Q.3:


```
clc
clear all

Data=load('foundry.txt');
Estimation_Data=Data(1:300,:);
Est_X=Estimation_Data(:,1:(size(Estimation_Data,2)-1));
Est_y=Estimation_Data(:,size(Estimation_Data,2));
Validation_Data=Data(301:size(Data,1),:);
Validation_X=Validation_Data(:,1:(size(Estimation_Data,2)-1));
Validation_y=Validation_Data(:,size(Estimation_Data,2));

% Pre-processing the data
% Scaling using range for each variable - The same line ! We're just
% shifting the data.
RangeEst=range(Est_X,1);
RangeVal=range(Validation_X,1);
for i=1:size(Est_X,1)
Pre1_Est_X(i,:)=(Est_X(i,:)-min(Est_X,[],1))./RangeEst;
end

for i=1:size(Validation_X,1)
Pre1_Validation_X(i,:)=(Validation_X(i,:)-min(Validation_X,[],1))./RangeVal;
end
```

Standard OLS Regression : model - using 300 data points and testing the model with the remaining

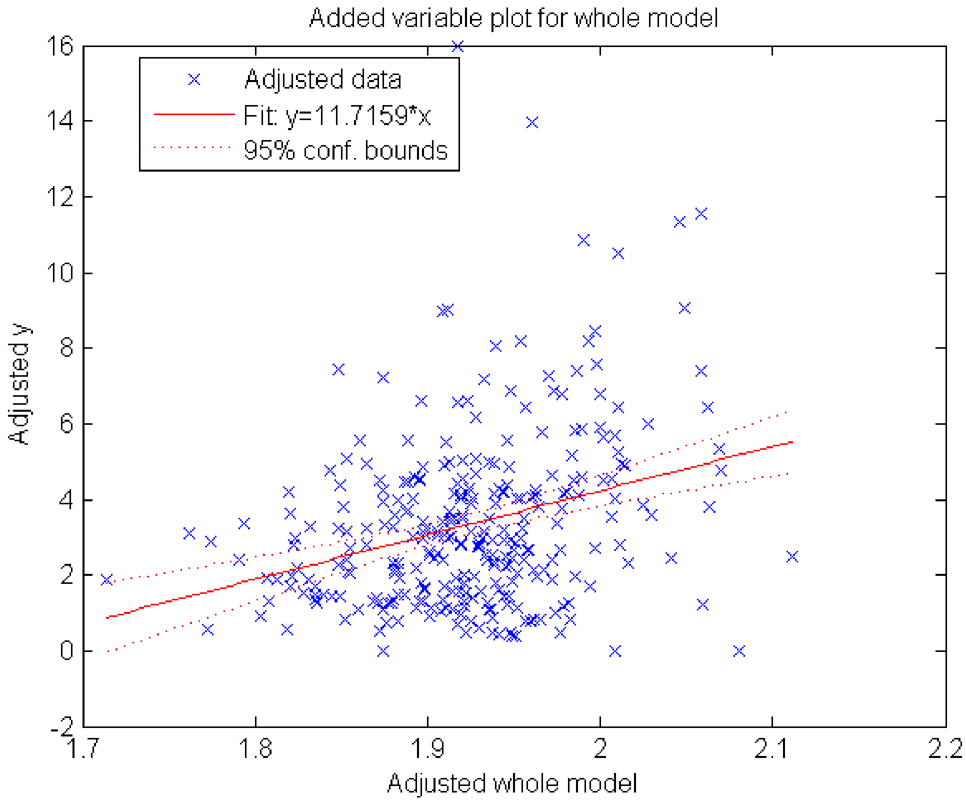
```
% Fitting linear model using standard ols (without weights)
mdl = fitlm(Est_X,Est_y)
% Plotting the model
figure(1)
plot(mdl)
CoeffOfDetermination = mdl.Rsquared.Ordinary % R squared Value

%Validating the model using TestData
Predicted_y=feval(mdl,Validation_X);
Pred_Error=Predicted_y-Validation_y;
MaxAbs_Pred_Error=max(abs(Pred_Error))
Standard_Deviation = std(Pred_Error)
figure(2)
hist(Pred_Error,15)
title('Validation Data Error Plot')
xlabel('Predicted value error')
ylabel('Frequency')

% Fitting linear model using standard ols (without weights)
mdl2 = fitlm(Pre1_Est_X,Est_y)
% Plotting the model
figure(3)
plot(mdl2)
CoeffOfDetermination = mdl2.Rsquared.Ordinary % R squared Value
```

