STAT 511: HW #7

Rumil Legaspi

4/19/2021

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
```

1a. Regress degree of brand liking on sweetness only. Write down the estimated regression model.

```
#Coding Sweetness as a dummy variable
cat_sweetness <- as.factor(Sweetness)

#Regressing Rating on new Sweetness dummy variable
Sweetness_only <- lm(Rating ~ cat_sweetness, data = brand_data)

summary(Sweetness_only)
```

```
##
## Call:
## lm(formula = Rating ~ cat_sweetness, data = brand_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.375  -7.312  -0.125   8.688  16.625
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      77.375      3.851  20.094 1.01e-11 ***
## cat_sweetness4    8.750      5.446   1.607     0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.89 on 14 degrees of freedom
## Multiple R-squared:  0.1557, Adjusted R-squared:  0.09539
## F-statistic: 2.582 on 1 and 14 DF,  p-value: 0.1304
```

Estimated Regression model:

$\hat{Y} = 77.375 + 8.750X$

```
#Checking how many levels there are in sweetness as a dummy
levels(cat_sweetness)
```

```
## [1] "2" "4"
```

```
typeof(cat_sweetness)
```

```
## [1] "integer"
```

```
attributes(cat_sweetness)
```

```
## $levels
## [1] "2" "4"
##
## $class
## [1] "factor"
```

So in this case since sweetness is a category with only 2 level (sweetness level 2 and sweetness level 4) we can think of our regression model like so:

We will use sweetness level 2 as our reference group

```
#the "2" indicates the level marked as "2" given from our level() function and now setting sweetness le
cat_sweetness_new <- relevel(cat_sweetness, ref = "2")

#Regressing
Sweetness_only2 <- lm(Rating ~ cat_sweetness_new, data = brand_data)

summary(Sweetness_only2)
```

```
##
## Call:
## lm(formula = Rating ~ cat_sweetness_new, data = brand_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.375  -7.312  -0.125   8.688  16.625
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)           77.375      3.851  20.094 1.01e-11 ***
## cat_sweetness_new4     8.750      5.446   1.607     0.13
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.89 on 14 degrees of freedom
## Multiple R-squared:  0.1557, Adjusted R-squared:  0.09539
## F-statistic: 2.582 on 1 and 14 DF,  p-value: 0.1304
```

After doing so we can now say

Table 1: Mean Rating at Level 2 or Level 4 Sweetness

$\hat{Y} = B_0 + B_1 X_1$	Sweetness = Level 2	Sweetness = Level 4
$\hat{Y} = 77.375 + 8.750X$	$\hat{Y} = 77.375 + 8.750(0)$ $= 77.375$ 77.375	$\hat{Y} = 77.375 + 8.750(1)$ $= 77.375 + 8.750$ 86.125

```
warnings()
```