**STAT 462 – Applied Regression Analysis**

**Fall 2017, Lab 11**

Prepare a short report with relevant output, your comments, and answers to the questions (this does not need to be exhaustive or polished, but should contain enough to show that you completed all tasks and analyses).

Submit the report at the end of the lab session.

Consider the dataset *greenhouse.txt*, that contains data about the effect of different planting methods on potato crop yields. In particular, you will consider the following variables:

*weight:* Total weight of tubers (potatoes)

*method:* Planting method used (hydroponic, aeroponic, pots and plant beds). Hydroponic is a method that doesn't involve soil, but uses some water solution as the growing medium. Aeroponic is a method that grows the plants directly in the air or some mist, without any growing medium.

*variety:* Variety of potatos (costanera, mariva and unica)

* Fit a linear regression model with outcome the total weight of potatoes. As predictor, consider the planting method used, after constructing a factor variable in R so that the level “bed” will be used as baseline.
  + Produce the summary of the fitted regression model.

> summary(lm\_fit)

Call:

lm(formula = weight ~ cat)

Residuals:

Min 1Q Median 3Q Max

-97.967 -27.842 -2.955 21.942 248.615

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 84.477 3.950 21.386 < 2e-16 \*\*\*

cathydroponic -62.254 5.586 -11.144 < 2e-16 \*\*\*

cataeroponic -9.792 5.586 -1.753 0.0803 .

catpot 24.790 5.586 4.438 1.13e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.27 on 476 degrees of freedom

Multiple R-squared: 0.3516, Adjusted R-squared: 0.3475

F-statistic: 86.05 on 3 and 476 DF, p-value: < 2.2e-16

* + What is the average weight for potatoes grown with hydroponic method?

**84.477-62.254=22.223**

* + Which planting method has an average weight significantly different from the baseline “bed”, at level 5%? What test are you using to answer to this question?

**We are using student t test with degree of freedom 476.**

**We observe that hydroponic method and pot method have an average weight significantly different from the baseline “bed” since the p-values of those method are smaller than α = 0.05.**

* + Does the planting method have any effect on weight, at level 5%? What test are you using to answer to this question?

**I am using F-test here.**

**Since the p-value < 2.2 × 10-16 < α = 0.05, we can reject the hull hypothesis and conclude that the planting method has effect on weight.**

* Now fit a linear regression model considering both the planting method and the variety of potatoes as categorical predictors. For the variety, use “Costanera” as a baseline. Do not consider any interaction between the two variables.
  + Produce the summary of the fitted regression model.

> summary(lm\_fit2)

Call:

lm(formula = weight ~ cat + cat2)

Residuals:

Min 1Q Median 3Q Max

-94.921 -28.158 -3.656 21.047 246.624

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 85.5333 4.8418 17.666 < 2e-16 \*\*\*

cathydroponic -62.2538 5.5908 -11.135 < 2e-16 \*\*\*

cataeroponic -9.7921 5.5908 -1.751 0.0805 .

catpot 24.7900 5.5908 4.434 1.15e-05 \*\*\*

cat2Unica -4.1019 4.8418 -0.847 0.3973

cat2Mariva 0.9344 4.8418 0.193 0.8471

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 43.31 on 474 degrees of freedom

Multiple R-squared: 0.3533, Adjusted R-squared: 0.3465

F-statistic: 51.79 on 5 and 474 DF, p-value: < 2.2e-16

* + What is the average weight for “Costanera” potatoes grown with hydroponic method?

**85.5333-62.2538=23.2795**

* + What is the average weight for “Unica” potatoes grown with hydroponic method?

**85.5333-4.1019=81.4314**

* + Does the variety of potatoes have any effect on weight, at level 5%, when the model takes into consideration the planting method (in other words, is the model with both predictors better than the model that only considers the planting method)? Write null and alternative hypothesis tested, distribution of the test statistics under the null hypothesis and p-value of the test.

