Question-1:

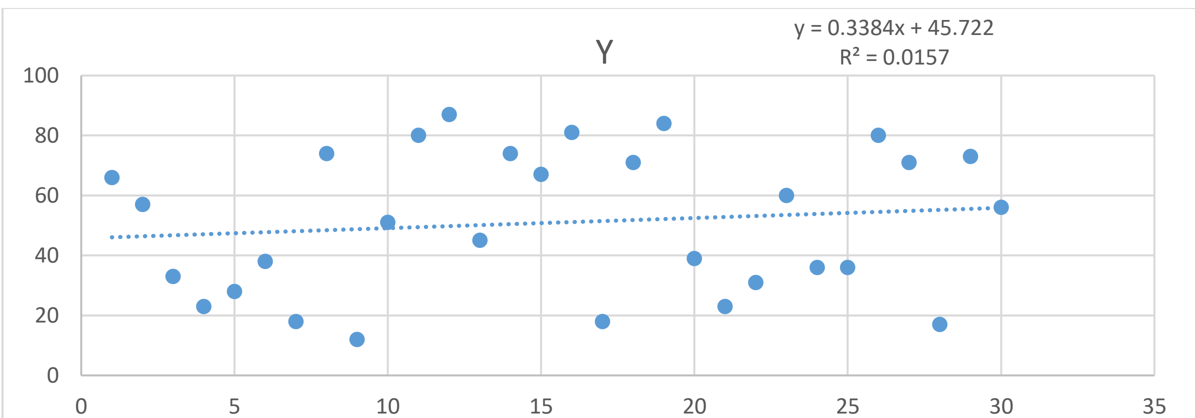
List down at least three main assumptions of linear regression and explain them in your own words. To explain an assumption, take an example or a specific use case to show why the assumption makes sense.

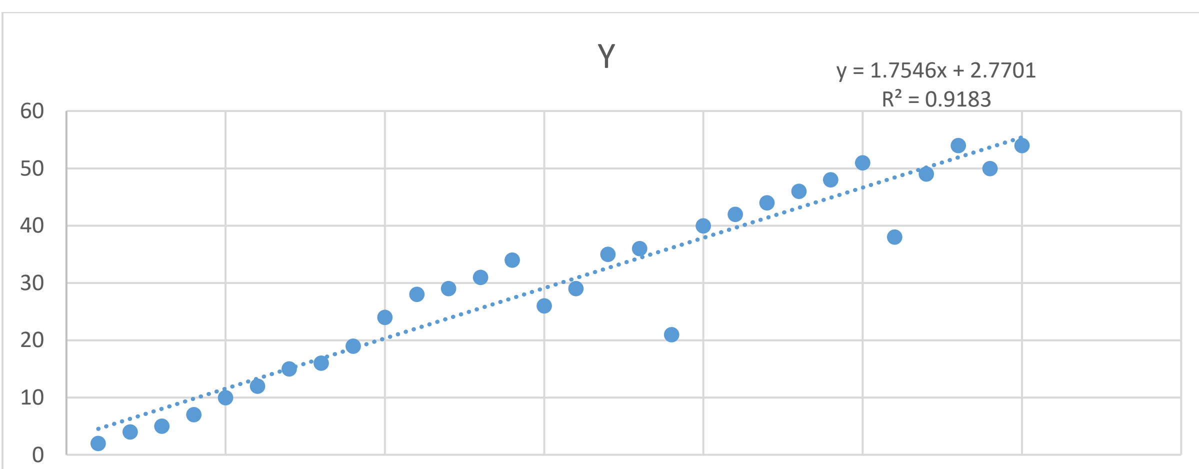
Solution:-
The three assumptions of linear regression are:-

1. Linearity
2. No Multi-collinearity
3. Homoscedasticity or absence of Heteroscedasticity

Linearity
Relationship between the dependent variable (response) & independent variable(s) aka predictors should be linear & additive. By Linear it implies change in one unit of independent variable is constant irrespective of value of it & additivity implies that effect of any independent variable is independent of other independent variables.

Graph below depicting scatter plots of 2 different response with one as non-linear & other linear pattern.





