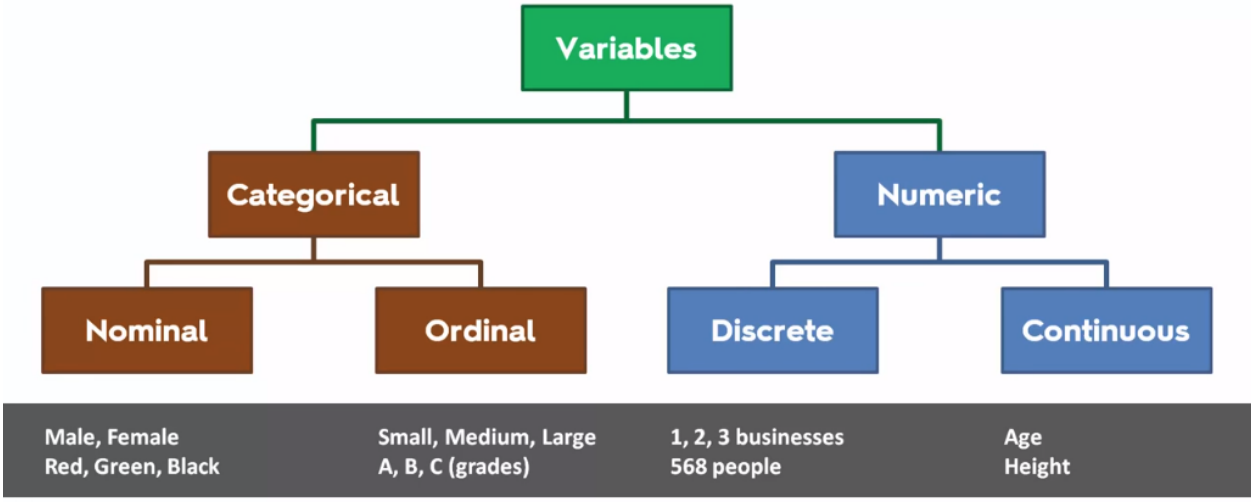
Types of Variables

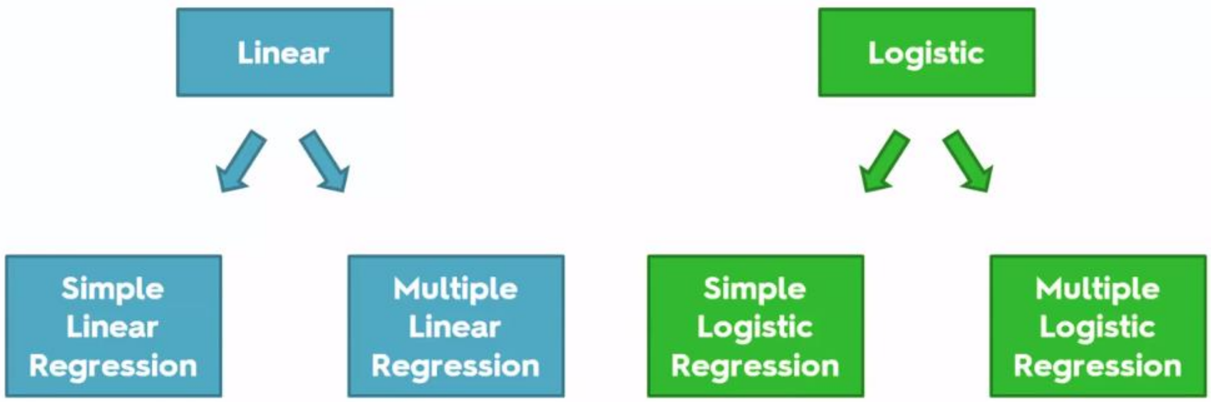


Regressions

In statistics, regression analysis is a statistical process for estimating the relationships among variables. ...

The focus is on the relationship between a dependent variable and one or more independent variables.

