

Homework 7

Statistical Computing, STAT 3675Q

Anthony Anderson

General Instructions

- Answer the questions by inserting R code and necessary comments. Your output must contain the R code (do not use the `echo=FALSE` option).
- After you complete the assignment, save it under the file name `LastName-FirstName-HW7.pdf`
- Then submit the compiled PDF file through HuskyCT by April 12, 2019, at 11:59 PM.

Data

The file `raw.txt` contains the following variables:

- `id`: Each ID represents a participant
- `Sex`: Participant's gender
- `Genotype`: Three types: `ff` or `fs` or `ss`
- `Activity`: enzyme activity of mannose-6-phosphate isomerase (MPI).

a. [2 points] Read `raw.txt` into R. Check the structure of your data set.

```
Data <- matrix(data=c("id", "Sex", "Genotype", "Activity",
1, "male", "ff", 1.884,
2, "male", "ff", 2.283,
3, "male", "fs", 2.396,
4, "female", "ff", 2.838,
5, "male", "fs", 2.956,
6, "female", "ff", 4.216,
7, "female", "ss", 3.620,
8, "female", "ff", 2.889,
9, "female", "fs", 3.550,
10, "male", "fs", 3.105,
11, "female", "fs", 4.556,
12, "female", "fs", 3.087,
13, "male", "ff", 4.939,
14, "male", "ff", 3.486,
15, "female", "ss", 3.079,
16, "male", "fs", 2.649,
17, "female", "fs", 1.943,
19, "female", "ff", 4.198,
20, "female", "ff", 2.473,
22, "female", "ff", 2.033,
24, "female", "fs", 2.200,
25, "female", "fs", 2.157,
26, "male", "ss", 2.801,
28, "male", "ss", 3.421,
29, "female", "ff", 1.811,
30, "female", "fs", 4.281,
32, "female", "fs", 4.772,
34, "female", "ss", 3.586,
36, "female", "ff", 3.944,
38, "female", "ss", 2.669,
```

```

39, "female", "ss", 3.050,
41, "male", "ss", 4.275,
43, "female", "ss", 2.963,
46, "female", "ss", 3.236,
48, "female", "ss", 3.673,
49, "male", "ss", 3.110), byrow=T, ncol=4)
colnames(Data)<- c("id","Sex","Genotype","Activity")
Data<- Data[-1,]
Data<-data.frame(Data)
Data$Activity<-as.numeric(as.character(Data$Activity), length=4)
str(Data)

```

```

## 'data.frame': 36 obs. of 4 variables:
## $ id : Factor w/ 36 levels "1","10","11",...: 1 11 19 26 32 33 34 35 36 2 ...
## $ Sex : Factor w/ 2 levels "female","male": 2 2 2 1 2 1 1 1 2 ...
## $ Genotype: Factor w/ 3 levels "ff","fs","ss": 1 1 2 1 2 1 3 1 2 2 ...
## $ Activity: num 1.88 2.28 2.4 2.84 2.96 ...

```

b. [2 points] Create a table containing the sample size for each combination of sex and genotype.

```

SexGeno.table<- xtabs(~Data$Sex+Data$Genotype)
addmargins(SexGeno.table)

```

```

##          Data$Genotype
## Data$Sex ff fs ss Sum
## female  8  8  8  24
## male    4  4  4  12
## Sum     12 12 12  36

```

c. [4 points] Calculate the mean and standard deviation for each combination of sex and genotype.

```

library("doBy")

```

```

## Warning: package 'doBy' was built under R version 3.5.3

```

```

ByGroup.mean.sd<-summaryBy(Activity~Sex+Genotype,data=Data,FUN=c(mean,sd))
ByGroup.mean.sd

