**Linear Regression** analysis is widely used as a statistical tool to establish a relationship between two variables. One is predictor variable whose values are gathered from experiments and surveys and other is response variable whose value is derived from the predictor variable.

In Linear Regression there are two variables that are related through an equation, where exponent(power) of both these variables are 1. Mathematically a linear relationship represents a straight line when plotted as a graph. A nonlinear relationship where the exponent of any variable is not equal to 1 creates a curve. So,

Y = ax + b

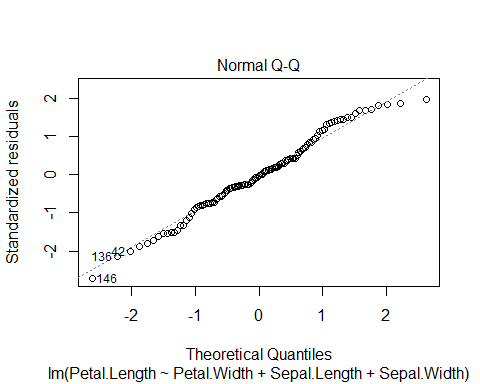
Where y is the response variable.

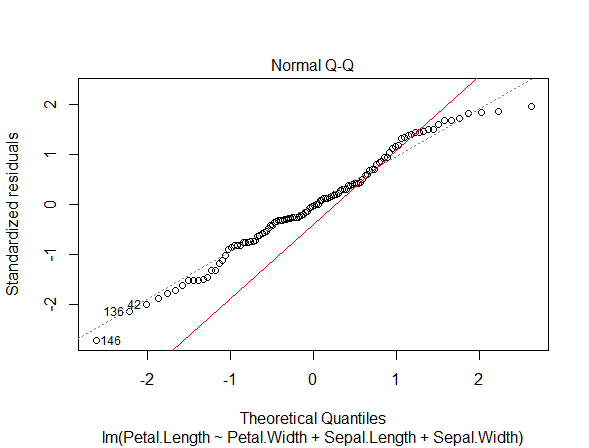
x is the predictor variable.

a and b are constants which are called the coefficients.

Let us understand Why, What and How in Linear Regression model. Firstly, Why Linear Regression? This regression model draws the relationship between the dependent and independent variables, where dependent variable is otherwise known as predictor variable and independent variable is known as response variable. Now let’s investigate What is Linear Regression? This is a statistical analysis that attempts to show the relationship between two variables. How Linear Regression works? With the data provided by the dataset, we draw a plot between independent and dependent variable and spot a mean value for those values. It selects the best fit line that passes through the mean of the data. But as we know there will be multiple lines that passes through that mean point. To solve this scenario, we move the line until we have least square distance from all the data points.

Let us understand the step by step process of establishing a linear regression model of a sample data.

