Linear Regression

Simple linear regression-

1. Statistical method that allows us to summarize and find relations between two continuous (quantitative) variables.
2. One variable is denoted as predictor, explanatory or independent variable.
3. Another variable is denoted as response, outcome or dependent variable.
4. We are interested in statistical relationship and not deterministic relationship.
5. Statistical relationships are not perfect relationships eg- relation between height and weight, driving speed and gas mileage…

Best Fit line-

1. In order to summarize relation between we need to find the best fitting line.
2. Common notations-

- yi - denotes the observed response for experimental unit *i*
- xi - denotes the predictor value for experimental unit *i*
- ŷi - is the predicted response (or fitted value) for experimental unit *i*
- Then, the equation for the best fitting line is:

ŷi = b0+b1xi

3. Experimental unit is the object or person on which the measurement is made.

Prediction error-

1. The predictions made wouldn't be perfectly correct there will be some prediction error (or residual error).
2. Size of prediction error-

$$\mathbf{ei=yi-\hat{y}i}$$

3. Line that fits data best is one which has **n- prediction errors** – one for each observed data point – **are as small as possible in overall sense**.
4. One way to minimize prediction error is to use "least squared error criteria", which says to "minimize the sum of squared prediction errors." That is-
   - The equation of the best fitting line is: ŷ i=b0+b1xi
   - We just need to find the values b_0 and b_1 that make the sum of the squared prediction errors the smallest it can be.
   - That is, we need to find the values b_0 and b_1 that minimize:

$$Q=\sum i=1\ (yi-\hat{y}\ i)^2$$