```

```

##      Sex Genotype Activity.mean Activity.sd
## 1 female      ff      3.05025    0.9599032
## 2 female      fs      3.31825    1.1445388
## 3 female      ss      3.23450    0.3617754
## 4 male       ff      3.14800    1.3745115
## 5 male       fs      2.77650    0.3168433
## 6 male       ss      3.40175    0.6348109

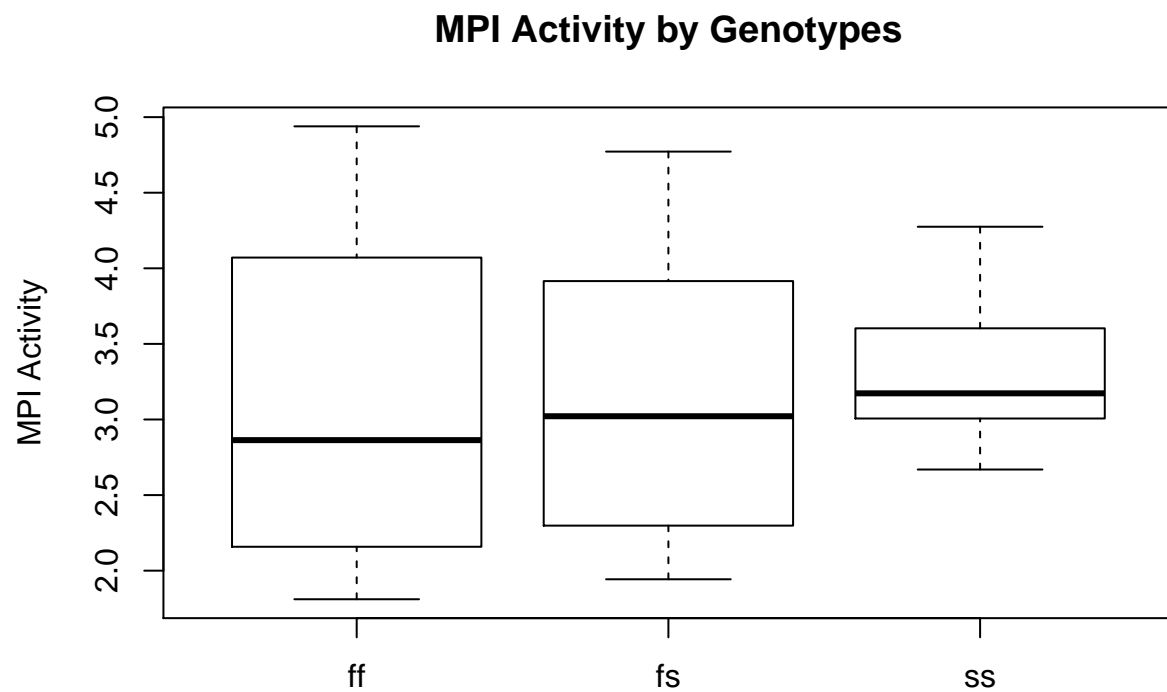
```

