## **1. Simple Linear Regression (SLR):**

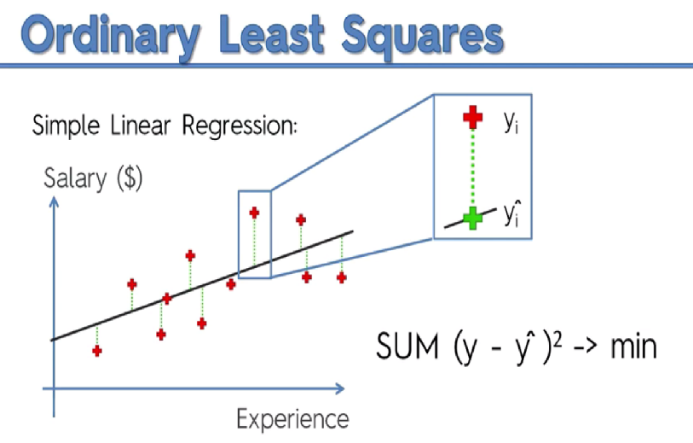
1. Simple linear regression is a statistical technique for finding the existence of an association relationship between dependent variable and independent variable.
2. SLR implies that there is only one independent variable and in the model.
3. Linear regression mean that relationship between response variable (y) and regression co-efficient () is linear.
4. **Example of SLR:** SLR draws all possible trend lines through the data as shown below and counts the sum of the squares every single time, records this temporarily and it find the minimum one out of those recorded data.



Fig 1. SLR for the Salary and experience explained

* – variation in y (Response variable y explained by regression model)
* – Variation explained by the model
* (- Intercept or Constant, – Coefficient, - independent variable)
* – Epsilon i explains variation in y not explained by the model.

1. Ordinary Least Square (OLS) is used to estimate regression parameters salary and experience. It will calculate and find best fitting line.



1. Model Evaluation:

* Higher value of is better fit, But be aware of spurious regression. Higher value is not necessarily a good indicator of correctness of the model it could be spurious regression. is also known as coefficient of determination.
* Hypothesis test for regression coefficient [T-test]: p-value is <0.05. We reject hypothesis and conclude that there is significant evidence suggesting a linear relationship between x and y.
* Test for overall model -Analysis of Variance(ANOVA) or F-test: In SLR, F-test =t-test. P-value is same for F-test & T-test.

1. **Extra Notes :**

SLR Model Building Steps:

1. Collect/Extract data: Which is time consuming and expensive step in the SLR model building.
2. Pre-process data: Ensure quality of data for issues such as reliability, completeness, usefulness, accuracy and missing data.
3. Divide Data into training and validation data set: Proportion of training set is usually 70% to 80%.
4. Perform descriptive analytics: It is goof to practice to perform descriptive analytics before moving to predictive analytics.

* Find variability in the model & visualization of data through box plot, scatterplot,etc.
* Box plot will show the outliers in the data
* Scatter plot reveal if there is any obvious relationship between 2 variables.

1. Define the functional form of relationship: Scatter plot may assist the modeler to define the right functional form. Analysis of residual will be useful in case the model uses an incorrect functional form.
2. Estimate the regression parameters: Ordinary Least Square (OLS) is used to estimate regression parameters. OLS provides the best linear unbiased estimate [BLUE].
3. Perform Regression model diagnostics: Before deploying the model it needs to be performed diagnostics test using all the assumptions. If the model assumptions are violated then remedial measure has to be used.
4. Validate model using validation data set.
5. Decide on the model deployment.

## **1.1.Python and R Libraries and Functions:**

R:

**Libraries:** library(ggplot2)

**Functions:**

lm(),predict()

ggplot(),geom\_point(),aes(),geom\_line()ggtitle(),xlab(),ylab()

Python:

**Libraries:** from sklearn.linear\_model import LinearRegression

**Functions:**

LinearRegression(),predict()

Plot related Libraries: scatter(),plot(),title(),xlabel(),ylabel(),show()