* Gathering sample data.
* Cleaning and splitting data.
* Creating relationship between variables.
* Using lm() function deriving responses.
* Plotting the train data and drawing a line that is straight.
* Predicting the test data value and validating.
* Linear-Regression.R
* navya
* 2020-11-05
* #reading the data.  
  mydata <- iris  
  #Showing the data.  
  mydata
* ## Sepal.Length Sepal.Width Petal.Length Petal.Width Species  
  ## 1 5.1 3.5 1.4 0.2 setosa  
  ## 2 4.9 3.0 1.4 0.2 setosa  
  ## 3 4.7 3.2 1.3 0.2 setosa  
  ## 4 4.6 3.1 1.5 0.2 setosa  
  ## 5 5.0 3.6 1.4 0.2 setosa  
  ## 6 5.4 3.9 1.7 0.4 setosa  
  ## 7 4.6 3.4 1.4 0.3 setosa  
  ## 8 5.0 3.4 1.5 0.2 setosa  
  ## 9 4.4 2.9 1.4 0.2 setosa  
  ## 10 4.9 3.1 1.5 0.1 setosa  
  ## 11 5.4 3.7 1.5 0.2 setosa  
  ## 12 4.8 3.4 1.6 0.2 setosa  
  ## 13 4.8 3.0 1.4 0.1 setosa  
  ## 14 4.3 3.0 1.1 0.1 setosa  
  ## 15 5.8 4.0 1.2 0.2 setosa  
  ## 16 5.7 4.4 1.5 0.4 setosa  
  ## 17 5.4 3.9 1.3 0.4 setosa  
  ## 18 5.1 3.5 1.4 0.3 setosa  
  ## 19 5.7 3.8 1.7 0.3 setosa  
  ## 20 5.1 3.8 1.5 0.3 setosa  
  ## 21 5.4 3.4 1.7 0.2 setosa  
  ## 22 5.1 3.7 1.5 0.4 setosa  
  ## 23 4.6 3.6 1.0 0.2 setosa  
  ## 24 5.1 3.3 1.7 0.5 setosa  
  ## 25 4.8 3.4 1.9 0.2 setosa  
  ## 26 5.0 3.0 1.6 0.2 setosa  
  ## 27 5.0 3.4 1.6 0.4 setosa  
  ## 28 5.2 3.5 1.5 0.2 setosa  
  ## 29 5.2 3.4 1.4 0.2 setosa  
  ## 30 4.7 3.2 1.6 0.2 setosa  
  ## 31 4.8 3.1 1.6 0.2 setosa  
  ## 32 5.4 3.4 1.5 0.4 setosa  
  ## 33 5.2 4.1 1.5 0.1 setosa  
  ## 34 5.5 4.2 1.4 0.2 setosa  
  ## 35 4.9 3.1 1.5 0.2 setosa  
  ## 36 5.0 3.2 1.2 0.2 setosa  
  ## 37 5.5 3.5 1.3 0.2 setosa  
  ## 38 4.9 3.6 1.4 0.1 setosa  
  ## 39 4.4 3.0 1.3 0.2 setosa  
  ## 40 5.1 3.4 1.5 0.2 setosa  
  ## 41 5.0 3.5 1.3 0.3 setosa  
  ## 42 4.5 2.3 1.3 0.3 setosa  
  ## 43 4.4 3.2 1.3 0.2 setosa  
  ## 44 5.0 3.5 1.6 0.6 setosa  
  ## 45 5.1 3.8 1.9 0.4 setosa  
  ## 46 4.8 3.0 1.4 0.3 setosa  
  ## 47 5.1 3.8 1.6 0.2 setosa  
  ## 48 4.6 3.2 1.4 0.2 setosa  
  ## 49 5.3 3.7 1.5 0.2 setosa  
  ## 50 5.0 3.3 1.4 0.2 setosa  
  ## 51 7.0 3.2 4.7 1.4 versicolor  
  ## 52 6.4 3.2 4.5 1.5 versicolor  
  ## 53 6.9 3.1 4.9 1.5 versicolor  
  ## 54 5.5 2.3 4.0 1.3 versicolor  
  ## 55 6.5 2.8 4.6 1.5 versicolor  
  ## 56 5.7 2.8 4.5 1.3 versicolor  
  ## 57 6.3 3.3 4.7 1.6 versicolor  
  ## 58 4.9 2.4 3.3 1.0 versicolor  
  ## 59 6.6 2.9 4.6 1.3 versicolor  
  ## 60 5.2 2.7 3.9 1.4 versicolor  
  ## 61 5.0 2.0 3.5 1.0 versicolor  
  ## 62 5.9 3.0 4.2 1.5 