d. [2 points] Create side by side boxplots of MPI activity for the three types of **Genotype**.

```

boxplot(Activity~Genotype, data=Data, ylab="MPI Activity", main="MPI Activity by Genotypes")

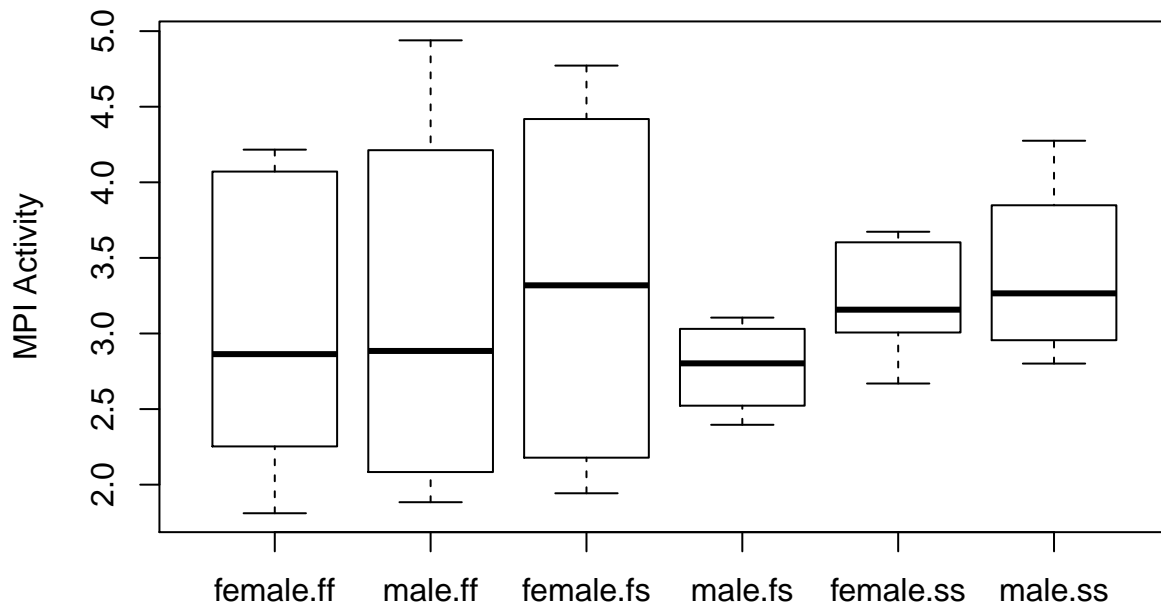
```



e. [2 points] Create side by side boxplots of MPI activity for all factorial combinations (interaction) of **Genotype** and **Sex**.

```
boxplot(Activity~Sex+Genotype,data=Data, ylab="MPI Activity",  
        main="MPI Activity by Genotype and Sex")
```

MPI Activity by Genotype and Sex



f. [5 points] Conduct a two-way ANOVA which includes sex, genotype and their interaction as independent variables. Interpret your results.

```
twoway.aov<-aov(Activity~Sex*Genotype, data=Data)
twoway.lm<-lm(Activity~Sex*Genotype, data=Data)
summary(twoway.aov)
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Sex         1  0.068   0.0681   0.086  0.771
## Genotype    2  0.277   0.1386   0.175  0.840
## Sex:Genotype 2  0.815   0.4073   0.515  0.603
## Residuals   30 23.714   0.7905
```

All of the F-statistic's p-values are larger than .05 or .01, so are non significant. None of these variables are statistically significant in predicting MPI Activity.

