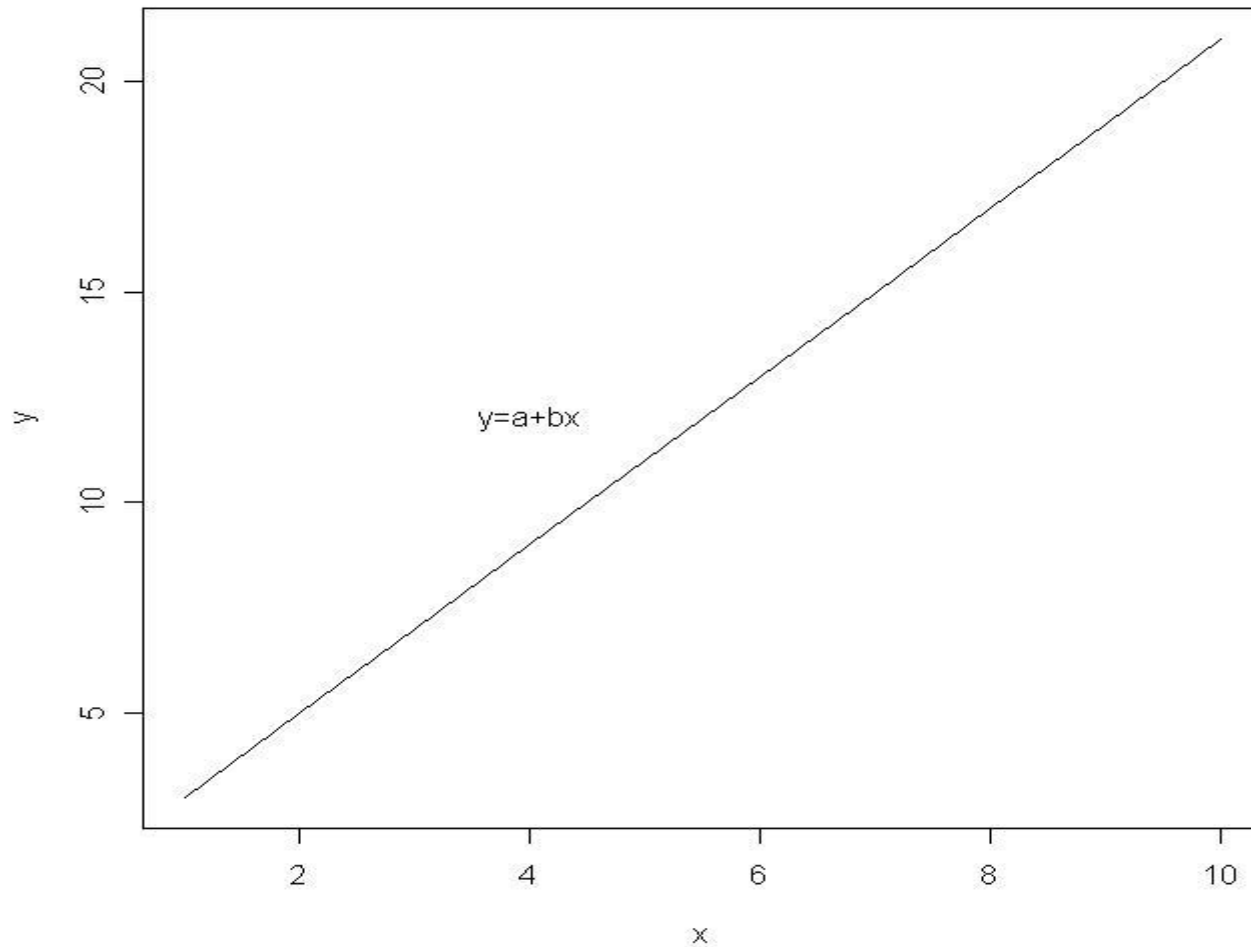


Regression

Will Doyle

Flashback Time: Defining a Line



Regression: Line Fitting

- Regression is a means of fitting a line to an observed set of data
- In two dimensional space, the line is fit according to $y=a+bx$
- In more dimensions, the same logic applies, but the line can not be drawn simply

