ESE\_56

Jeevan ~ 1740256

30/09/2019

**AIM:**

**A wine company collected a data on various factors to study about the quality of wine on the basis of various factors. The data set is as follows:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *y* | *x*2 | *x*3 | *x*4 | *x*5 | *x*6 | *x*7 | *x*8 | *x*9 | *x*10 |
| 19.2 | 3.85 | 66 | 9.35 | 5.65 | 2.40 | 3.25 | 0.33 | 19 | 0.065 |
| 18.3 | 3.73 | 79 | 11.15 | 6.95 | 3.15 | 3.80 | 0.36 | 21 | 0.076 |
| 17.1 | 3.88 | 73 | 9.40 | 5.75 | 2.10 | 3.65 | 0.40 | 18 | 0.073 |
| 17.3 | 3.86 | 99 | 12.85 | 7.70 | 3.90 | 3.80 | 0.35 | 22 | 0.076 |
| 16.8 | 3.98 | 75 | 8.55 | 5.05 | 2.05 | 3.00 | 0.49 | 12 | 0.060 |
| 16.5 | 3.85 | 61 | 10.30 | 6.20 | 2.50 | 3.70 | 0.38 | 20 | 0.074 |
| 15.8 | 3.93 | 66 | 4.90 | 2.75 | 1.20 | 1.55 | 0.29 | 11 | 0.031 |
| 15.2 | 3.66 | 86 | 6.40 | 4.00 | 1.50 | 2.50 | 0.27 | 19 | 0.050 |
| 15.2 | 3.91 | 78 | 5.80 | 3.30 | 1.40 | 1.90 | 0.40 | 9 | 0.038 |
| 14.0 | 3.47 | 178 | 3.60 | 2.25 | 0.75 | 1.50 | 0.37 | 8 | 0.030 |
| 14.0 | 3.91 | 81 | 3.90 | 2.15 | 1.00 | 1.15 | 0.32 | 7 | 0.023 |
| 13.8 | 3.75 | 108 | 5.80 | 3.20 | 1.60 | 1.60 | 0.38 | 8 | 0.032 |
| 13.6 | 3.90 | 92 | 5.40 | 2.85 | 1.55 | 1.30 | 0.44 | 6 | 0.026 |
| 12.8 | 3.92 | 96 | 5.00 | 2.70 | 1.40 | 1.30 | 0.35 | 7 | 0.026 |
| 18.5 | 3.87 | 89 | 9.15 | 5.60 | 1.95 | 3.65 | 0.46 | 16 | 0.073 |
| 17.3 | 3.97 | 59 | 10.25 | 6.10 | 2.40 | 3.70 | 0.40 | 19 | 0.074 |
| 16.3 | 3.76 | 22 | 8.20 | 5.00 | 1.85 | 3.15 | 0.25 | 25 | 0.063 |
| 16.3 | 3.76 | 77 | 8.35 | 5.05 | 1.90 | 3.15 | 0.37 | 17 | 0.063 |
| 16.0 | 3.98 | 58 | 10.15 | 6.00 | 2.60 | 3.40 | 0.38 | 18 | 0.068 |
| 16.0 | 3.88 | 85 | 6.85 | 4.10 | 1.50 | 2.60 | 0.33 | 16 | 0.052 |
| 15.7 | 3.75 | 120 | 8.80 | 5.50 | 1.85 | 3.65 | 0.39 | 19 | 0.073 |
| 15.5 | 3.98 | 94 | 5.45 | 3.05 | 1.50 | 1.55 | 0.41 | 8 | 0.031 |
| 15.3 | 3.69 | 122 | 8.00 | 5.05 | 1.90 | 3.15 | 0.27 | 23 | 0.063 |
| 15.3 | 3.77 | 144 | 5.60 | 3.35 | 1.10 | 2.25 | 0.36 | 12 | 0.045 |
| 14.8 | 3.74 | 10 | 7.90 | 4.75 | 1.95 | 2.80 | 0.25 | 23 | 0.056 |
| 14.3 | 3.76 | 100 | 5.55 | 3.25 | 1.15 | 2.10 | 0.34 | 12 | 0.042 |
| 14.3 | 3.91 | 73 | 4.65 | 2.70 | 0.95 | 1.75 | 0.36 | 10 | 0.035 |
| 14.2 | 3.60 | 301 | 4.25 | 2.40 | 1.25 | 1.15 | 0.42 | 6 | 0.023 |
| 14.0 | 3.76 | 104 | 8.70 | 5.10 | 2.25 | 2.85 | 0.34 | 17 | 0.057 |
| 13.8 | 3.90 | 67 | 7.40 | 4.40 | 1.60 | 2.80 | 0.45 | 13 | 0.056 |
| 12.5 | 3.80 | 89 | 5.35 | 3.15 | 1.20 | 1.95 | 0.32 | 12 | 0.039 |
| 11.5 | 3.65 | 192 | 6.35 | 3.90 | 1.25 | 2.65 | 0.63 | 8 | 0.053 |

