MATH 588

HW5

Md Ismail Hossain

3/06/2022

Question 1

a

```
library(Sleuth3)
pyg = case1302
# Number of rows and columns
dim(pyg)
```

```
## [1] 29  3
```

```
# Head of the data
head(pyg)
```

```
##   Company     Treat Score
## 1      C1 Pygmalion  80.0
## 2      C1   Control  63.2
## 3      C1   Control  69.2
## 4      C2 Pygmalion  83.9
## 5      C2   Control  63.1
## 6      C2   Control  81.5
```

```
names(pyg) <- tolower(names(pyg))
summary(aov(score~company*treat,pyg))
```

```
##               Df Sum Sq Mean Sq F value Pr(>F)
## company        9  671.0    74.6   1.437 0.2990
## treat          1  338.9   338.9   6.530 0.0309 *
## company:treat  9  311.5    34.6   0.667 0.7221
## Residuals      9  467.0    51.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b

```
summary(aov(score~company,pyg))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## company      9    671   74.55   1.268  0.315
## Residuals   19   1117   58.81
```

c

```
summary(aov(score~company+treat,pyg))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## company      9  671.0    74.6   1.724 0.1556
## treat        1  338.9   338.9   7.835 0.0119 *
## Residuals   18  778.5    43.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

μ_0 : There is no treatment effect.
μ_1 : Treatment effect is significant.

At 5% level of significance we can reject the null hypothesis because the calculated p-value is 0.0119. So, treatment effect is statistically significant.

d

$MS_R = 43.3$ is the best estimate of the residual.

