STAT 511: HW #5 Q:3 & 4

Rumil Legaspi

5 April 2021

Multiple Regression & Brand Preference Dataset

Setting up workspace

```
library(nortest)
library(olsrr)
library(car)
library(lmtest)
library(MASS)
library(tidyverse)

setwd("C:/Users/RUMIL/Desktop/APU/STAT 511 - Millie Mao (Applied Regression Analysis)/Week 10/Week 10")

brand_data = read.table(file = "Brand.txt", header = FALSE, sep = "")

View(brand_data)

# #Adding headers
names(brand_data) <- c("Rating", "Moisture", "Sweetness")

# names(bank_data) <- c("", "")

#Defining dependent and independent vars
Rating = brand_data$Rating #Y
Moisture = brand_data$Moisture #X1
Sweetness = brand_data$Sweetness #X2
```

```
#Regressing Rating (response) on Moisture (explanatory) and Sweetness (explanatory).
#Then summarizing our model
brand_lm <- lm(Rating ~ Moisture + Sweetness, data = brand_data)
summary(brand_lm)
```

```
##
## Call:
## lm(formula = Rating ~ Moisture + Sweetness, data = brand_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.400 -1.762  0.025  1.587  4.200
```

```
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.6500     2.9961  12.566 1.20e-08 ***
## Moisture      4.4250     0.3011  14.695 1.78e-09 ***
## Sweetness     4.3750     0.6733   6.498 2.01e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.693 on 13 degrees of freedom
## Multiple R-squared:  0.9521, Adjusted R-squared:  0.9447
## F-statistic: 129.1 on 2 and 13 DF,  p-value: 2.658e-09
```

3a. What is the fitted value of the degree of brand liking when moisture content is at 5 and sweetness is at 4?

```r
new_brand_data <- data.frame(Moisture = 5, Sweetness = 4)

ci_brand <- predict(brand_lm, new_brand_data, interval = "confidence", level = 0.95)

pi_brand <- predict(brand_lm, new_brand_data, interval = "prediction", level = 0.95)

ci_brand
```

```
##      fit      lwr      upr
## 1 77.275 74.84094 79.70906
```

```r
pi_brand
```

```
##      fit      lwr      upr
## 1 77.275 70.96788 83.58212
```

The following results show us that the when moisture content is at 5 and sweetness level is at 4, our fitted value will be 77.275.

3b. Obtain a 95% interval estimate of the average degree of brand liking when moisture content

is at 5 and sweetness is at 4. Interpret this interval.

```r
#Printing ci_brand
ci_brand
```

```
##      fit      lwr      upr
## 1 77.275 74.84094 79.70906
```

Confidence Interval Interpretation when Moisture = 5 and Sweetness = 4:

This 95% confidence interval when Moisture = 5 and Sweetness = 4 **is from 74.84094 to 79.70906.**

When the Moisture = 5 and Sweetness = 4, with 95% confidence we can expect our confidence interval to capture the **average(true mean)** of Rating (response variable).

3c. Obtain a 95% interval estimate of a future degree of brand liking when moisture content is at