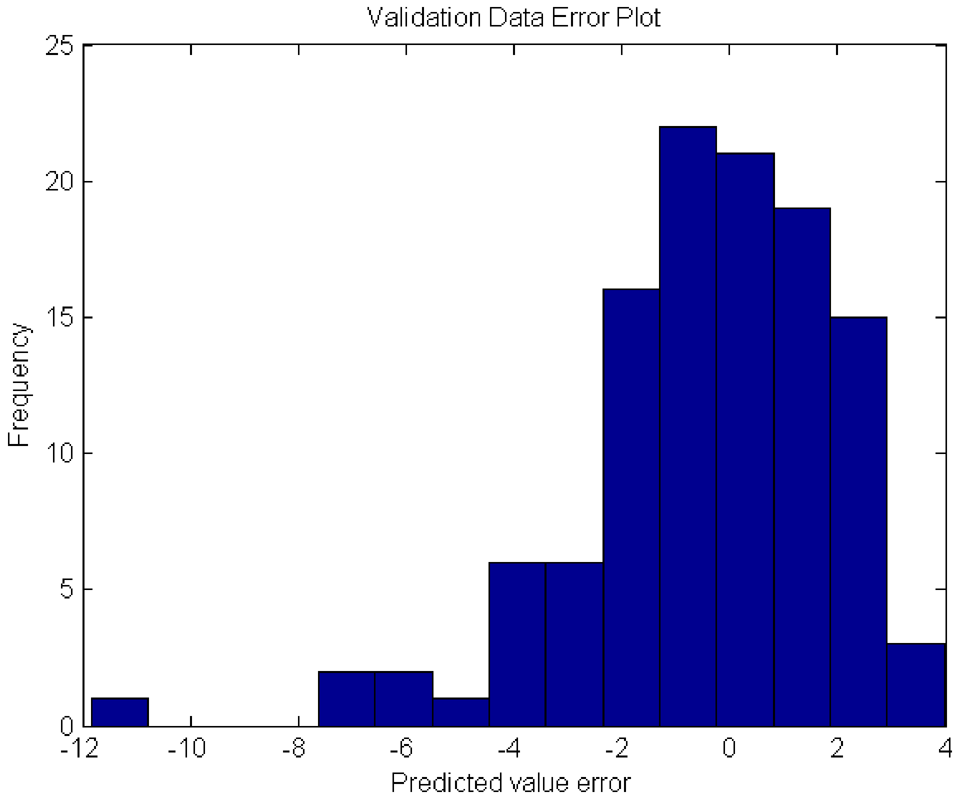
```
%Validating the model using TestData
Predicted_y=feval(mdl2,Pre1_Validation_X);
Pred_Error=Predicted_y-Validation_y;
MaxAbs_Pred_Error=max(abs(Pred_Error))
Standard_Deviation = std(Pred_Error)
figure(4)
hist(Pred_Error,15)
title('Validation Data Error Plot')
xlabel('Predicted value error')
ylabel('Frequency')
```

