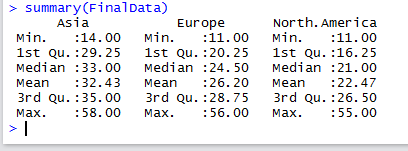
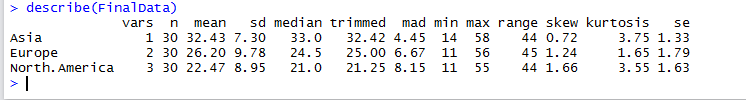
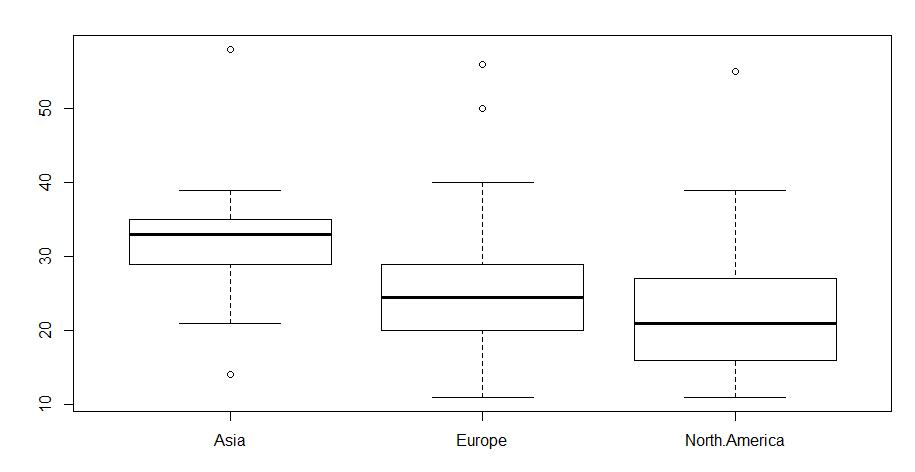
ISQS 5347 – Advanced Statistical Methods

Final Exam

1. A transportation strategist wanted to compare the traffic congestion levels across three continents: Asia, Europe, and North America. The accompanying Excel file contains the congestion levels for 30 cities in each continent (data extracted from the Tom Tom Traffic Index, www.tomtom.com). Congestion level is measured as the percent increase in overall travel time when compared to an uncongested situation. (15 points)
   1. Run summary statistics for each of the continents and create a box plot for each. Do the data look like they could have come from a normal distribution?





Boxplots for all three vehicle types:

After seeing the box plots, it is reasonable to believe that these data could have been drawn from a normal distribution

* 1. Regardless of your assessment in part a, conduct an ANOVA test at the 0.05 level of significance whether the means of traffic congestion are equal in all three continents.

An ANOVA test was conducted to compare mean chest compression for the three continents.

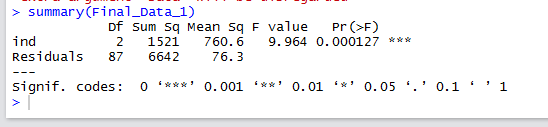
Hypotheses: H0: µ1 = µ2 = µ3

H0: Not all µj are equal

Decision rule: Reject H0 if FSTAT > FCRIT

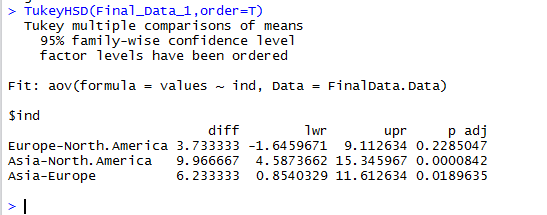
Reject H0 if p-value < α



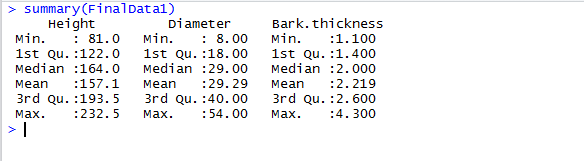


The FSTAT for the fitted model is 9.964 with a p-value of 0.00127. Therefore, I will reject the null hypothesis. There is sufficient evidence to conclude that congestion levels differ between the three continents.

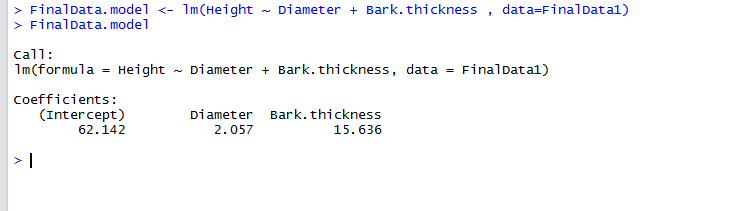
* 1. If your ANOVA results cause you to reject the null hypothesis, conduct the Tukey HSD procedure at the 0.05 level of significance to determine which mean comparisons differ.