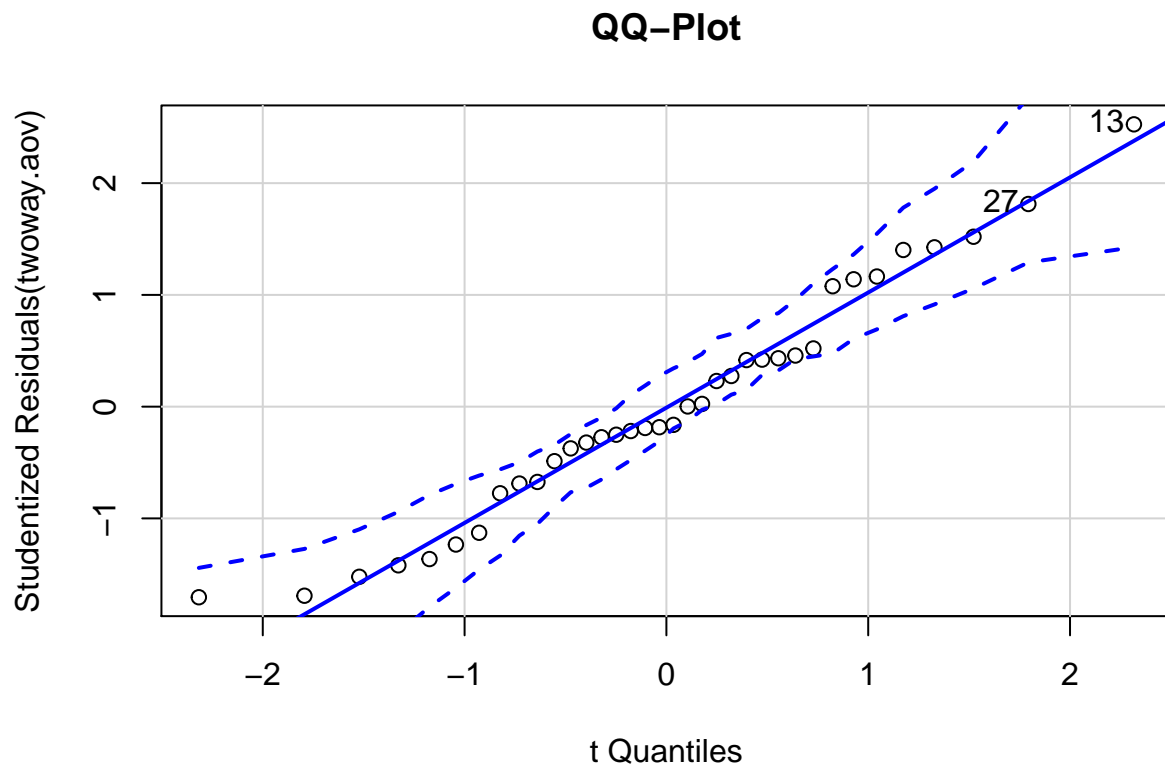
g. [5 points] What are the assumptions necessary for the ANOVA model? Assess the assumptions using **graphical** methods and **formal** statistical tests, where possible.

```
library(car)
```

```
## Warning: package 'car' was built under R version 3.5.3
```

```
## Loading required package: carData
```

```
qqPlot(twoway.aov, main="QQ-Plot")
```



```
## [1] 13 27
```

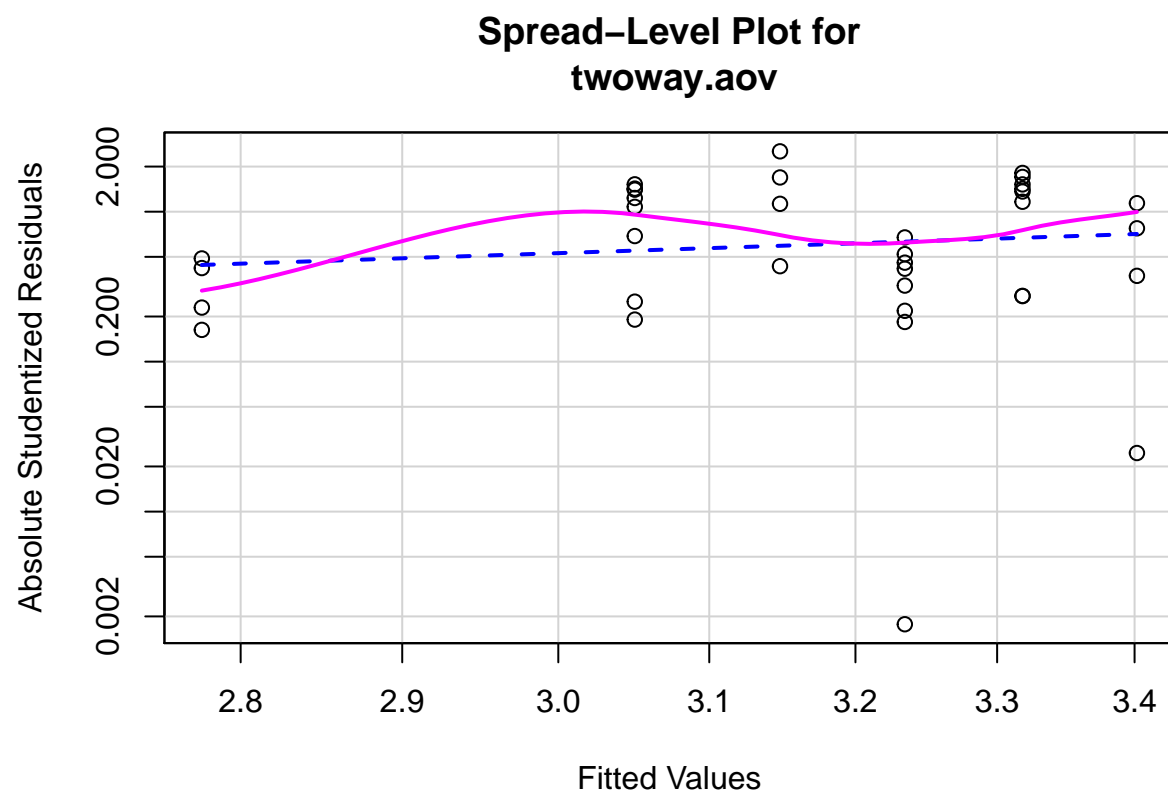
```
shapiro.test(twoway.lm$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  twoway.lm$residuals
## W = 0.97002, p-value = 0.426
```

```
ncvTest(twoway.lm)
```

```
## Non-constant Variance Score Test
## Variance formula: ~ fitted.values
## Chisquare = 0.4555531, Df = 1, p = 0.49971
```

```
spreadLevelPlot(twoway.aov)
```