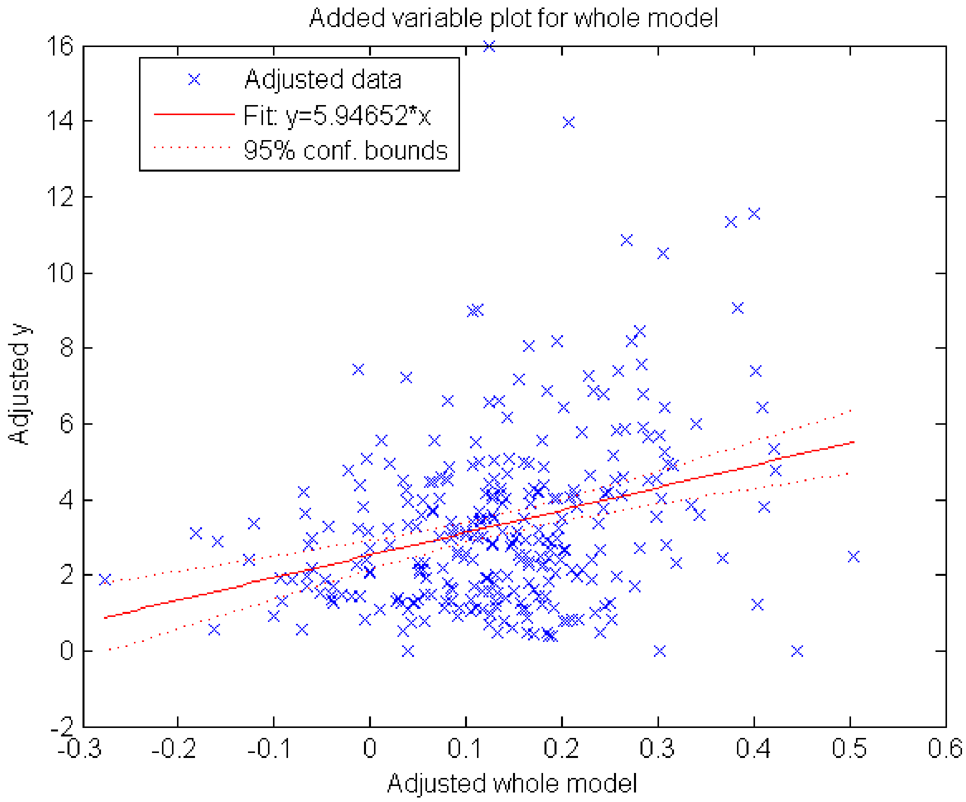
```
For normal data :
mdl =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12

Estimated Coefficients:
                   Estimate        SE         tStat       pValue
                   _________    _________    ________    _________

    (Intercept)      -19.206        24.55     -0.7823      0.43468
    x1               -0.8038       1.1561    -0.69529      0.48744
    x2               -1.0786       0.5795     -1.8613     0.063724
    x3              0.016928    0.0054416      3.1108    0.0020536
    x4              -0.21542      0.15118     -1.4249      0.15527
    x5               0.35679      0.34044       1.048       0.2955
    x6                2.8977       1.7246      1.6802     0.094011
    x7                11.224       10.541      1.0648      0.28785
    x8              0.056322      0.08127     0.69302      0.48886
    x9                0.8312       1.1846     0.70165      0.48346
    x10            0.0033919     0.013696     0.24765      0.80458
    x11             -0.30063       1.1258    -0.26704      0.78963
    x12             -0.34872      0.20822     -1.6747     0.095078


Number of observations: 300, Error degrees of freedom: 287
Root Mean Squared Error: 2.23
R-squared: 0.0939,  Adjusted R-Squared 0.056
F-statistic vs. constant model: 2.48, p-value = 0.00425

CoeffOfDetermination =

    0.0939


MaxAbs_Pred_Error =

   11.8396


Standard_Deviation =

    2.4209
```