e

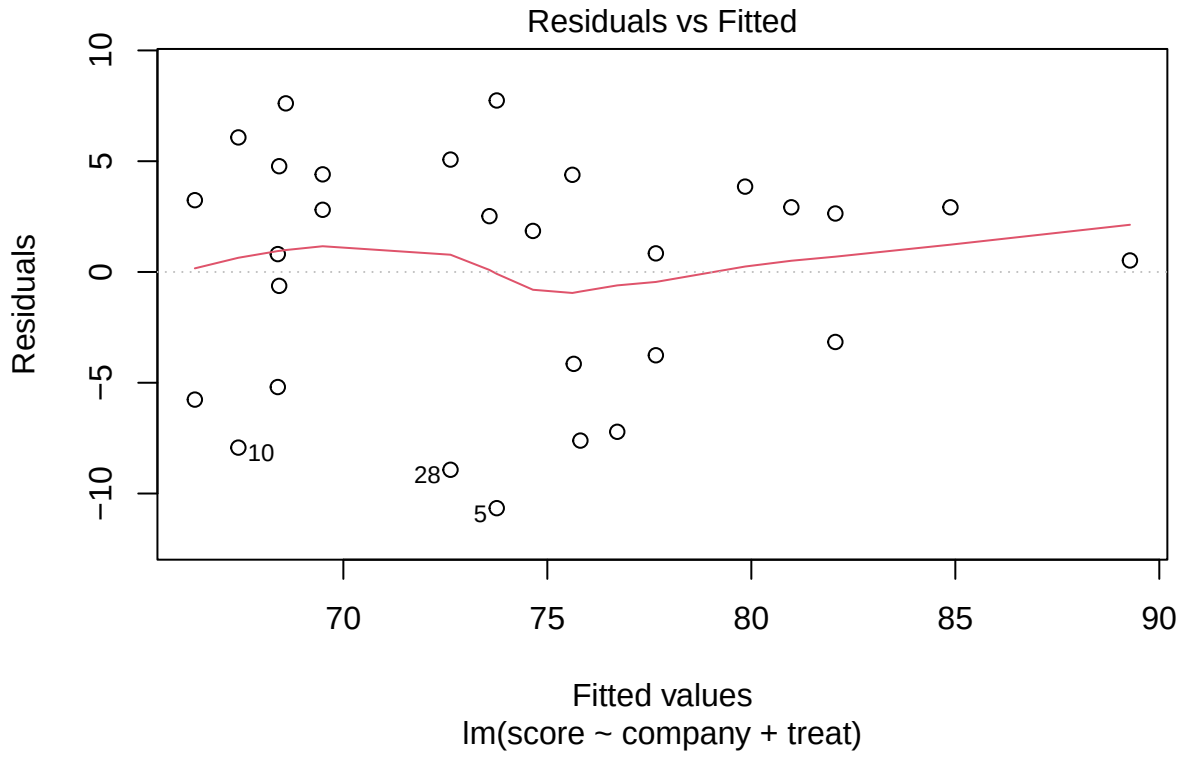
```
summary(aov(score~treat+company,pyg))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## treat        1  327.3   327.3   7.569 0.0131 *
## company      9  682.5    75.8   1.753 0.1484
## Residuals   18  778.5    43.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

So, the p-value for treatment is 0.0131 if we put treatment before company.

f

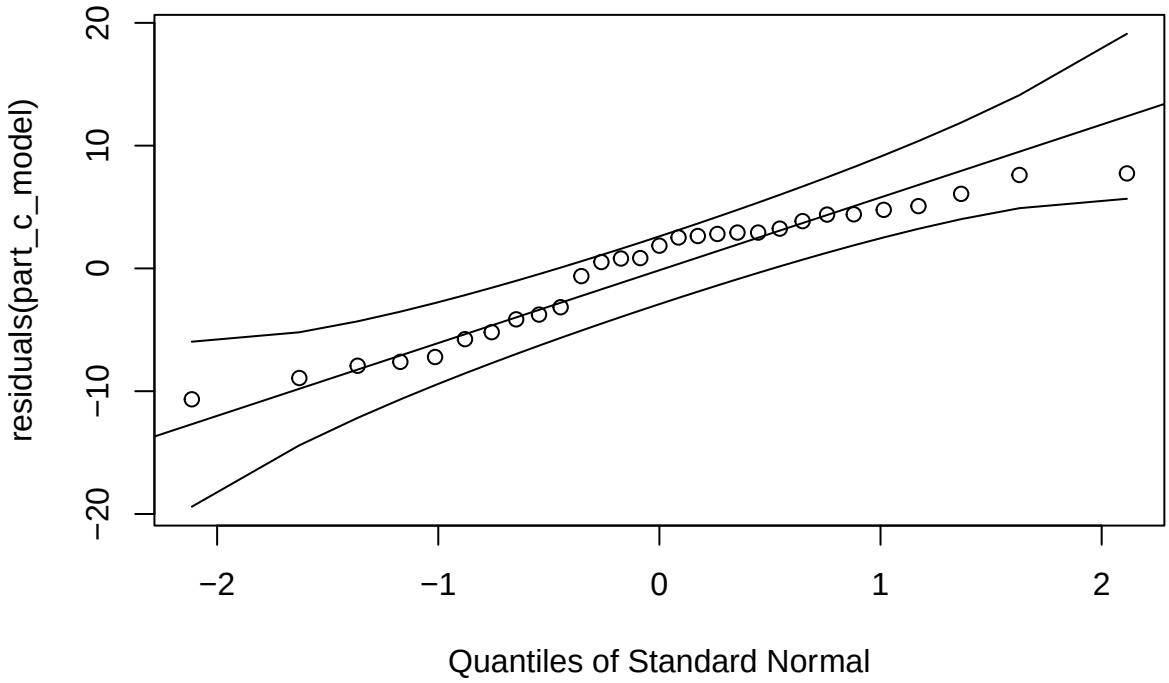
```
part_c_model = lm(score~company+treat,pyg)

plot(part_c_model, which=c(1,1))
```



The residual Vs fitted plot seems good because we are not observing any unusual pattern in the upper part and lower part of the zero line.

g

```
source("http://www.stat.cmu.edu/~hseltman/files/qqn.R")
qqn(residuals(part_c_model))
```



Other than very few observation in lower and upper tail the plot looks fine.

h