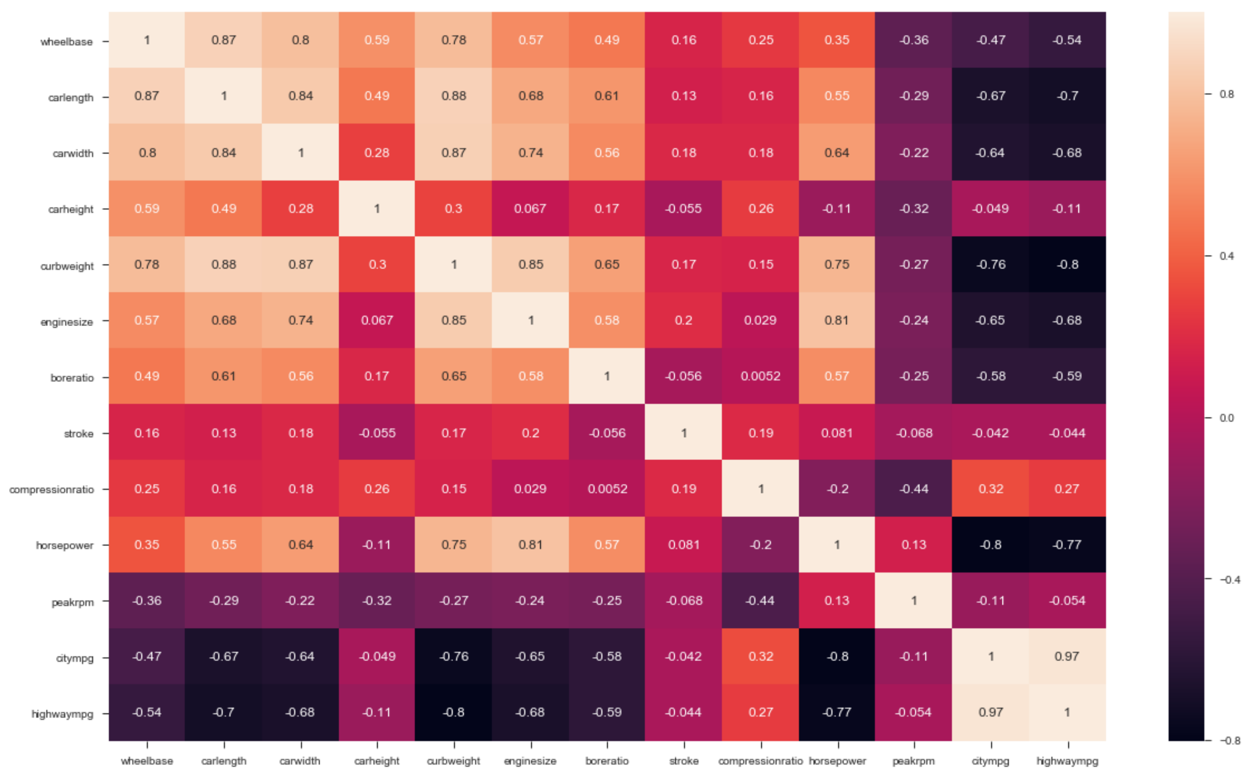
We observed that response & predictor(s) which follows linearity is better fit for Regression Analysis as R-Square measure is great for linear model indicating it can explain the response more accurately.

No Multi-collinearity

This phenomenon exists when Independent variables are found to be highly or moderately correlated. It becomes a challenge to figure out true relationship of independent variables aka predictors with response variable.

It can be detected by below methods:-

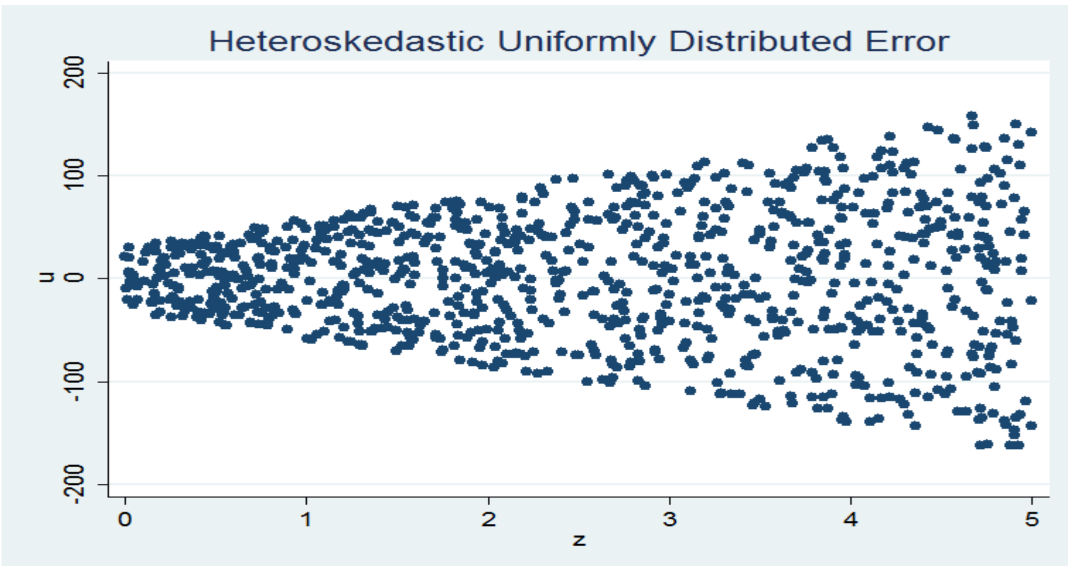
a. Scatter plots of independent variables.
b. Correlation matrix among all independent variables.
c. VIF (variance inflation factor) – where in one of the independent variable is used as dependent variable & rest others as dependent variable. Then try to fit the model & get r-square VIF = 1/(1-r square) any value less than 4 is good & more than it should be considered for removing as an predictor.