1. Measuring the height of a California redwood tree is very difficult because they may grow to heights of over 300 feet. Botanists familiar with these trees theorize that the height of a California redwood tree is related to other characteristics of a tree, including the diameter of the tree at the chest height of a person (in inches) and the thickness of the bark of the tree (in inches). The accompanying file contains height, diameter at chest height, and bark thickness for a sample of 21 California redwood trees. (15 points)



* 1. State the multiple regression equation that predicts the height of a tree based on the tree’s diameter and bark thickness.



* 1. Interpret the meaning of the slopes in this equation.

Yi = C + M1 X1 + M2 X2

C is the intercept

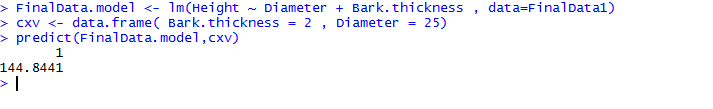
M1 is the slope of the regression line

M2 is the slope of the regression line

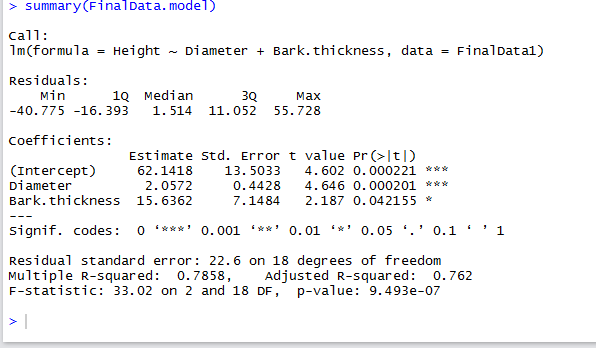
X1 is the value for X for observation 1

X2 is the value for X for observation 2

* 1. Predict the mean height for a tree that has a chest height diameter of 25 inches and a bark thickness of 2 inches.



* 1. Interpret the meaning of the adjusted r2 in this problem.



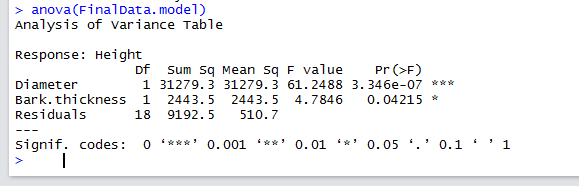
The adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model, basically adjusted R square improves the model. Generally adjusted R square is lower than that of the R square.

* 1. Conduct an F Test (α=0.05) to determine the significance of the regression equation.

F test for the Slope: -

Null Hypothesis: - Slope of the line is 0, i.e. No connection between dependent and Independent variables.

Alternate Hypothesis: - Slope of the line is not equal to 0, i.e. Connection between dependent and Independent variables.



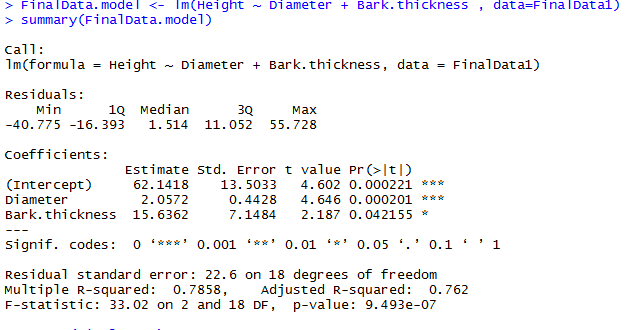


So, Conclusion states that there is evidence that the height of the tree is effected by the Diameter and Bark Thickness.

* 1. Conduct t-tests (α=0.05) to evaluate the significance of each of the independent variables in the model.

H0: βj = 0 (no linear relationship between Xi and Y)

H1: βj ≠ 0 (linear relationship does exist between Xi and Y)





α = .05

The test statistic for each variable falls in the rejection region (p-values < .05)

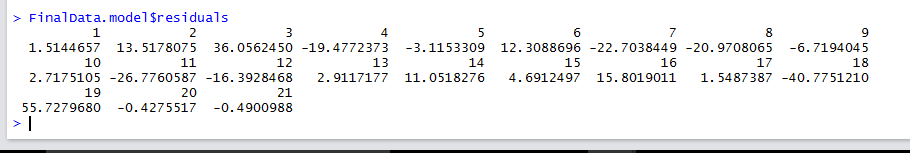
Also

or thickness = tstat = 2.187 and diameter = tstat = 4.646 From the above tstat and t α/2