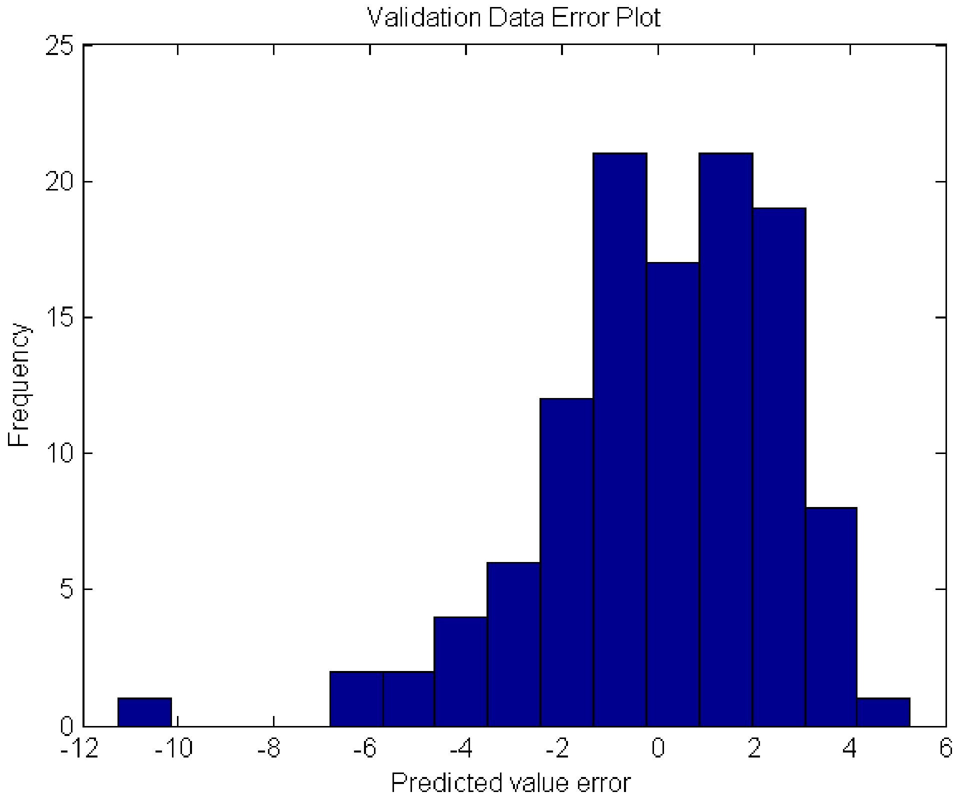
For scaled data based on range:

```
mdl2 =


Linear regression model:
    y ~ 1 + x1 + x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10 + x11 + x12

Estimated Coefficients:
                   Estimate      SE        tStat       pValue  
                   ________    _______    ________    _________

    (Intercept)      2.5242     1.0471      2.4106     0.016554
    x1             -0.64304    0.92485    -0.69529      0.48744
    x2               -3.007     1.6155     -1.8613     0.063724
    x3               3.7562     1.2075      3.1108    0.0020536
    x4               -1.187    0.83301     -1.4249      0.15527
    x5              0.82419    0.78641       1.048       0.2955
    x6               1.7096     1.0175      1.6802     0.094011
    x7               1.6836     1.5811      1.0648      0.28785
    x8              0.65146    0.94003     0.69302      0.48886
    x9               1.1803     1.6822     0.70165      0.48346
    x10             0.31206     1.2601     0.24765      0.80458
    x11            -0.30063     1.1258    -0.26704      0.78963
    x12             -1.3949    0.83289     -1.6747     0.095078


Number of observations: 300, Error degrees of freedom: 287
Root Mean Squared Error: 2.23
R-squared: 0.0939,  Adjusted R-Squared 0.056
F-statistic vs. constant model: 2.48, p-value = 0.00425

CoeffOfDetermination =

    0.0939


MaxAbs_Pred_Error =

   11.2518


Standard_Deviation =

    2.5011
```


Added variable plot for whole model
Adjusted data
Fit: y=11.7159*x
95% conf. bounds
Adjusted y
Adjusted whole model
16
14
12
10
8
6
4
2
0
-2
1.7
1.8
1.9
2
2.1
2.2


Validation Data Error Plot
Frequency
Predicted value error
25
20
15
10
5
0
-12
-10
-8
-6
-4
-2
0
2
4





Published with MATLAB® R2014a

Therefore we can see that on scaling the data by range the result does not vary for the case of OLS as is expected.