71 \, x \]
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.
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.

Example

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

\[
R^2 = 0.984 \\
\hat{y} = -62.8 + 3.48x
\]
**Interpretation**

- There is a curved pattern in the residuals.
  - under predicts
  - over predicts
  - under predicts

**Interpretation**

- Although the straight line does a very good job explaining the variation in stopping distance, a curved relationship model would do even better.
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 be influential points, that is, they can greatly influence what the intercept and the slope of the least squares line will be.