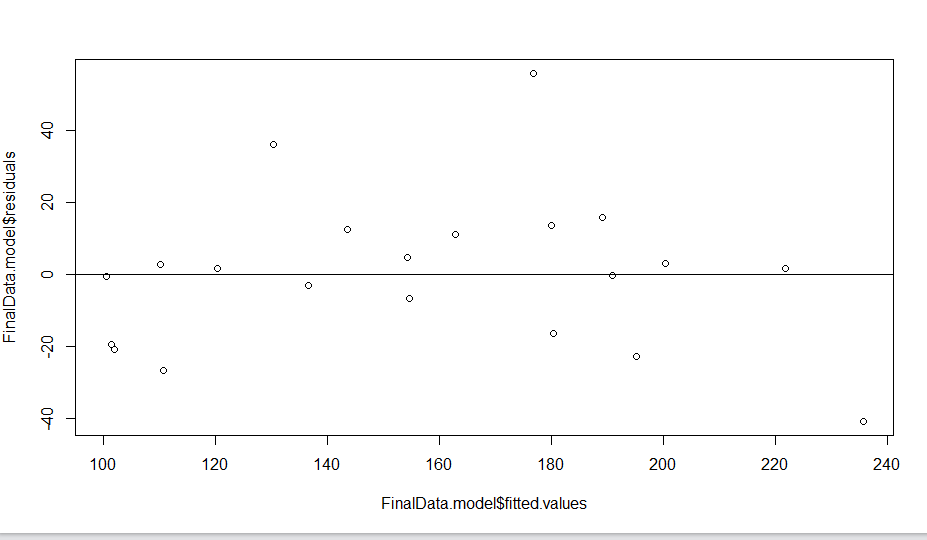
Conclusion: -

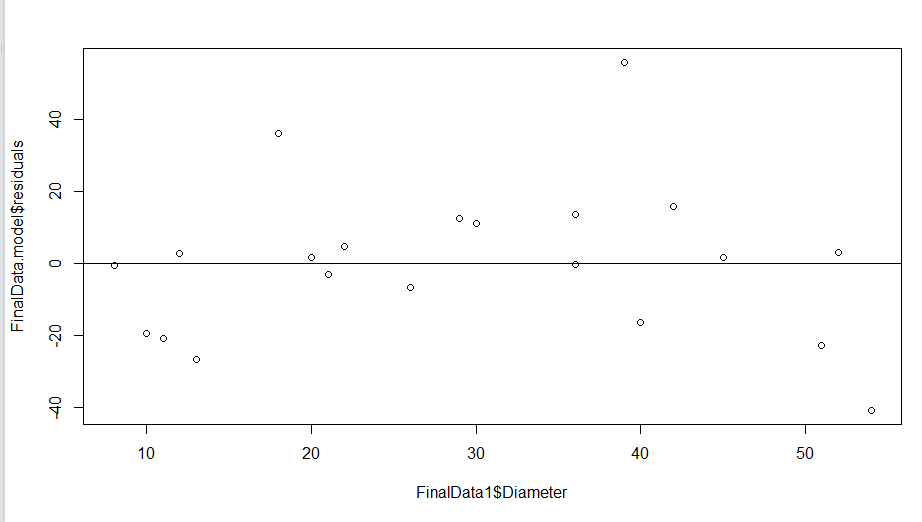
There is evidence that both Diameter and Bark thickness effect the Height at α = .05.

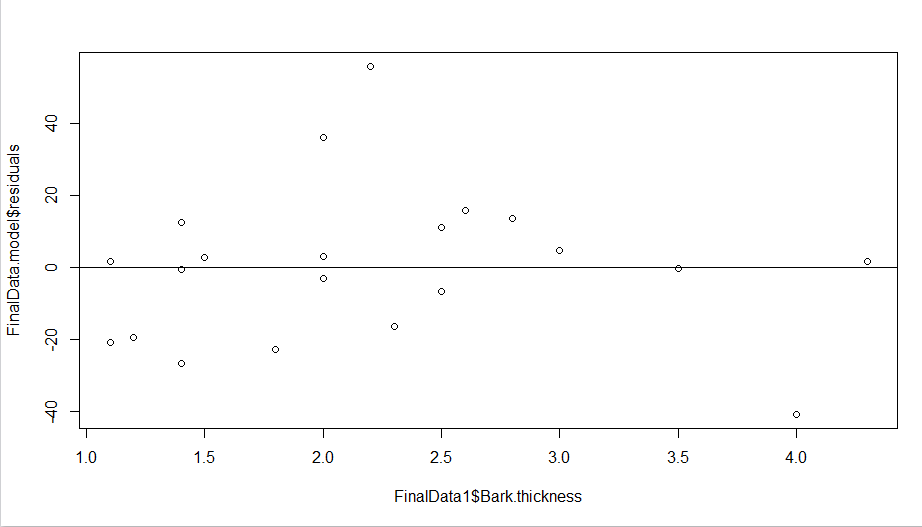
* 1. Perform a residual analysis on the results and determine whether the regression assumptions are valid.