```
summary(aov(score~company+treat,pyg))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## company      9  671.0    74.6   1.724 0.1556
## treat        1  338.9   338.9   7.835 0.0119 *
## Residuals   18  778.5    43.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(score~treat,pyg))
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## treat        1  327.3   327.3   6.049 0.0206 *
## Residuals   27 1461.0    54.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When we are using the company+treatment model then the degree of freedom for residual become lesser than the model where we are only use treatment. As a result when we are dividing the residual sum of square value

using it's degree of freedom then the calculated sum of square become lower for company+treatment model. Finally when we try to find the F-statistic, by dividing all the mean square using residual sum of square it finding a higher F-statistics value for treatment in treatment+company model. So we are getting a smaller p-value for the treatment+company model. This smaller p-value increasing the power because residual mean square is nothing but the σ^2 which we want as smaller as possible. So, smaller p-value indicating a more power of the analysis.

Question 2

```
stp <- read.delim("E:/NMT MS/Spring 22/MATH 588/Home_Work/Spring-2022---MATH-588-01-Advanced-Data-Analys


dim(stp) # 30 6
```

```
## [1] 30  6
```

```
sapply(stp, class)
```

```
##     Order     Block    Height Frequency    RestHR        HR
## "numeric" "numeric" "numeric" "numeric" "numeric" "integer"
```

```
# Order Block Height Frequency RestHR HR
# "integer" "integer" "integer" "integer" "integer" "integer"
stp$Block = factor(stp$Block)
stp$Height = factor(stp$Height, labels=c("Low","High"))
stp$Frequency = factor(stp$Frequency, labels=c("Low","Med","High"))
summary(stp)
```

```
##      Order       Block  Height   Frequency     RestHR            HR
##  Min.   : 1.00   1:5   Low :15   Low :10   Min.   :60.00   Min.   : 75.0
##  1st Qu.: 8.25   2:5   High:15   Med :10   1st Qu.:72.75   1st Qu.: 93.0
##  Median :15.50   3:5             High:10   Median :81.00   Median : 99.0
##  Mean   :15.50   4:5                       Mean   :80.00   Mean   :107.4
##  3rd Qu.:22.75   5:5                       3rd Qu.:87.00   3rd Qu.:122.2
##  Max.   :30.00   6:5                       Max.   :96.00   Max.   :153.0
```

a

```
with(stp, table(Height, Frequency, Block))
```

```
## , , Block = 1
##
##       Frequency
## Height Low Med High
##   Low    1   1    1
##   High   1   1    0
##
## , , Block = 2
##
##       Frequency
## Height Low Med High
##   Low    1   1    1
##   High   1   0    1
##
## , , Block = 3
```

```
##
##       Frequency
## Height Low Med High
##   Low    0   1    1
##   High   1   1    1
##
## , , Block = 4
##
##       Frequency
## Height Low Med High
##   Low    1   0    1
##   High   1   1    1
##
## , , Block = 5
##
##       Frequency
## Height Low Med High
##   Low    1   1    1
##   High   0   1    1
##
## , , Block = 6
##
##       Frequency
## Height Low Med High
##   Low    1   1    0
##   High   1   1    1
```

In each block we observed 1 observation per cell and the missing observation is located in different position.

b

```
with(stp, table(Height, Frequency))
```

```
##       Frequency
## Height Low Med High
##   Low    5   5    5
##   High   5   5    5
```

Looks like it's a "Balanced" design.

c

```
summary(aov(HR~Block+Frequency+Height,stp))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Block        5   4511     902   16.20 1.37e-06 ***
## Frequency    2   3035    1518   27.26 1.46e-06 ***
## Height       1   3406    3406   61.17 1.18e-07 ***
## Residuals   21   1169      56
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(HR~Frequency+Block+Height,stp))
```