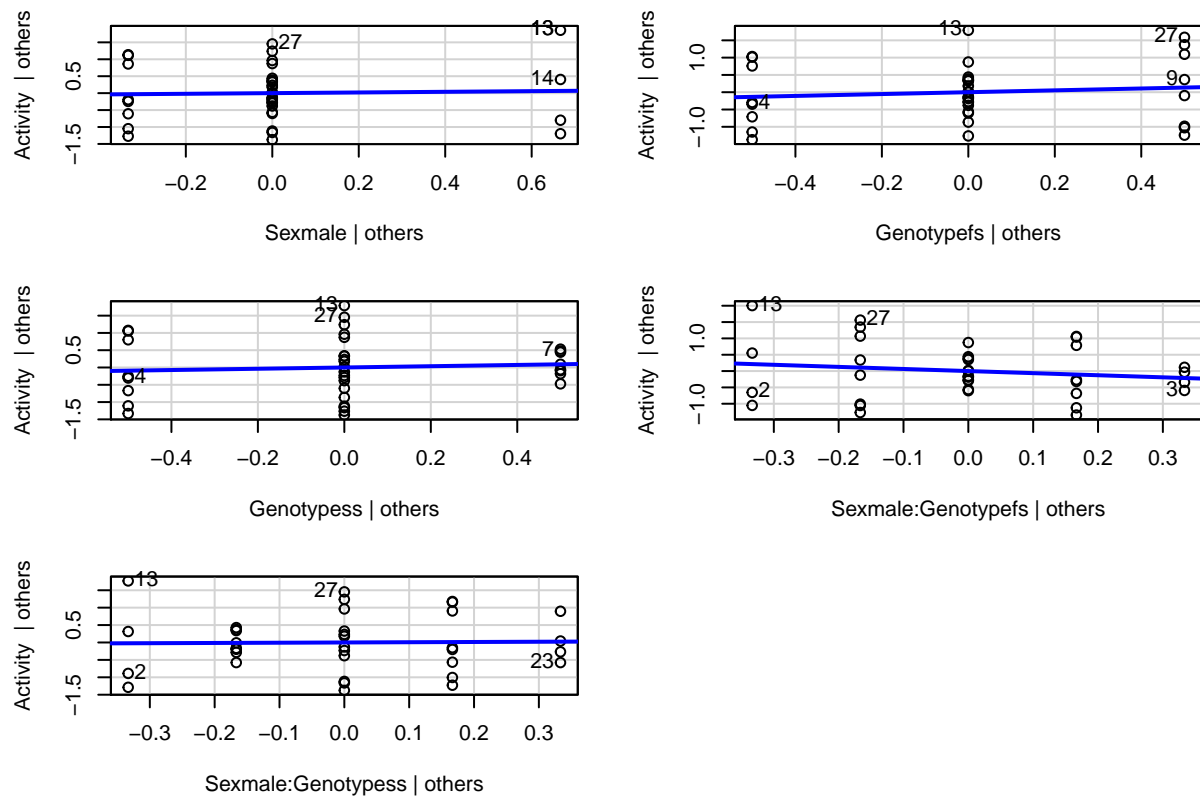
```
##
## Suggested power transformation: -1.34063
```

```
vif(twoway.lm)
```

```
##           GVIF Df  GVIF^(1/(2*Df))
## Sex           3.00  1      1.732051
## Genotype       2.25  2      1.224745
## Sex:Genotype   5.25  2      1.513700
```

```
avPlots(twoway.aov)
```

