Stat*3240, Assignment
Graham Eckel
0679576

Q2

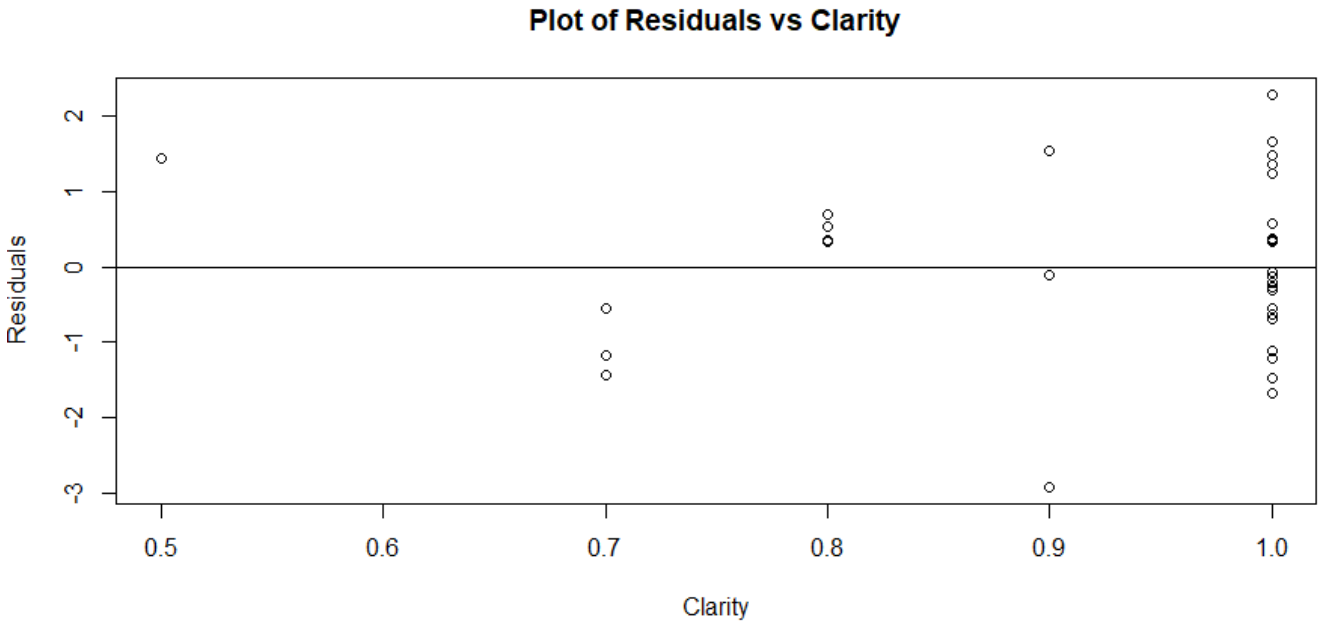
```
set.seed(2019-09-27)
dir = "E:\\...\\Applied Regression Analysis\\"
file1 = "3240_F19_wine.csv"
dfWine = read.table(file=paste(dir,file1, sep=""), header=TRUE, sep=',')

Clarity = dfWine$Clarity
Aroma = dfWine$Aroma
Body = dfWine$Body
Flavour = dfWine$Flavor
Oakiness = dfWine$Oakiness
Quality = dfWine$Quality

mlrWine = lm(Quality~Clarity+Aroma+Body+Flavour+Oakiness)
summary(mlrWine)
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-2.9315 -0.6519 -0.0888  0.6135  2.2931

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.2133     2.7308   1.177  0.24997
Clarity       2.4846     2.0243   1.227  0.23069
Aroma         0.4775     0.3055   1.563  0.13015
Body          0.4089     0.3844   1.064  0.29722
Flavour       1.1543     0.3543   3.257  0.00312 **
Oakiness     -0.6502     0.3125  -2.081  0.04747 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.259 on 26 degrees of freedom
Multiple R-squared:  0.6891,    Adjusted R-squared:  0.6293
F-statistic: 11.53 on 5 and 26 DF,  p-value: 6.148e-06
```