*y*: quality rating (20 maximum)

*x*2: pH

*x*3: Total *SO*2 (ppm)

*x*4: color density

*x*5: wine color

*x*6: polymeric pigment color

*x*7: anthocyanin color

*x*8: total anthocyanins (g/L)

*x*9: degree of ionization of anthocyanins (percent)

*x*10: ionized anthocyanins (percent)

Use R package to analyze the data and give your comments to the following:

1. Use Backward Elimination method to obtain the best regression model.
2. Obtain the adjusted R square.
3. Give your comment about the significance of the overall model.

**ANALYSIS:**

library(readxl)

## Warning: package 'readxl' was built under R version 3.5.2

library(GGally)

## Warning: package 'GGally' was built under R version 3.5.3

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.5.2

ESE\_data <- read\_excel("C:/Users/Jeevan/Desktop/Christ University/Statistics/Linear Regression/ESE\_data.xlsx")  
# View(ESE\_data)  
attach(ESE\_data)  
fit = lm(y~.,data = ESE\_data)  
step(fit,direction = "backward")

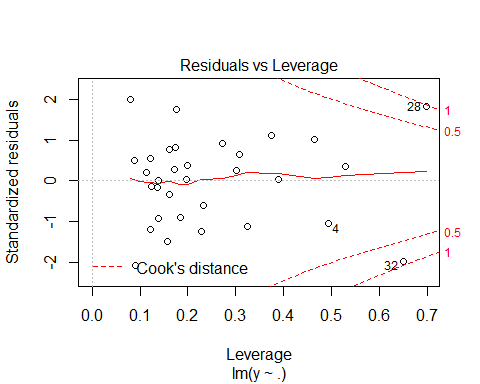
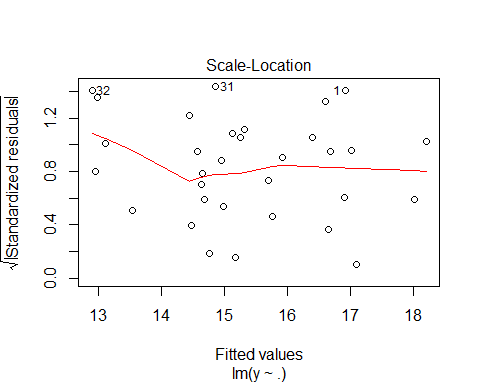
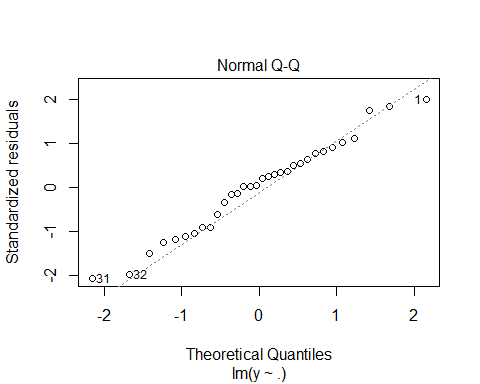
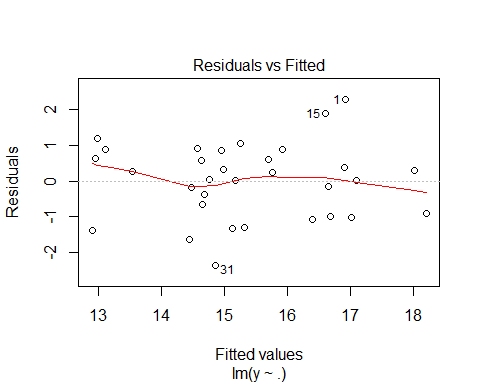
## Start: AIC=18.33  
## y ~ x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9 + x10  
##   
##   
## Step: AIC=18.33  
## y ~ x2 + x3 + x4 + x5 + x6 + x7 + x8 + x9  
##   
##   
## Step: AIC=18.33  
## y ~ x2 + x3 + x4 + x5 + x6 + x8 + x9  
##   
## Df Sum of Sq RSS AIC  
## - x6 1 0.2015 34.618 16.516  
## - x3 1 0.5004 34.917 16.791  
## <none> 34.417 18.330  
## - x9 1 2.4917 36.908 18.566  
## - x4 1 2.7006 37.117 18.747  
## - x2 1 4.4836 38.900 20.248  
## - x8 1 4.9992 39.416 20.670  
## - x5 1 6.6016 41.018 21.945  
##   
## Step: AIC=16.52  
## y ~ x2 + x3 + x4 + x5 + x8 + x9  
##   
## Df Sum of Sq RSS AIC  
## - x3 1 0.4087 35.027 14.892  
## <none> 34.618 16.516  
## - x9 1 2.5250 37.143 16.769  
## - x2 1 5.8289 40.447 19.496  
## - x4 1 6.4706 41.089 20.000  
## - x8 