Simple Linear Regression

**Simple linear regression**

- Measures the strength of an association between two variables, but cannot be used to demonstrate causation
- The adjective *simple* refers to the fact that this regression is one of the simplest in statistics.
- Used to calculate the equation for a straight line
- Can be used to fit a predictive model to an observed data set of $y$ and $X$ values
- Examines the relation between a normally distributed variable and 1 or more continuous predictor or independent variables

**Assumptions for Inference in Linear Regression**

- “$Y$”s are independent
- “$Y$”s are normally distributed for each value of $x$
- Normal distributions at different values of $x$ all have the same variance

**Prediction v. Relation**

Regression provides specific quantitative predictions that more precisely explains relations among variables
- Regression also allows the use of one piece of information to make predictions about something else

**Testing hypothesis.** We generally test a null hypothesis using the evidence (data) we have, but does not prove or disprove anything. This is the hallmark of scientific discovery in general not statistics per se.
For example: The problem of using biased sample is not the problem of statistics. It is the human error we bring in either due to a biased instrument for measuring a characteristic of interest or the study or sampling design we use.

**Best Fit**

All the values for y and x are considered in a scatterplot and the line that fits the best is the line assigned for the regression. The best fit is determined by the smallest total distance from each of the points on the scatterplot to the assigned line.