R – Logistic Regression

- The Logistic Regression is a regression model in which the response variable (dependent variable) has categorical values such as True/False or 0/1.

- It actually measures the probability of a binary response as the value of response variable based on the mathematical equation relating it with the predictor variables.

$$p = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p)}}$$

- Following is the description of the parameters used:

- **p** is the response variable.

- **x** is the predictor variable.

- **a** and **b** are the coefficients which are numeric constants.

- The function used to create the regression model is the **glm()** function.
- Syntax
- glm(formula,data,family)
- Following is the description of the parameters used:
- **formula** is the symbol presenting the relationship between the variables.
- **data** is the data set giving the values of these variables
- **family** is R object to specify the details of the model. It's value is binomial for logistic regression.

- #Problem:#By use of the logistic regression equation of vehicle transmission in the data set mtcars, estimate the probability of a vehicle being fitted with a manual transmission if it has a 120hp engine and weights 2800 lbs.
- #Solution:We apply the function glm to a formula that describes the transmission type (am) by the horsepower (hp) and weight (wt).
- # This creates a generalized linear model (GLM) in the binomial family.

- mtcars
- am.glm = glm(formula=am ~ hp + wt, data=mtcars, family=binomial)
- am.glm
- #We then wrap the test parameters inside a data frame newdata.
- newdata = data.frame(hp=120, wt=2.8)
- #apply the function predict to the generalized linear model am.glm along with newdata.
- #We will have to select response prediction type in order to obtain the predicted probability.
- predict(am.glm, newdata, type="response")

- #Answer:For an automobile with 120hp engine and 2800 lbs weight,
- the probability of it being fitted with a manual transmission is about 64%.

- In the summary as the p-value in the last column is more than 0.05 for the variables "cyl“ and "hp", we consider them to be insignificant in contributing to the value of the variable "am". Only weight (wt) impacts the "am" value in this regression model.