> anova(lm\_fit2)

Analysis of Variance Table

Response: weight

Df Sum Sq Mean Sq F value Pr(>F)

cat 3 483322 161107 85.9050 <2e-16 \*\*\*

cat2 2 2297 1148 0.6123 0.5425

Residuals 474 888945 1875

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**H0: variety of potatoes have no effect on weight.**

**H1: variety of potatoes have significant effect on weight.**

**Test-statistic = F test = F2,474 = 0.6123**

**p-value=0.5425 > α =0.05**

**Thus, we fail to reject null hypothesis and conclude that variety of potatoes does not have effect on weight.**

* Finally, fit a linear regression model considering both categorical predictors and their interaction.
  + Produce the summary of the fitted regression model.

> summary(lm\_fit3)

Call:

lm(formula = weight ~ cat + cat2 + cat:cat2)

Residuals:

Min 1Q Median 3Q Max

-88.548 -17.871 -3.492 15.626 227.944

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 87.548 6.280 13.940 < 2e-16 \*\*\*

cathydroponic -61.413 8.881 -6.915 1.54e-11 \*\*\*

cataeroponic -47.395 8.881 -5.336 1.48e-07 \*\*\*

catpot 53.495 8.881 6.023 3.46e-09 \*\*\*

cat2Unica -18.465 8.881 -2.079 0.038158 \*

cat2Mariva 9.255 8.881 1.042 0.297925

cathydroponic:cat2Unica 18.615 12.560 1.482 0.139001

cataeroponic:cat2Unica 66.860 12.560 5.323 1.59e-07 \*\*\*

catpot:cat2Unica -28.022 12.560 -2.231 0.026152 \*

cathydroponic:cat2Mariva -21.139 12.560 -1.683 0.093047 .

cataeroponic:cat2Mariva 45.949 12.560 3.658 0.000283 \*\*\*

catpot:cat2Mariva -58.092 12.560 -4.625 4.85e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 39.72 on 468 degrees of freedom

Multiple R-squared: 0.4629, Adjusted R-squared: 0.4502

F-statistic: 36.66 on 11 and 468 DF, p-value: < 2.2e-16

* + What is the average weight for “Costanera” potatoes grown with hydroponic method?

**87.548-61.413=26.135**

* + What is the average weight for “Unica” potatoes grown with hydroponic method?

**87.548-61.413-18.465+18.615=26.285**

* + Is the interaction between the variety of potatoes and the planting method significant in this last model? Write null and alternative hypothesis tested, distribution of the test statistics under the null hypothesis and p-value of the test.

> anova(lm\_fit3)

Analysis of Variance Table

Response: weight

Df Sum Sq Mean Sq F value Pr(>F)

cat 3 483322 161107 102.1204 <2e-16 \*\*\*

cat2 2 2297 1148 0.7279 0.4835

cat:cat2 6 150619 25103 15.9120 <2e-16 \*\*\*

Residuals 468 738326 1578

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

**H0: interaction between variety of potatoes and planting method has no effect on weight.**

**H1: interaction between variety of potatoes and planting method has significant effect on weight.**

**Test-statistic = F test = F6,468 = 15.9120**

**p-value < 2 × 10-16 < α =0.05**

**Thus, we reject null hypothesis and conclude that interaction between variety of potatoes and planting method has significant effect on weight.**

R code:

setwd("//udrive.win.psu.edu/Users/j/q/jql5883/Desktop/math462")

getwd()

greenhouse=read.csv("greenhouse.txt",header=T,sep="")

attach(greenhouse)

cat=factor(method,level=c("bed","hydroponic","aeroponic","pot"))

lm\_fit=lm(weight~cat)

summary(lm\_fit)

cat2=factor(variety,level=c("Costanera","Unica","Mariva"))

lm\_fit2=lm(weight~cat+cat2)

summary(lm\_fit2)

anova(lm\_fit2)

lm\_fit3=lm(weight~cat+cat2+cat:cat2)

summary(lm\_fit3)

anova(lm\_fit3)