1 6.8089 41.427 20.262  
## - x5 1 8.4697 43.088 21.520  
##   
## Step: AIC=14.89  
## y ~ x2 + x4 + x5 + x8 + x9  
##   
## Df Sum of Sq RSS AIC  
## <none> 35.027 14.892  
## - x9 1 4.1030 39.130 16.437  
## - x8 1 6.7293 41.756 18.515  
## - x4 1 6.7524 41.779 18.533  
## - x2 1 7.1559 42.183 18.841  
## - x5 1 9.0351 44.062 20.235

##   
## Call:  
## lm(formula = y ~ x2 + x4 + x5 + x8 + x9, data = ESE\_data)  
##   
## Coefficients:  
## (Intercept) x2 x4 x5 x8   
## -2.2538 5.4205 -3.5469 7.1611 -12.4590   
## x9   
## -0.2566

summary(fit)

##   
## Call:  
## lm(formula = y ~ ., data = ESE\_data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.3594 -0.9179 0.1362 0.6940 2.2845   
##   
## Coefficients: (2 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.243486 14.099636 -0.372 0.7132   
## x2 6.146082 3.475846 1.768 0.0897 .  
## x3 0.004546 0.007696 0.591 0.5602   
## x4 -2.957382 2.155036 -1.372 0.1827   
## x5 6.579426 3.066486 2.146 0.0422 \*  
## x6 -0.663794 1.771016 -0.375 0.7111   
## x7 NA NA NA NA   
## x8 -14.488077 7.759581 -1.867 0.0741 .  
## x9 -0.260609 0.197705 -1.318 0.1999   
## x10 NA NA NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.198 on 24 degrees of freedom  
## Multiple R-squared: 0.6459, Adjusted R-squared: 0.5426   
## F-statistic: 6.254 on 7 and 24 DF, p-value: 0.0003045

plot(fit)



**INTERPRETATION:**

The analysis has been applied on the data and backward elimination has been used.

The adjusted R – squared value obtained is ***0.5426.***

Doing a summary of the model tells us that ***x5 which is wine*** color is significant as it’s p – value is less than 0.05. This tells us that we should reject our null hypothesis and accept the alternate hypothesis which tells us that there is a significant difference between the color of wine and the other variables.

***------------------------------------------------------------------------------------------------------***