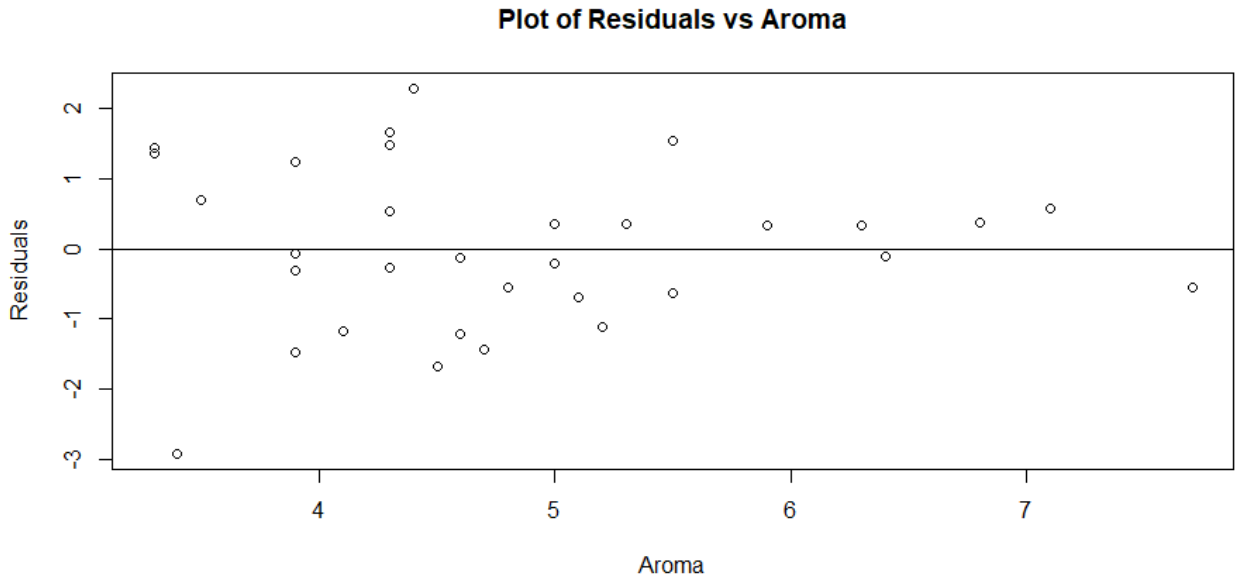
a)
There is a positive linear relationship between the Aroma of the wine and the Quality of the wine. Holding all other variables constant, for every one unit increase in the rating of the wine’s aroma, there will be a 0.4775 unit increase in the rating of the quality of the wine.

b)

```
resWine = resid(mlrWine)
resWineClarity = plot(Clarity, resWine, ylab = "Residuals", main = "Plot of Residuals vs Clarity")
abline(0,0)
resWineAroma = plot(Aroma, resWine, ylab = "Residuals", main = "Plot of Residuals vs Aroma")
abline(0,0)
resWineBody = plot(Body, resWine, ylab = "Residuals", main = "Plot of Residuals vs Body")
abline(0,0)
resWineFlavour = plot(Flavour, resWine, ylab = "Residuals", main = "Plot of Residuals vs
Flavour")
abline(0,0)
resWineOakiness = plot(Oakiness, resWine, ylab = "Residuals", main = "Plot of Residuals vs
Oakiness")
abline(0,0)
```

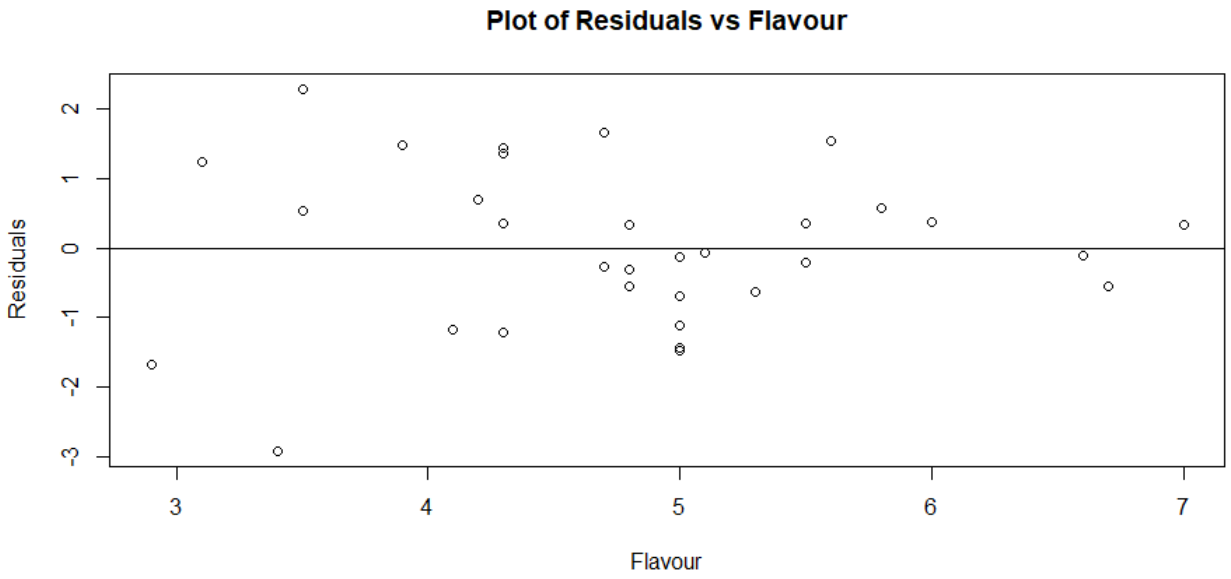


Slight indication of a problem with increasing variance.