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

```
## Frequency    2   3728    1864   33.48 2.94e-07 ***
## Block        5   3818     764   13.71 5.13e-06 ***
## Height       1   3406    3406   61.17 1.18e-07 ***
## Residuals   21   1169      56
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(aov(HR~Block+Height+Frequency,stp))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Block        5   4511     902   16.20 1.37e-06 ***
## Height       1   3406    3406   61.17 1.18e-07 ***
## Frequency    2   3035    1518   27.26 1.46e-06 ***
## Residuals   21   1169      56
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Among these 3 combination we observed that the Height and Residuals are unchanged. The Block and Frequency sum of square changed when we are changing the position.

d

$df_B = df_{Block} + df_H + df_F = 5 + 1 + 2 = 8$
$SS_B = SS_{Block} + SS_H + SS_F = 4511 + 3406 + 3035 = 10952$
$MS_B = SS_B/df_B = 10952/8 = 1369$
$df_T = df_B + df_W = 8 + 21 = 29$
$SS_T = SS_B + SS_W = 10952 + 1169.2 = 12121.2$

e

```
summary(aov(HR~Block+Frequency*Height,stp))
```

```
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## Block             5   4511     902  19.794 6.12e-07 ***
## Frequency         2   3035    1518  33.297 6.17e-07 ***
## Height            1   3406    3406  74.733 5.20e-08 ***
## Frequency:Height  2    303     152   3.327   0.0577 .
## Residuals        19    866      46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

f

```
mi=aov(HR~Block+Frequency+Height+Frequency:Height, stp)
coefficients(mi)
```

```
##              (Intercept)                   Block2                   Block3
##                    81.90                    -3.50                    -5.25
##                   Block4                   Block5                   Block6
##                    23.00                    16.25                    -7.25
##             FrequencyMed            FrequencyHigh               HeightHigh
##                    12.25                    20.00                    20.50
##  FrequencyMed:HeightHigh FrequencyHigh:HeightHigh
##                    -6.00                     9.75
```

```
sqrt(vcov(mi)["FrequencyHigh:HeightHigh", "FrequencyHigh:HeightHigh"])
```

```
## [1] 6.162813
```

```
summary.lm(mi)
```

```
##
## Call:
## aov(formula = HR ~ Block + Frequency + Height + Frequency:Height,
##     data = stp)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.400  -4.775   0.225   4.100   9.350
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 81.900      4.088  20.035 3.09e-14 ***
## Block2                      -3.500      4.358  -0.803 0.431813
## Block3                      -5.250      4.358  -1.205 0.243094
## Block4                      23.000      4.358   5.278 4.29e-05 ***
## Block5                      16.250      4.358   3.729 0.001423 **
## Block6                      -7.250      4.358  -1.664 0.112580
## FrequencyMed                12.250      4.358   2.811 0.011151 *
## FrequencyHigh               20.000      4.358   4.590 0.000200 ***
## HeightHigh                  20.500      4.358   4.704 0.000154 ***
## FrequencyMed:HeightHigh     -6.000      6.163  -0.974 0.342497
## FrequencyHigh:HeightHigh     9.750      6.163   1.582 0.130137
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.751 on 19 degrees of freedom
## Multiple R-squared:  0.9286, Adjusted R-squared:  0.891
## F-statistic:  24.7 on 10 and 19 DF,  p-value: 8.423e-09
```

```
qt(0.975, 19)
```

```
## [1] 2.093024
```

g

```
summary(aov(HR~Block+Frequency+Height,stp))
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Block        5   4511     902   16.20 1.37e-06 ***
## Frequency    2   3035    1518   27.26 1.46e-06 ***
## Height       1   3406    3406   61.17 1.18e-07 ***
## Residuals   21   1169      56
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

h

```
#install.packages("gmodels")
library(gmodels)
```

```r
levels(stp$Frequency)
```

```
## [1] "Low"  "Med"  "High"
```

```r
s0 = aov(HR~Block+Frequency+Height,stp)
contr = rbind(HvsML = c(-1/2, -1/2, 1), MvsL = c(-1, 1, 0))
round( fit.contrast(s0, "Frequency", contr, conf.int=0.95), 3)
```

```
##                Estimate Std. Error t value Pr(>|t|) lower CI upper CI
## FrequencyHvsML    20.25      2.950   6.865    0.000   14.116   26.384
## FrequencyMvsL      9.25      3.406   2.716    0.013    2.167   16.333
## attr(,"class")
## [1] "fit_contrast"
```