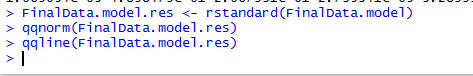


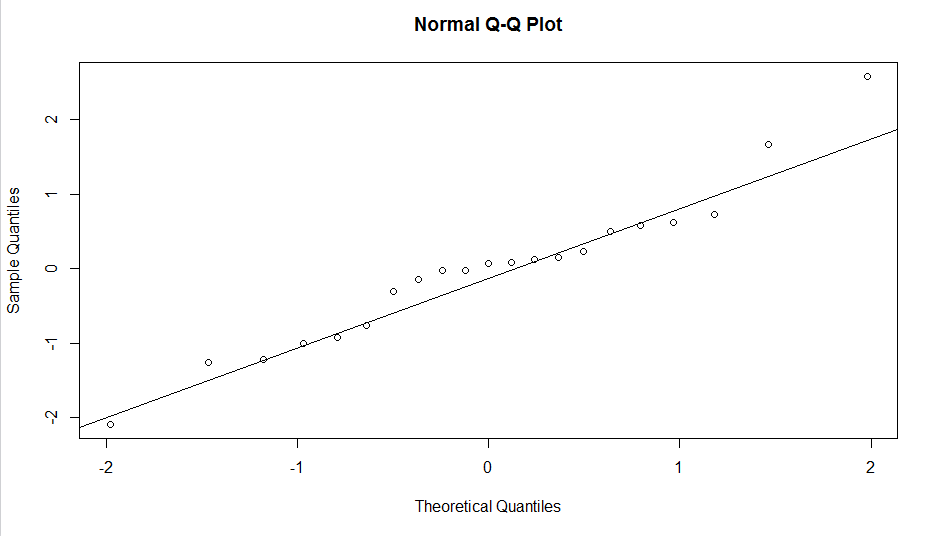
The majority of residual values fall within ±20 units of the center line across the width of the plots. Therefore, violations of the constant variance assumption are likely not severe. A QQ plot of the residuals is also shown. While the data points do not lie on a straight line, there is not a pattern indicating excessive departures from normality







r



Residual Analysis for Linearity: - It is Linear

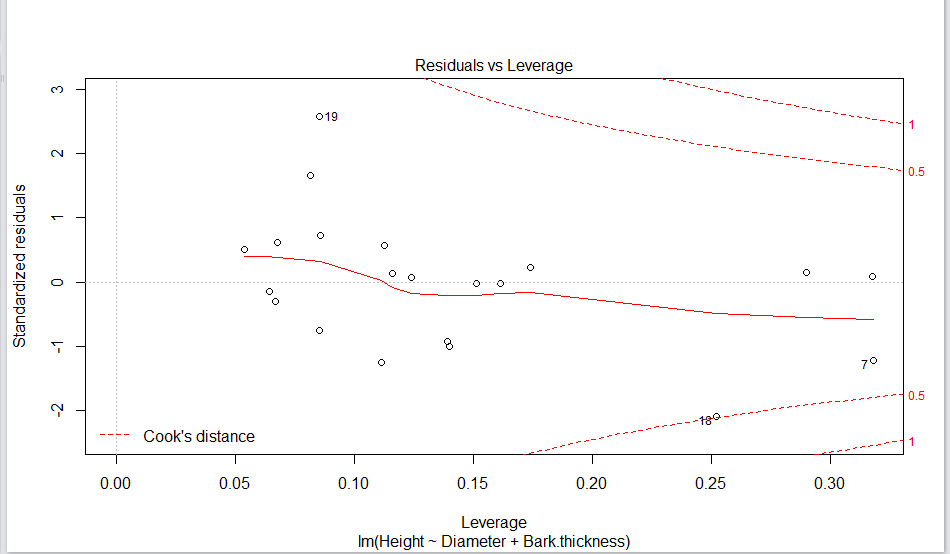
Residual Analysis for Independence: - No Cyclical Pattern Independent

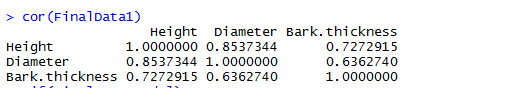
Residual Analysis for Equal Variance: - Constant variance

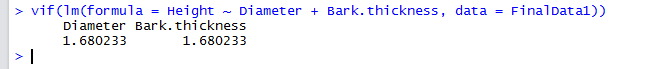
* 1. Evaluate the model for potential outliers and multicollinearity.

potential outliers: -

The leverage plot of Cook’s distance for the data show a potential outlier. Observation #18 that is close to 0.5. All other points have Cook’s distance within 0.5.







Since the regression coefficients are reasonable and VIFs are low, it appears that the independent variables are explaining enough separate variance in the dependent variable so that the high correlation between the independent variables does not cause excessive multicollinearity.