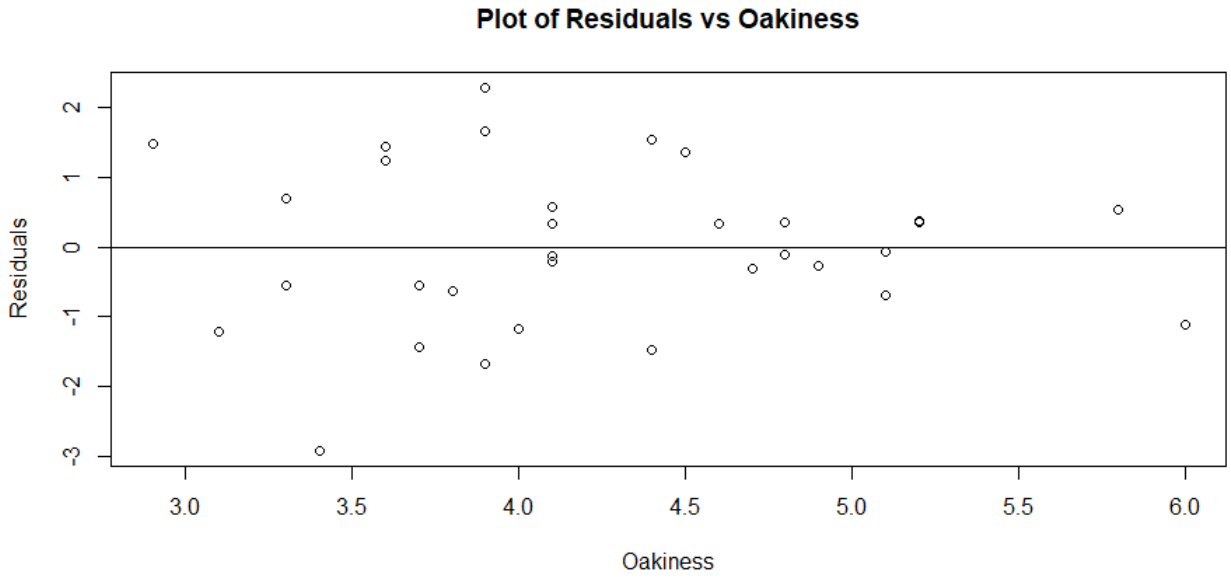


Indication of a problem with decreasing variance.

No problems.



No problems.



Slight indication of a problem with decreasing variance.

c)

anova(mlrWine)
qf(0.975, df1 = 5, df2 = 26) = 3.1048

```
Analysis of Variance Table

Response: Quality
           Df Sum Sq Mean Sq F value   Pr(>F)
Clarity     1  0.000   0.000  0.0000 0.996444
Aroma       1 64.061  64.061 40.4448  9.8e-07 ***
Body        1  5.669   5.669  3.5790 0.069703 .
Flavour     1 14.704  14.704  9.2835 0.005251 **
Oakiness    1  6.856   6.856  4.3287 0.047468 *
Residuals  26 41.182   1.584
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

We are testing to see if any of the variables are linearly related to Quality. So, we set this as our null hypothesis.

Null hypothesis: none of the variables are linearly related to Quality
Ho: B1 = B2 = B3 = B3 = B4 = B5 = 0

Alternative Hypothesis:
Ha: At least one independent variable =/= 0

Looking at our independent variables, we see that Aroma, Flavour and Oakiness have a p-value less than our significance level of 5% and large enough F test statistics. Particularly in the case of Aroma.There exists a linear relationship between the Quality of the wine and the Aroma, Flavour and Oakiness of the wine. As such, we have evidence to reject the null hypothesis. It can be said that as these independent variables increase - Aroma, Flavour and Oakiness, the Quality of the wine tends to increase as well.

d)

summary(mlrWine)
qt(0.975, df = 26) = 2.0555

```
Residuals:
    Min      1Q  Median      3Q     Max
-2.9315 -0.6519 -0.0888  0.6135  2.2931

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.2133     2.7308   1.177  0.24997
Clarity       2.4846     2.0243   1.227  0.23069
Aroma         0.4775     0.3055   1.563  0.13015
Body          0.4089     0.3844   1.064  0.29722
Flavour       1.1543     0.3543   3.257  0.00312 **
Oakiness     -0.6502     0.3125  -2.081  0.04747 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.259 on 26 degrees of freedom
Multiple R-squared:  0.6891,    Adjusted R-squared:  0.6293
F-statistic: 11.53 on 5 and 26 DF,  p-value: 6.148e-06
```

We are testing to see if there exists a linear relationship between the Quality of the wine and the aroma of the wine after all Clarity, Body, Flavour and Oakiness are included in the linear model.

Null Hypothesis: After including all the independent variables in our linear model, there does not exist a linear relationship between Aroma and Quality.
Ho: B1 = 0

Alternative Hypothesis:
Ha: B1 =/= 0

At a 5% significance, our threshold t-value for a two tailed test with 26 degrees of freedom is 2.0555. We see that Aroma has a p value larger than our significance. Aroma also has a t value less than our threshold value. In conclusion, there is evidence to support the null hypothesis that Aroma and Quality are not linearly related after all the independent variables are included in our model.