Regression lines

- In regression we postulate a model of the world of the following linear (line) form:

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

Where:

y_i = the dependent variable for case i

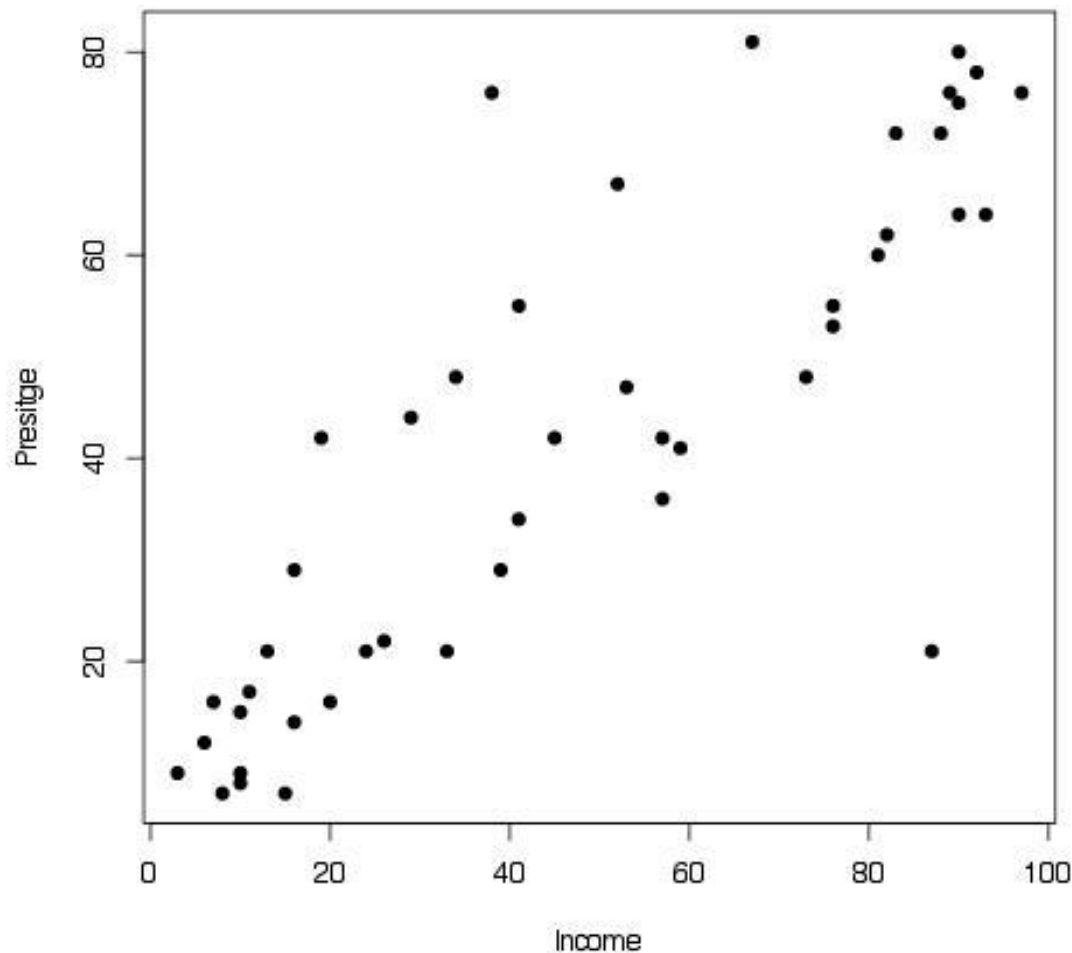
x_i = the independent variable for case i

