Chapter 3: Regression Wisdom

Skills

• Understand the dangers of extrapolation
• Identify outliers and influential points
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.
Examples
Summary

Plot the data!
Strength, direction, association, unusual features
If linear:
$r$ and $r^2$ as support
Residual plot – looking for no pattern
Histogram of residuals – looking for approximately normal
Use the regression model to make a prediction