versicolor  
  ## 63 6.0 2.2 4.0 1.0 versicolor  
  ## 64 6.1 2.9 4.7 1.4 versicolor  
  ## 65 5.6 2.9 3.6 1.3 versicolor  
  ## 66 6.7 3.1 4.4 1.4 versicolor  
  ## 67 5.6 3.0 4.5 1.5 versicolor  
  ## 68 5.8 2.7 4.1 1.0 versicolor  
  ## 69 6.2 2.2 4.5 1.5 versicolor  
  ## 70 5.6 2.5 3.9 1.1 versicolor  
  ## 71 5.9 3.2 4.8 1.8 versicolor  
  ## 72 6.1 2.8 4.0 1.3 versicolor  
  ## 73 6.3 2.5 4.9 1.5 versicolor  
  ## 74 6.1 2.8 4.7 1.2 versicolor  
  ## 75 6.4 2.9 4.3 1.3 versicolor  
  ## 76 6.6 3.0 4.4 1.4 versicolor  
  ## 77 6.8 2.8 4.8 1.4 versicolor  
  ## 78 6.7 3.0 5.0 1.7 versicolor  
  ## 79 6.0 2.9 4.5 1.5 versicolor  
  ## 80 5.7 2.6 3.5 1.0 versicolor  
  ## 81 5.5 2.4 3.8 1.1 versicolor  
  ## 82 5.5 2.4 3.7 1.0 versicolor  
  ## 83 5.8 2.7 3.9 1.2 versicolor  
  ## 84 6.0 2.7 5.1 1.6 versicolor  
  ## 85 5.4 3.0 4.5 1.5 versicolor  
  ## 86 6.0 3.4 4.5 1.6 versicolor  
  ## 87 6.7 3.1 4.7 1.5 versicolor  
  ## 88 6.3 2.3 4.4 1.3 versicolor  
  ## 89 5.6 3.0 4.1 1.3 versicolor  
  ## 90 5.5 2.5 4.0 1.3 versicolor  
  ## 91 5.5 2.6 4.4 1.2 versicolor  
  ## 92 6.1 3.0 4.6 1.4 versicolor  
  ## 93 5.8 2.6 4.0 1.2 versicolor  
  ## 94 5.0 2.3 3.3 1.0 versicolor  
  ## 95 5.6 2.7 4.2 1.3 versicolor  
  ## 96 5.7 3.0 4.2 1.2 versicolor  
  ## 97 5.7 2.9 4.2 1.3 versicolor  
  ## 98 6.2 2.9 4.3 1.3 versicolor  
  ## 99 5.1 2.5 3.0 1.1 versicolor  
  ## 100 5.7 2.8 4.1 1.3 versicolor  
  ## 101 6.3 3.3 6.0 2.5 virginica  
  ## 102 5.8 2.7 5.1 1.9 virginica  
  ## 103 7.1 3.0 5.9 2.1 virginica  
  ## 104 6.3 2.9 5.6 1.8 virginica  
  ## 105 6.5 3.0 5.8 2.2 virginica  
  ## 106 7.6 3.0 6.6 2.1 virginica  
  ## 107 4.9 2.5 4.5 1.7 virginica  
  ## 108 7.3 2.9 6.3 1.8 virginica  
  ## 109 6.7 2.5 5.8 1.8 virginica  
  ## 110 7.2 3.6 6.1 2.5 virginica  
  ## 111 6.5 3.2 5.1 2.0 virginica  
  ## 112 6.4 2.7 5.3 1.9 virginica  
  ## 113 6.8 3.0 5.5 2.1 virginica  
  ## 114 5.7 2.5 5.0 2.0 virginica  
  ## 115 5.8 2.8 5.1 2.4 virginica  
  ## 116 6.4 3.2 5.3 2.3 virginica  
  ## 117 6.5 3.0 5.5 1.8 virginica  
  ## 118 7.7 3.8 6.7 2.2 virginica  
  ## 119 7.7 2.6 6.9 2.3 virginica  
  ## 120 6.0 2.2 5.0 1.5 virginica  
  ## 121 6.9 3.2 5.7 2.3 virginica  
  ## 122 5.6 2.8 4.9 2.0 virginica  
  ## 123 7.7 2.8 6.7 2.0 virginica  
  ## 124 6.3 2.7 4.9 1.8 virginica  
  ## 125 6.7 3.3 5.7 2.1 virginica  
  ## 126 7.2 3.2 6.0 1.8 virginica  
  ## 127 6.2 2.8 4.8 1.8 virginica  
  ## 128 6.1 3.0 4.9 1.8 virginica  
  ## 129 6.4 2.8 5.6 2.1 virginica  
  ## 130 7.2 3.0 5.8 1.6 