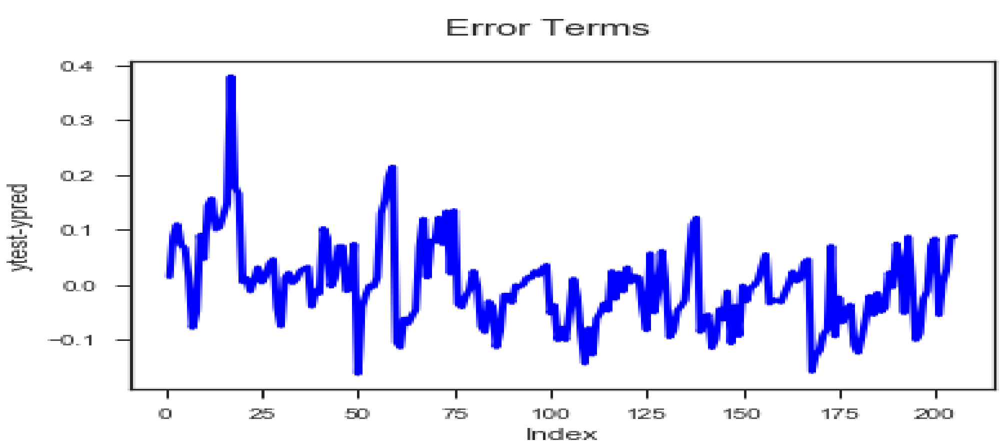
From the above co-relation matrix we can determine predictors which are moderately or highly correlated.

Homoscedasticity or absence of Heteroscedasticity

The error terms must have constant variance. A plot of standardized residuals versus predicted values can show whether points are equally distributed across all values of the independent variables. If Heteroscedasticity exists then the plot would exhibit a funnel shape pattern.



Homoscedasticity



Question-2:-

Explain the gradient descent algorithm in the following two parts:

Illustrate at least two iterations of the algorithm using the univariate function

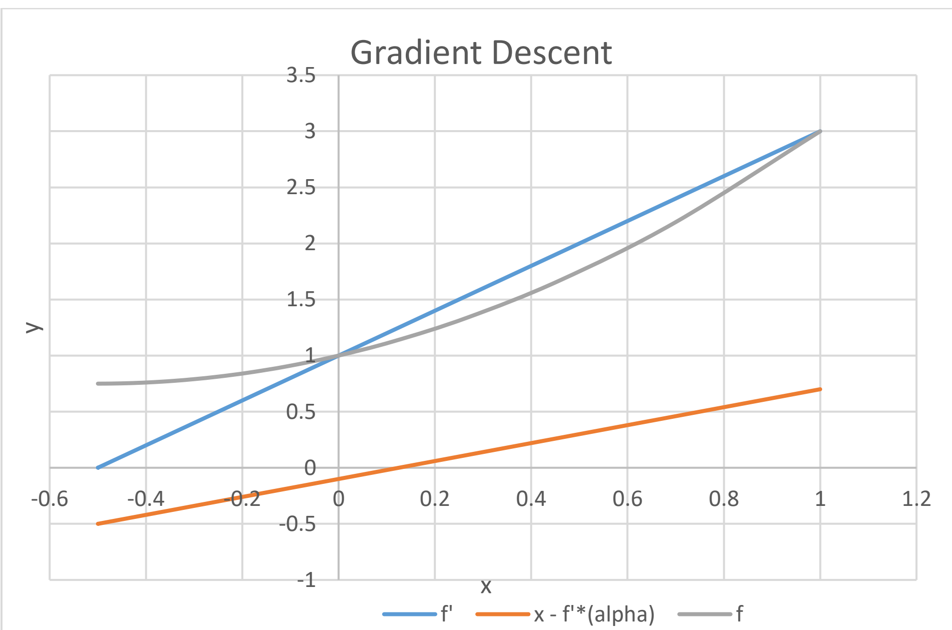
J(x) =x2+x+1. Assume a learning rate η=0.1 and an initial guess x0=1 and demonstrate that the iterations converge towards the minima. Also, report the minima (which you can compute using the closed form solution).

