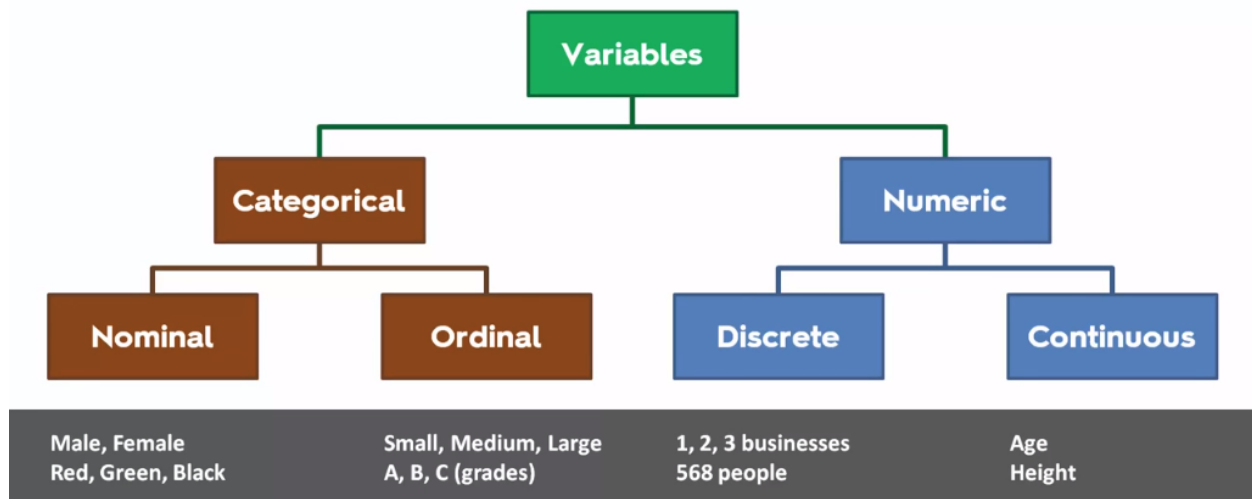


# Types of Variables

---



Data Science Training

© Kirill Ereminenko

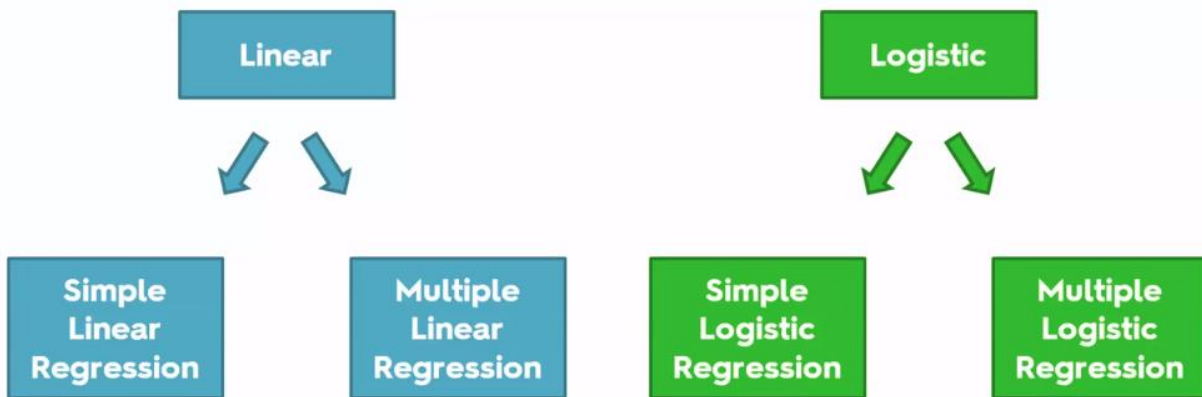
## Regressions

---

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**

$$y = b_0 + b_1 * x_1$$

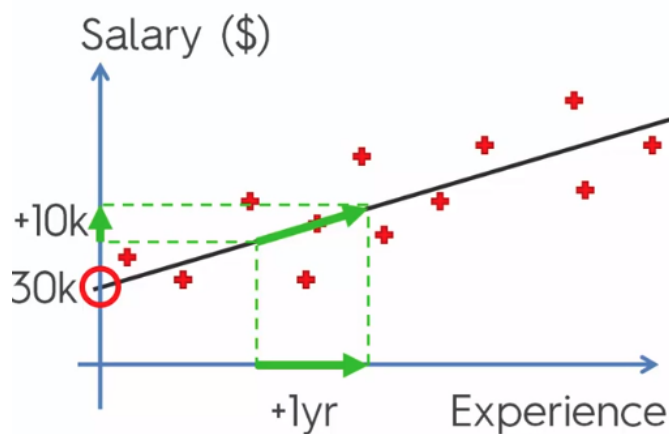
Constant (points to  $b_0$ )      Coefficient (points to  $b_1$ )  
 Dependent variable (DV) (points to  $y$ )      Independent variable (IV) (points to  $x_1$ )

**Multiple Linear Regression**

$$y = b_0 + b_1 * x_1 + b_2 * x_2 + \dots + b_n * x_n$$

Dependent variable (DV) (points to  $y$ )      Independent variables (IVs) (points to  $x_1, x_2, \dots, x_n$ )  
 Constant (points to  $b_0$ )

Simple Linear Regression:



$$y = b_0 + b_1 * x$$

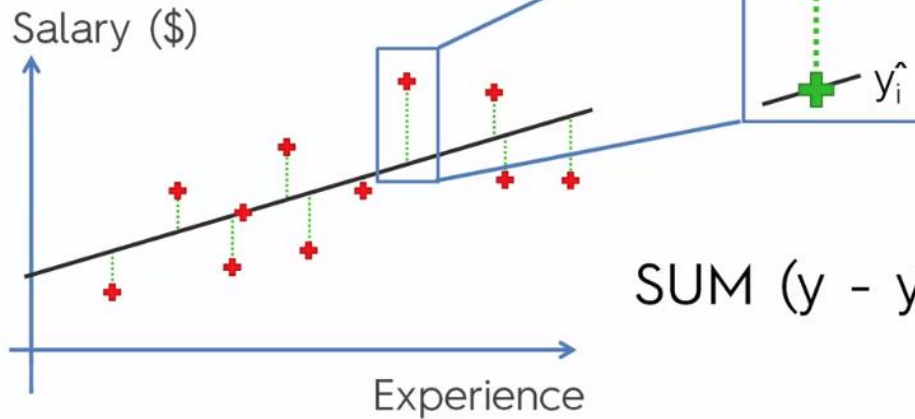
↓