-Wikipedia



Simple Linear Regression

Constant Coefficient

$$y = b_0 + b_1{*}x_1$$

Dependent variable (DV) Independent variable (IV)

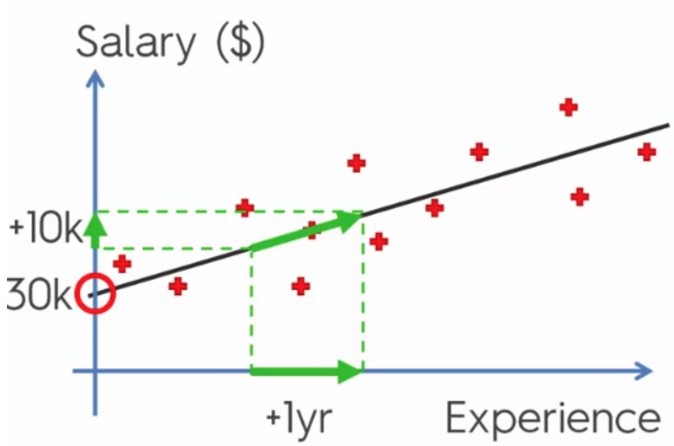
Multiple Linear Regression

Dependent variable (DV) Independent variables (IVs)

$$y = b_0 + b_1{*}x_1 + b_2{*}x_2 + \ldots + b_n{*}x_n$$

Constant

Simple Linear Regression:

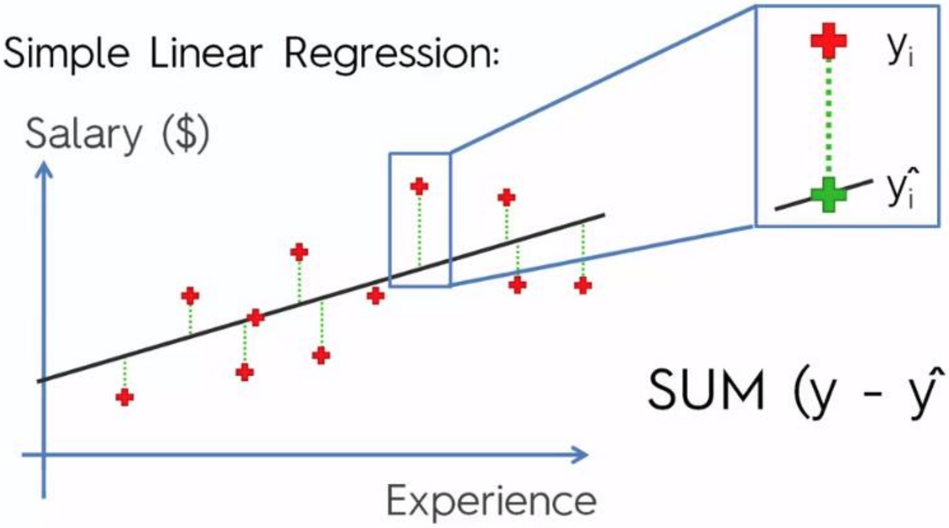


$$y = b_0 + b_1{*}x$$

$$\text{Salary} = b_0 + b_1{*}\text{Experience}$$

Ordinary Least Squares

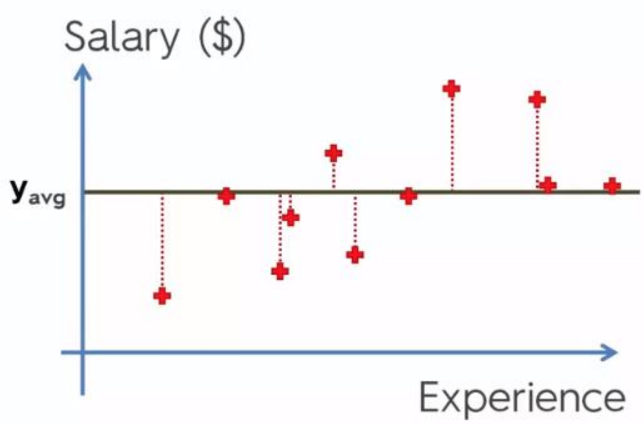
Simple Linear Regression:



$$\text{SUM } (y - \hat{y})^2 \rightarrow \min$$

R Squared

Simple Linear Regression:



$$SS_{res} = \text{SUM } (y_i - \hat{y_i})^2$$

$$SS_{tot} = \text{SUM } (y_i - y_{avg})^2$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Adjusted R^2

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

R^2 – Goodness of fit
(greater is better)

$y = b_0 + b_1 * x_1$

$y = b_0 + b_1 * x_1 + b_2 * x_2$

Problem:

$\boxed{+ b_3 * x_3}$

SS_{res} -> Min

R^2 will never decrease

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

$$\text{Adj } R^2 = 1 - (1 - R^2)\frac{n - 1}{n - p - 1}$$

p - number of regressors

n - sample size