5 and sweetness is at 4. Interpret this interval

```
pi_brand
```

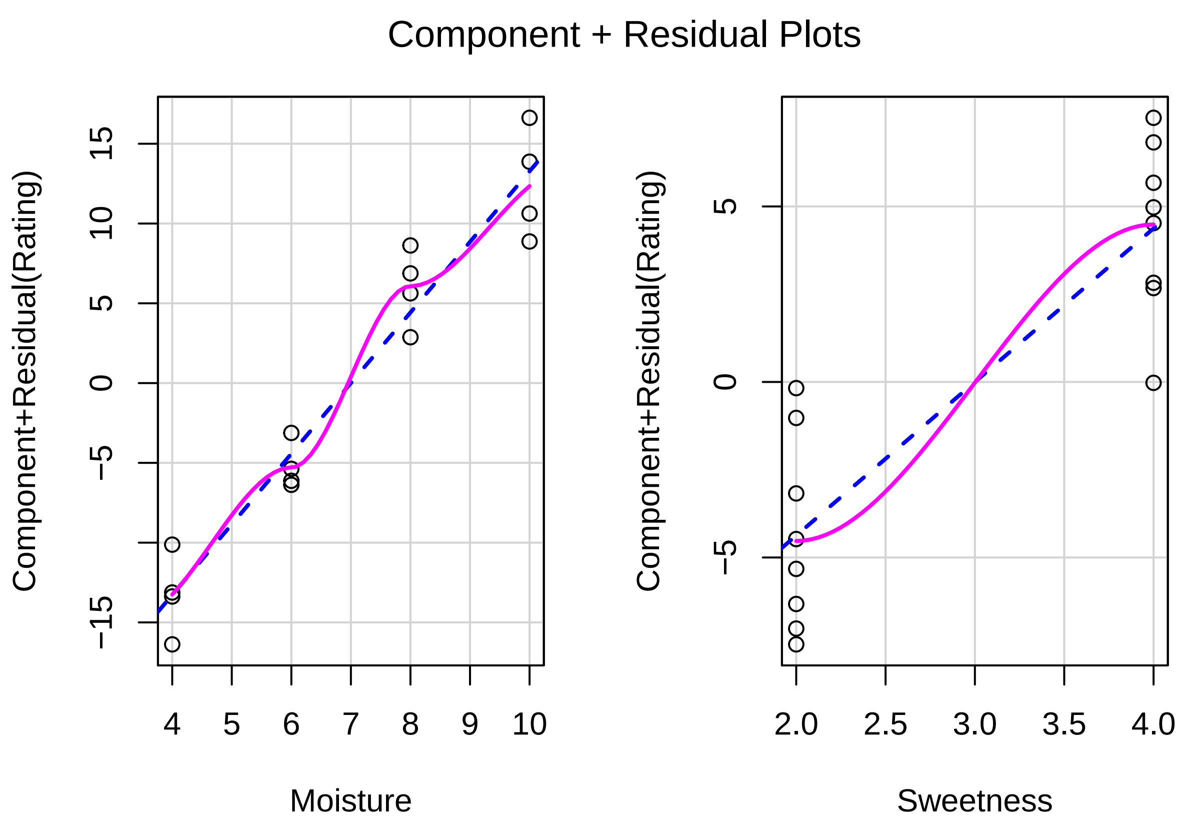
```
##      fit      lwr      upr
## 1 77.275 70.96788 83.58212
```

From the results we can predict with 95% confidence that when the moisture content of the product is at 5 and the sweetness level is 4, the future rating will fall somewhere between 70.96 and 83.58. In other words, somewhere between a rating of roughly 71% to 84%.

4. Analyzing the residuals from the MLR in Question 1

4a. Use the "Component Plus Residual" (CPR) plots to conclude on linearity assumption.

```
car::crPlots(brand_lm)
```

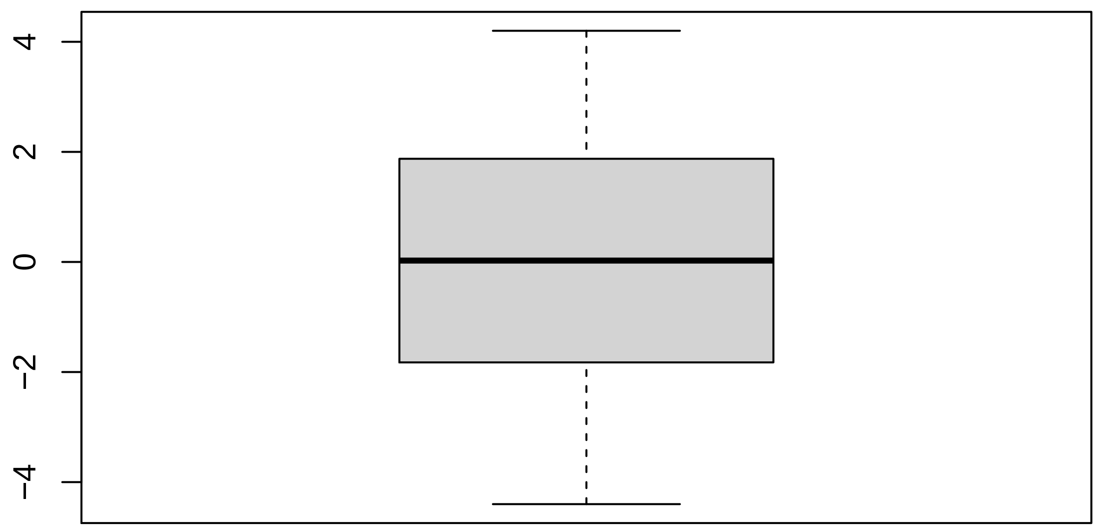


From these plots, we can say that the smoothed pink lines don't seem to be straying to far from the blue dashed lines indicating that the these plots are **approximately linear** and thereby satisfying the linearity assumption.