Solution:-

1. Gradient Descent

f(x)	x^2+x+1	**alpha**	0.1
f'(x)	2x+1		
x0	1		

x	**f'**	**x - f'*(alpha)**	**f**
1	3	0.7	3
0.7	2.4	0.46	2.19
0.46	1.92	0.268	1.6716



2. Closed Loop: - f'(x) equated to 0 would give minimization & f''(*x) > 0 (add comment)
   f(x) = x2 + x + 1
   f'(x) = 2x + 1
   f''(x) = 2

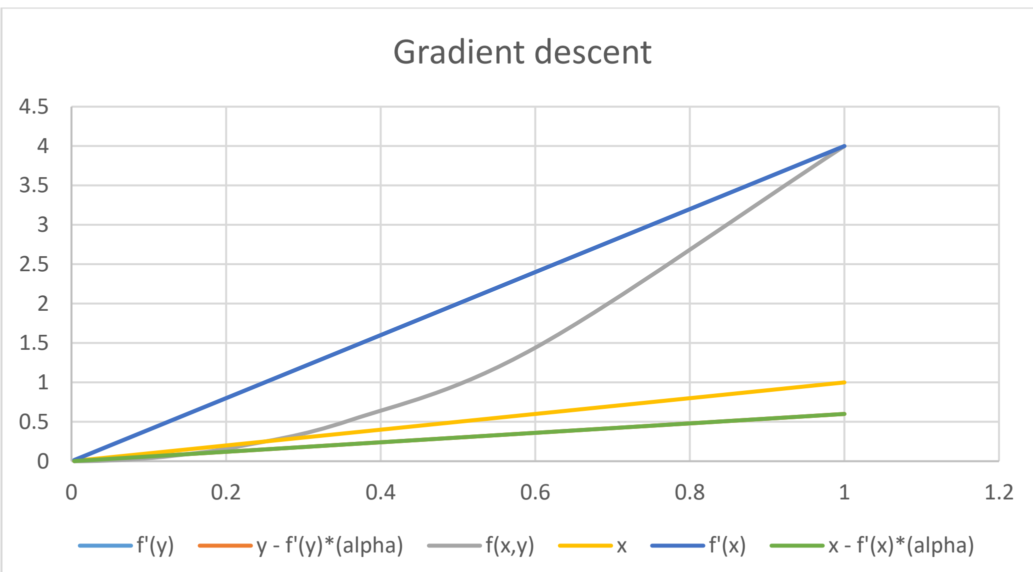
   So on equating 2x+1=0, we get x = -1/2 i.e. -0.5 & f''(x) = 2

Illustrate at least two iterations of the algorithm using the bivariate function of two independent variables

J(x,y)=x2+2xy+y2. Assume a learning rate η=0.1 and an initial guess (x0,y0)=(1,1) . Report the minima and show that the solution converges towards it.

f(x,y)	x^2+2xy+y^2					
f'(x)	2x+2y		**alpha**	0.1		
f'(y)	2x+2y					
x0	1	**y0**	1			

y	**f'(y)**	**y - f'(y)*(alpha)**	**f(x,y)**	**x**	**f'(x)**	**x - f'(x)*(alpha)**
1	4	0.6	4	1	4	0.6
0.6	2.4	0.36	1.44	0.6	2.4	0.36
0.36	1.44	0.216	0.5184	0.36	1.44	0.216



From closed loop we know partial derivative w.r.t. to x & y equated to 0 would give minima so,
∂F/∂x = 2x + 2y = 0
∂F/∂y = 2x + 2y = 0
∂2F/∂x2 = 2 i.e. > 0
∂2F/∂y2 = 2 i.e. > 0
From the above equation it is evident it has multiple minima's one of them is f(0,0) other ones are x = -y for all possible values of x & y.