Added-Variable Plots



Assumptions: Equal variance is reasonable from the `ncvTest` output's high p-value, Normality is reasonable from the `shapiro-wilk`'s high p-value, Independence is reasonable from the low VIF of Sex and Genotype, and Linearity is reasonable from the linear relationships in the `avPlot` output.

h. [4 points] Compare the mean MPI for all possible pairs (note that there are 6 sex-genotype combinations, so the number of pairwise comparisons is $6 \times 5/2 = 15$). Use Tukey's HSD procedure and interpret your results

```
TukeyHSD(twoway.aov)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Activity ~ Sex * Genotype, data = Data)
##
## $Sex
##           diff          lwr          upr      p adj
## male-female -0.09225 -0.7342113 0.5497113 0.7711798
##
## $Genotype
##           diff          lwr          upr      p adj
## fs-ff 0.05483333 -0.8399734 0.9496401 0.9875021
## ss-ff 0.20741667 -0.6873901 1.1022234 0.8362403
## ss-fs 0.15258333 -0.7422234 1.0473901 0.9074857
##
## $`Sex:Genotype`
##           diff          lwr          upr      p adj
```

```
## male:ff-female:ff      0.09775 -1.558238 1.753738 0.9999712
## female:fs-female:ff    0.26800 -1.084108 1.620108 0.9900169
## male:fs-female:ff     -0.27375 -1.929738 1.382238 0.9956835
## female:ss-female:ff    0.18425 -1.167858 1.536358 0.9982708
## male:ss-female:ff      0.35150 -1.304488 2.007488 0.9863961
## female:fs-male:ff      0.17025 -1.485738 1.826238 0.9995569
## male:fs-male:ff       -0.37150 -2.283670 1.540670 0.9908872
## female:ss-male:ff      0.08650 -1.569488 1.742488 0.9999843
## male:ss-male:ff        0.25375 -1.658420 2.165920 0.9984769
## male:fs-female:fs     -0.54175 -2.197738 1.114238 0.9159152
## female:ss-female:fs   -0.08375 -1.435858 1.268358 0.9999634
## male:ss-female:fs      0.08350 -1.572488 1.739488 0.9999868
## female:ss-male:fs      0.45800 -1.197988 2.113988 0.9571582
## male:ss-male:fs        0.62525 -1.286920 2.537420 0.9160754
## male:ss-female:ss      0.16725 -1.488738 1.823238 0.9995937
```

All the p-values are almost 1, suggesting there is no significant difference between any of these effects. None of the contrasts are significant.

i. [4 points] Produce an interaction plot using the function `plotmeans` in the `gplots` package. Enhance your plot (color of lines, meaningful titles, axis labels, ...).

```
library("gplots")
```

```
## Warning: package 'gplots' was built under R version 3.5.3
```

```
##
```

```
## Attaching package: 'gplots'
```

```
## The following object is masked from 'package:stats':
```

```
##
```

```
##      lowess
```

```
plotmeans(Activity~interaction(Sex,Genotype,sep=" "),data=Data, main="Interaction Plot of Sex and Genot,
```


Interaction Plot of Sex and Genotype

