Regression Modelling Part-1

Dr Nisha Arora

Monday, June 06, 2016

# Simple Linear Regression

To import text file

data1 <- read.table("regression\_sample\_data1.txt",  
 header = TRUE)

To view the data set

data1

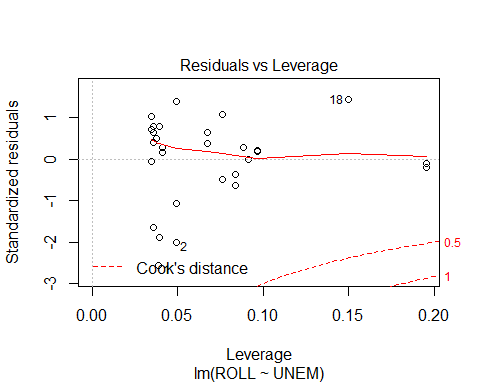
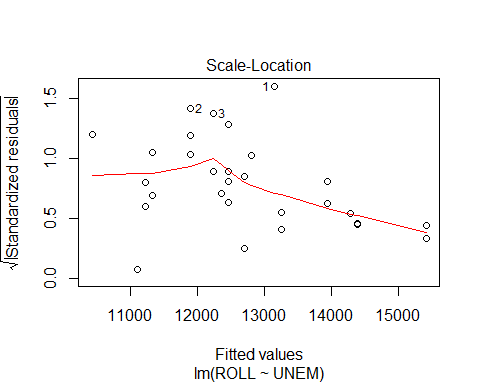
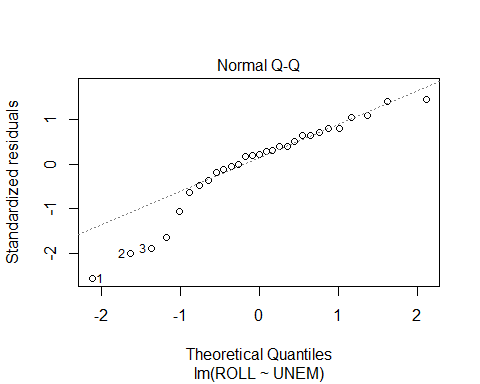
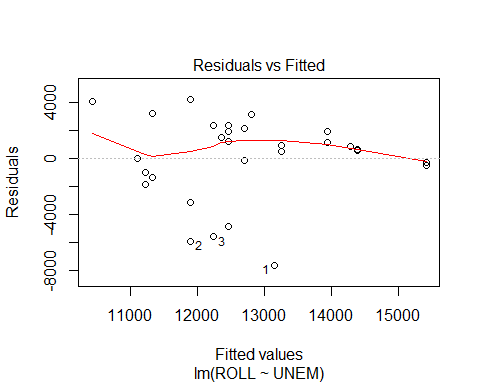
## S\_NO YEAR ROLL UNEM HGRAD INC  
## 1 1 1 5501 8.1 9552 1923  
## 2 2 2 5945 7.0 9680 1961  
## 3 3 3 6629 7.3 9731 1979  
## 4 4 4 7556 7.5 11666 2030  
## 5 5 5 8716 7.0 14675 2112  
## 6 6 6 9369 6.4 15265 2192  
## 7 7 7 9920 6.5 15484 2235  
## 8 8 8 10167 6.4 15723 2351  
## 9 9 9 11084 6.3 16501 2411  
## 10 10 10 12504 7.7 16890 2475  
## 11 11 11 13746 8.2 17203 2524  
## 12 12 12 13656 7.5 17707 2674  
## 13 13 13 13850 7.4 18108 2833  
## 14 14 14 14145 8.2 18266 2863  
## 15 15 15 14888 10.1 19308 2839  
## 16 16 16 14991 9.2 18224 2898  
## 17 17 17 14836 7.7 18997 3123  
## 18 18 18 14478 5.7 19505 3195  
## 19 19 19 14539 6.5 19800 3239  
## 20 20 20 14395 7.5 19546 3129  
## 21 21 21 14599 7.3 19117 3100  
## 22 22 22 14969 9.2 18774 3008  
## 23 23 23 15107 10.1 17813 2983  
## 24 24 24 14831 7.5 17304 3069  
## 25 25 25 15081 8.8 16756 3151  
## 26 26 26 15127 9.1 16749 3127  
## 27 27 27 15856 8.8 16925 3179  
## 28 28 28 15938 7.8 17231 3207  
## 29 29 29 16081 7.0 16816 3345

To predict the fall enrollment (ROLL) using the unemployment rate (UNEM)

m1=lm(ROLL~UNEM, data=data1)  
summary(m1)

