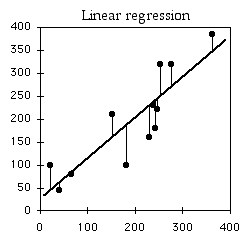
**Simple Linear Regression:**

It is a statistical method that allows us to study relationships between two continuous variables.

* variable **x** , which is the **predictor, explanatory, or independent** variable.
* variable **y**, which is the **response, outcome, or dependent** variable.



line equation

Y=β0+β1X

β1(Slope) & β0(Intercept)

How to find the best fit line(finding coefficients β1 & β0 ) ?

A **line of best fit**is a straight line that is the best approximation of the given set of data.

We can find the coefficients of the line using the following approaches :

* **Ordinary Least Square Method**

**To find the best possible line we first have to minimize the error in our model .**

**The following are the steps :**

* + 1. Import libraries
    2. Split the data into test and train
    3. Check for missing data
    4. Minimizing the sum of all squared errors of each point on the train and calculate (β1 & β0)

Math and formulas in the following links :

* + - 1. <https://are.berkeley.edu/courses/EEP118/current/derive_ols.pdf>
      2. <http://users.stat.ufl.edu/~winner/sta6208/notes1.pdf>
    1. Predict on the test data
    2. Check the goodness of the model (Root Mean Squared Error or Coefficient of Determination(R-square Score)
* **Gradient Descent Approach**

**Gradient Descent is an optimization algorithm.**

**The following are the steps:**

* + 1. Import libraries
    2. Split the data into test and train
    3. Check for missing data
    4. Implement The cost function known as mean squared error or MSE on train to determine how far we are from describing the real relationship.(high cost value – far from describing , low cost value – close to describing)
    5. We use gradient descent on train to minimize the cost .

math on gradient descent and cost in the following link:

* + - 1. <http://mccormickml.com/2014/03/04/gradient-descent-derivation/>
    1. Train the data to get the best β1 & β0 values
    2. Predict on test data
    3. Check the goodness of the model (Root Mean Squared Error or Coefficient of Determination(R-square Score)

We can also use the **sklearn.linear\_regression** library in python to fit the model to the train data.