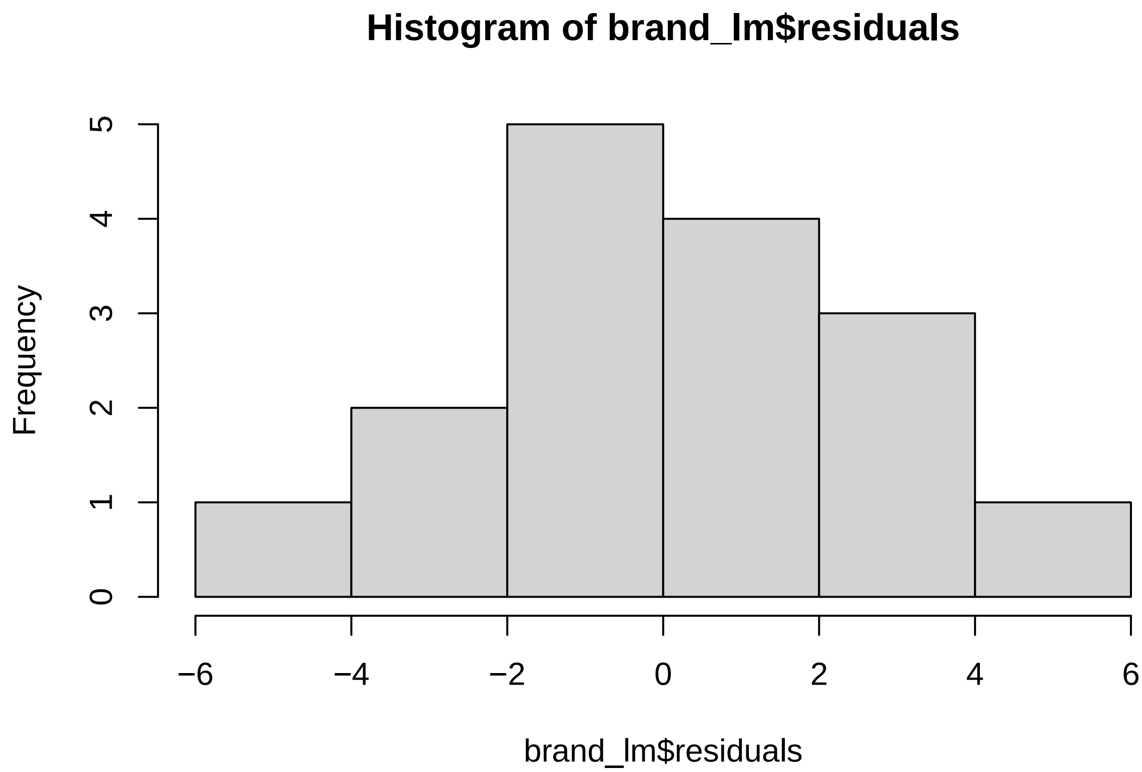
4b. Draw boxplot and histogram of residuals to conclude on normality assumption.

```r
#Plotting a boxplot and histogram

boxplot(brand_lm$residuals)
```




```r
hist(brand_lm$residuals)
```



After looking at these plots we see two things: - The boxplot is symmetrical and do not see any outliers - The histograms fairly resembles a normal distribution for our residuals. Knowing these, we can conclude our assumption of normality is not violated.

4c. Conduct the Breusch-Pagan test to check if the equal variance assumption is satisfied.

```
bptest(brand_lm, studentize = FALSE)
```

```
##
##  Breusch-Pagan test
##
## data:  brand_lm
## BP = 1.0422, df = 2, p-value = 0.5939
```

Null Hypothesis: Equal variance assumption is satisfied

Alternative Hypothesis: Equal variance assumption is violated

The Breusch-Pagan test gives us a high p value which means we fail reject the null hypothesis and that there is **no** issue with our equal variance assumption.