$$\text{Salary} = \text{b}_0 + \text{b}_1 * \text{Experience}$$

(Note: In the original image,  $b_0$  is circled in red and  $b_1$  is circled in green, corresponding to the 30k intercept and the slope of 10k per year respectively.)

# Ordinary Least Squares

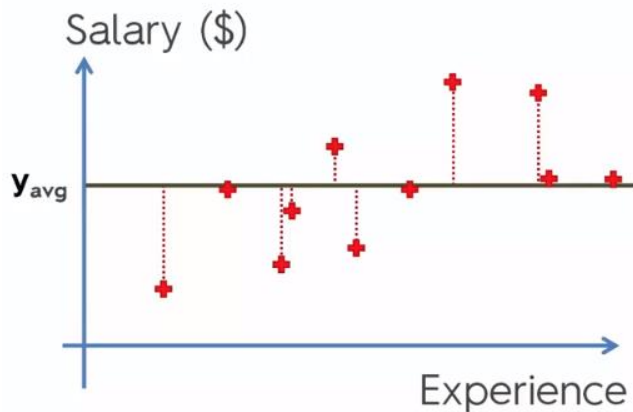
Simple Linear Regression:



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

## R Squared

Simple Linear Regression:



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

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

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

# Adjusted R<sup>2</sup>

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

R<sup>2</sup> – 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:**

$$+ b_3 * x_3$$

SS<sub>res</sub> → Min

R<sup>2</sup> 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