α = the intercept of the line describing the relationship between x and y

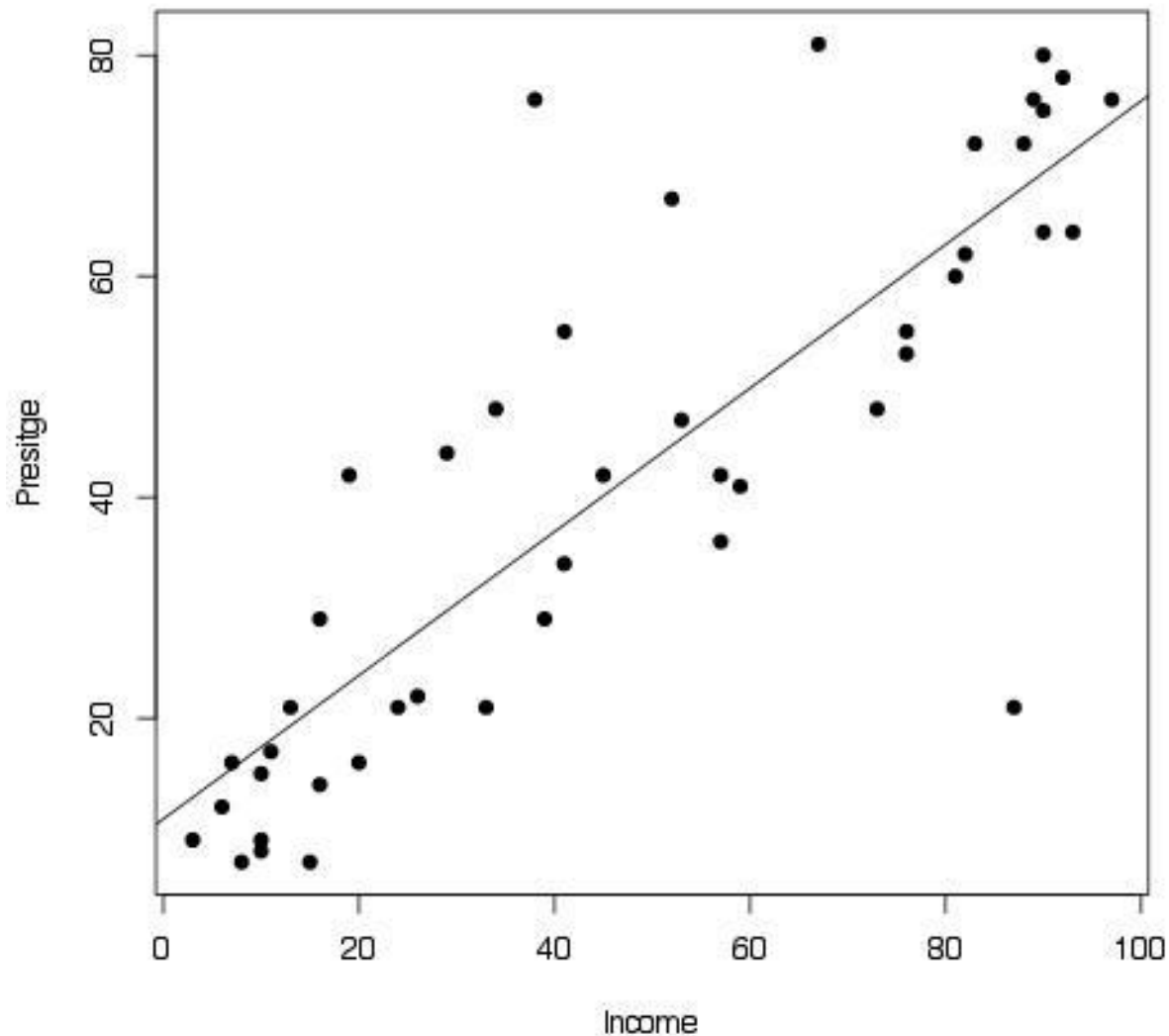
β = the slope of the line describing the relationship between x and y

ε_i = an error term showing how far the fitted line is from x and y

Duncan data on income and prestige



Duncan Data With Regression



Ordinary Least Squares

- Ordinary Least Squares (OLS) is the most common (and usually best) form of regression when the dependent variable is continuous
- OLS proceeds by finding a line that minimizes the sum of squared errors (sse) between the line and the points in the data
- OLS depends on a key set of assumptions

Assumptions under OLS

- Y is continuous
- There are fewer independent variables than there are cases in the data
- The errors are distributed normally, and have a mean of 0.
- Error terms are not correlated with one another, nor are they correlated with any of the regressors