STAT 511: Assignment #6

Rumil Legaspi

4/10/2021

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
```

a. Fit a standardized multiple regression model where all variables are centered and scaled.

```
#Scaling coefficients
scaled_Rating <- scale(Rating)
scaled_Moisture <- scale(Moisture)
scaled_Sweetness<-scale(Sweetness)

#putting scaled coefficients into a lm, now the results are scaled,
scaled_lm <- lm(scaled_Rating ~ 0 + scaled_Moisture + scaled_Sweetness, data = brand_data)
scaled_lm
```

```
##
## Call:
## lm(formula = scaled_Rating ~ 0 + scaled_Moisture + scaled_Sweetness,
##     data = brand_data)
##
## Coefficients:
##  scaled_Moisture  scaled_Sweetness
##           0.8924            0.3946
```

```
summary(scaled_lm)
```

```
##
## Call:
## lm(formula = scaled_Rating ~ 0 + scaled_Moisture + scaled_Sweetness,
##     data = brand_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38423 -0.15391  0.00218  0.13863  0.36677
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## scaled_Moisture   0.89239    0.05852  15.250 4.09e-10 ***
## scaled_Sweetness  0.39458    0.05852   6.743 9.43e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 0.2266 on 14 degrees of freedom
## Multiple R-squared:  0.9521, Adjusted R-squared:  0.9452
## F-statistic:   139 on 2 and 14 DF,  p-value: 5.82e-10
```

The estimated intercept coefficient will be zero if all standardized, so we can remove it from the model.

b. Interpret the partial slope coefficient $\hat{\beta}_1$ in the standardized regression model.

c. Find the correlation matrix of this dataset. Is there any multicollinearity issue?

d. Use the anova() function in R to test if sweetness (X_2) should be removed from the multiple

linear regression, i.e., test the difference between the full model and the reduced model.