##   
## Call:  
## lm(formula = ROLL ~ UNEM, data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7640.0 -1046.5 602.8 1934.3 4187.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3957.0 4000.1 0.989 0.3313   
## UNEM 1133.8 513.1 2.210 0.0358 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3049 on 27 degrees of freedom  
## Multiple R-squared: 0.1531, Adjusted R-squared: 0.1218   
## F-statistic: 4.883 on 1 and 27 DF, p-value: 0.03579

plot(m1)

 Other useful functions

# Other useful functions   
  
coefficients(m1);coef(m1) # model coefficients

## (Intercept) UNEM   
## 3957.029 1133.826

## (Intercept) UNEM   
## 3957.029 1133.826

confint(m1, level=0.95) # CIs for model parameters

## 2.5 % 97.5 %  
## (Intercept) -4250.43684 12164.495  
## UNEM 81.01336 2186.638

fitted(m1) # predicted values

## 1 2 3 4 5 6 7 8   
## 13141.02 11893.81 12233.96 12460.72 11893.81 11213.51 11326.90 11213.51   
## 9 10 11 12 13 14 15 16   
## 11100.13 12687.49 13254.40 12460.72 12347.34 13254.40 15408.67 14388.22   
## 17 18 19 20 21 22 23 24   
## 12687.49 10419.83 11326.90 12460.72 12233.96 14388.22 15408.67 12460.72   
## 25 26 27 28 29   
## 13934.69 14274.84 13934.69 12800.87 11893.81

residuals(m1) # residuals

## 1 2 3 4 5 6   
## -7640.01599 -5948.80787 -5604.95554 -4904.72065 -3177.80787 -1844.51254   
## 7 8 9 10 11 12   
## -1406.89509 -1046.51254 -16.12998 -183.48577 491.60145 1195.27935   
## 13 14 15 16 17 18   
## 1502.66190 890.60145 -520.66712 602.77589 2148.51423 4058.16536   
## 19 20 21 22 23 24   
## 3212.10491 1934.27935 2365.04446 580.77589 -301.66712 2370.27935   
## 25 26 27 28 29   
## 1146.30612 852.15845 1921.30612 3137.13168 4187.19213

anova(m1) # anova table

## Analysis of Variance Table  
##   
## Response: ROLL  
## Df Sum Sq Mean Sq F value Pr(>F)   
## UNEM 1 45407767 45407767 4.8829 0.03579 \*  
## Residuals 27 251084710 9299434   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Multiple Linear Regression

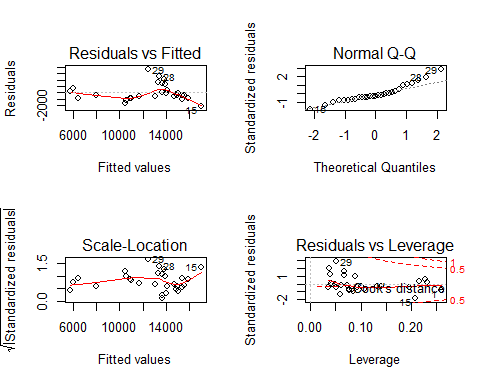
## Two predictor model

To predict the fall enrollment (ROLL) using the unemployment rate (UNEM) & number of spring high school graduates (HGRAD)

m2=lm(ROLL~UNEM+HGRAD, data=data1)  
summary(m2)

##   
## Call:  
## lm(formula = ROLL ~ UNEM + HGRAD, data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2102.2 -861.6 -349.4 374.5 3603.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -8.256e+03 2.052e+03 -4.023 0.00044 \*\*\*  
## UNEM 6.983e+02 2.244e+02 3.111 0.00449 \*\*   
## HGRAD 9.423e-01 8.613e-02 10.941 3.16e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1313 on 26 degrees of freedom  
## Multiple R-squared: 0.8489, Adjusted R-squared: 0.8373   
## F-statistic: 73.03 on 2 and 26 DF, p-value: 2.144e-11

par(mfrow=c(2,2)); plot(m2)



## Three predictor model (Full Model)

To predict the fall enrollment (ROLL) using the unemployment rate (UNEM), number of spring high school graduates (HGRAD) & per capita income (INC)

m3=lm(ROLL~UNEM+HGRAD+INC, data=data1); summary(m3)

##   
## Call:  
## lm(formula = ROLL ~ UNEM + HGRAD + INC, data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1148.84 -489.71 -1.88 387.40 1425.75   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.153e+03 1.053e+03 -8.691 5.02e-09 \*\*\*  
## UNEM 4.501e+02 1.182e+02 3.809 0.000807 \*\*\*  
## HGRAD 4.065e-01 7.602e-02 5.347 1.52e-05 \*\*\*  
## INC 4.275e+00 4.947e-01 8.642 5.59e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 670.4 on 25 degrees of freedom  
## Multiple R-squared: 0.9621, Adjusted R-squared: 0.9576   
## F-statistic: 211.5 on 3 and 25 DF, p-value: < 2.2e-16

layout(matrix(c(1,2,3,4),2,2)); plot(m3)