virginica  
  ## 131 7.4 2.8 6.1 1.9 virginica  
  ## 132 7.9 3.8 6.4 2.0 virginica  
  ## 133 6.4 2.8 5.6 2.2 virginica  
  ## 134 6.3 2.8 5.1 1.5 virginica  
  ## 135 6.1 2.6 5.6 1.4 virginica  
  ## 136 7.7 3.0 6.1 2.3 virginica  
  ## 137 6.3 3.4 5.6 2.4 virginica  
  ## 138 6.4 3.1 5.5 1.8 virginica  
  ## 139 6.0 3.0 4.8 1.8 virginica  
  ## 140 6.9 3.1 5.4 2.1 virginica  
  ## 141 6.7 3.1 5.6 2.4 virginica  
  ## 142 6.9 3.1 5.1 2.3 virginica  
  ## 143 5.8 2.7 5.1 1.9 virginica  
  ## 144 6.8 3.2 5.9 2.3 virginica  
  ## 145 6.7 3.3 5.7 2.5 virginica  
  ## 146 6.7 3.0 5.2 2.3 virginica  
  ## 147 6.3 2.5 5.0 1.9 virginica  
  ## 148 6.5 3.0 5.2 2.0 virginica  
  ## 149 6.2 3.4 5.4 2.3 virginica  
  ## 150 5.9 3.0 5.1 1.8 virginica
* #verifying the relationship between the variables.  
  plot(mydata)
* 
* #installing the packages to split the data.  
  require(caTools)
* ## Loading required package: caTools
* #setting the seed to shuffle the data.  
  set.seed(1234)  
  #spliting the data 75% and 25% as train and test dataset.  
  sam <- sample.split(mydata$Species, SplitRatio = .75)  
  train\_data <- subset(mydata, sam == TRUE)  
  test\_data <- subset(mydata, sam == FALSE)  
  #using linear regression model function.  
  linmod <- lm(Petal.Length~Petal.Width+Sepal.Length+Sepal.Width,data = train\_data)  
  #Plotting the data of linmod in line graph.  
  plot(linmod)
* 
* #A straight line passing through the point.  
  abline(linmod, col = 'red')
* ## Warning in abline(linmod, col = "red"): only using the first two of 4 regression  
  ## coefficients



#predicting the test dataset values.  
pmod <- predict(linmod,test\_data)  
pmod

## 5 11 14 16 26 28 29 36   
## 1.359502 1.585835 1.062353 1.688857 1.709212 1.560096 1.618381 1.592642   
## 39 40 47 50 53 58 60 61   
## 1.282284 1.547227 1.314086 1.534357 4.936960 3.177979 3.811693 3.482274   
## 66 72 74 81 86 90 92 100   
## 4.645874 4.245025 4.096248 3.753684 4.270488 3.992951 4.277231 3.960406   
## 113 116 117 122 123 124 131 135   
## 5.816748 5.713112 5.156955 4.930685 6.424935 5.189501 6.062694 4.510371   
## 137 140 142 149   
## 5.674163 5.829618 6.127171 5.454232

**References:**

1. <https://www.youtube.com/watch?v=1-URCcgTBf4&ab_channel=edureka%21>
2. <https://stackoverflow.com/questions/17200114/how-to-split-data-into-training-testing-sets-using-sample-function>
3. <https://www.youtube.com/watch?v=2Sb1Gvo5si8&ab_channel=Simplilearn>
4. <https://www.tutorialspoint.com/r/r_linear_regression.htm>