Statistical Methods – COSC 6323 - HomeWork-5

By Dinesh Narlakanti (2083649)

INTRODUCTION

More than 2,50,000 participants from Texas A&M University, University of Houston and University of California, Irvine. The data in this document has information about the participants who recorded their perinasal perspiration values, heart beat rate(Chest and Wrist) and breathe rate while performing different tasks i.e; Resting Baseline(RB), Single task(ST), Priming(PM), Relaxing Video(RV), Dual Task(DT) and Presentation Session(PR).

This report purely concentrates on:
i) Checking whether there is significance difference between the three treatments (ST, DT and PR).
ii) Comparing pp__qc values of all the treatments before and after normalization.
iii) Determining most stressful treatment using posthoc method.

GETTING STARTED WITH THE HOMEWORK

Step-1 Installing required packages and importing data.

```
psy_data <- read.csv("C:/Users/ndine/OneDrive/Desktop/Physiological Data.csv")
```

Step-2 Subsetting based on RB, ST, DT and PR treatments and removing NAs for both pp__qc and chest heart rate. Later log transforming pp__qc columns for Q1.

```
RB <- subset(psy_data, psy_data$Treatment == "RB")
RB <- RB[!is.na(RB$PP_QC),]
RB_before <- RB$PP_QC
RB$PP_QC <- log(RB$PP_QC)

PR <- subset(psy_data, psy_data$Treatment == "PR")
PR <- PR[!is.na(PR$PP_QC),]
PR_before <- PR$PP_QC
PR$PP_QC <- log(PR$PP_QC)


ST <- subset(psy_data, psy_data$Treatment == "ST")
ST <- ST[!is.na(ST$PP_QC),]
ST_before <- ST$PP_QC
ST$PP_QC <- log(ST$PP_QC)

DT <- subset(psy_data, psy_data$Treatment == "DT")
DT <- DT[!is.na(DT$PP_QC),]
DT_before <- DT$PP_QC
DT$PP_QC <- log(DT$PP_QC)

RB_2 <- subset(psy_data, psy_data$Treatment == "RB")
RB_2 <- RB_2[!is.na(RB_2$Chest_HR_QC),]

ST_2 <- subset(psy_data, psy_data$Treatment == "ST")
ST_2 <- ST_2[!is.na(ST_2$Chest_HR_QC),]

DT_2 <- subset(psy_data, psy_data$Treatment == "DT")
DT_2 <- DT_2[!is.na(DT_2$Chest_HR_QC),]
```

```
PR_2 <- subset(psy_data, psy_data$Treatment == "PR")
PR_2 <- PR_2[!is.na(PR_2$Chest_HR_QC),]
```

Step-3 Aggregating by participant id and calculating mean of pp_qc values for Q1 and chest heart rate for Q2.

```
RB_pp <- aggregate(RB$PP_QC, by = list(RB$Participant_ID), FUN = mean)
PR_pp <- aggregate(PR$PP_QC, by = list(PR$Participant_ID), FUN = mean)
ST_pp <- aggregate(ST$PP_QC, by = list(ST$Participant_ID), FUN = mean)
DT_pp <- aggregate(DT$PP_QC, by = list(DT$Participant_ID), FUN = mean)

RB_chest <- aggregate(RB_2$Chest_HR_QC, by = list(RB_2$Participant_ID),
                      FUN = mean)
PR_chest <- aggregate(PR_2$Chest_HR_QC, by = list(PR_2$Participant_ID),
                      FUN = mean)
ST_chest <- aggregate(ST_2$Chest_HR_QC, by = list(ST_2$Participant_ID),
                      FUN = mean)
DT_chest <- aggregate(DT_2$Chest_HR_QC, by = list(DT_2$Participant_ID),
                      FUN = mean)
```

Step-4 Taking common participants between RB and other treatments.

```
PR_RB <- inner_join(RB_pp, PR_pp, by = 'Group.1')
ST_RB <- inner_join(RB_pp, ST_pp, by = 'Group.1')
DT_RB <- inner_join(RB_pp, DT_pp, by = 'Group.1')

PR_RB_chest <- inner_join(RB_chest, PR_chest, by = 'Group.1')
ST_RB_chest <- inner_join(RB_chest, ST_chest, by = 'Group.1')
DT_RB_chest <- inner_join(RB_chest, DT_chest, by = 'Group.1')
```

Step-5 Subtracting other treatments' pp_qc values from RB for Q1 and subtracting other treatments' chest rate from RB chest rate.

```
PR_RB$aroused <- PR_RB$x.y - PR_RB$x.x
ST_RB$aroused <- ST_RB$x.y - ST_RB$x.x
DT_RB$aroused <- DT_RB$x.y - DT_RB$x.x

PR_RB$Treatment <- 'PR'
ST_RB$Treatment <- 'ST'
DT_RB$Treatment <- 'DT'

PR_RB_chest$aroused <- PR_RB_chest$x.y - PR_RB_chest$x.x
ST_RB_chest$aroused <- ST_RB_chest$x.y - ST_RB_chest$x.x
DT_RB_chest$aroused <- DT_RB_chest$x.y - DT_RB_chest$x.x

PR_RB_chest$Treatment <- 'PR'
ST_RB_chest$Treatment <- 'ST'
DT_RB_chest$Treatment <- 'DT'
```

Step-6 Combining all the data and extracting factors from the combined data.

```
comby_data <- rbind(PR_RB, ST_RB, DT_RB)
treatment_factors <- factor(comby_data$Treatment)

comby_data_chest <- rbind(PR_RB_chest, ST_RB_chest, DT_RB_chest)
treatment_factors_chest <- factor(comby_data_chest$Treatment)
```

Step-7 Performing ANOVA tests for both Q1 and Q2

```
comb_res <- aov(comby_data$aroused ~ treatment_factors, data = comby_data)
summary(comb_res)
```

```
##                    Df Sum Sq Mean Sq F value Pr(>F)
## treatment_factors   2  2.404  1.2018   53.03 <2e-16 ***
## Residuals         184  4.170  0.0227
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
comb_res_chest <- aov(comby_data_chest$aroused ~ treatment_factors_chest,
                      data = comby_data_chest)
summary(comb_res_chest)
```

```
##                          Df Sum Sq Mean Sq F value   Pr(>F)
## treatment_factors_chest   2   3192  1596.0   24.72 4.89e-10 ***
## Residuals               154   9942    64.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Step-8 Posthoc tests

```
posthoc <- TukeyHSD(x = comb_res, "treatment_factors", conf.level=0.95)
posthoc
```

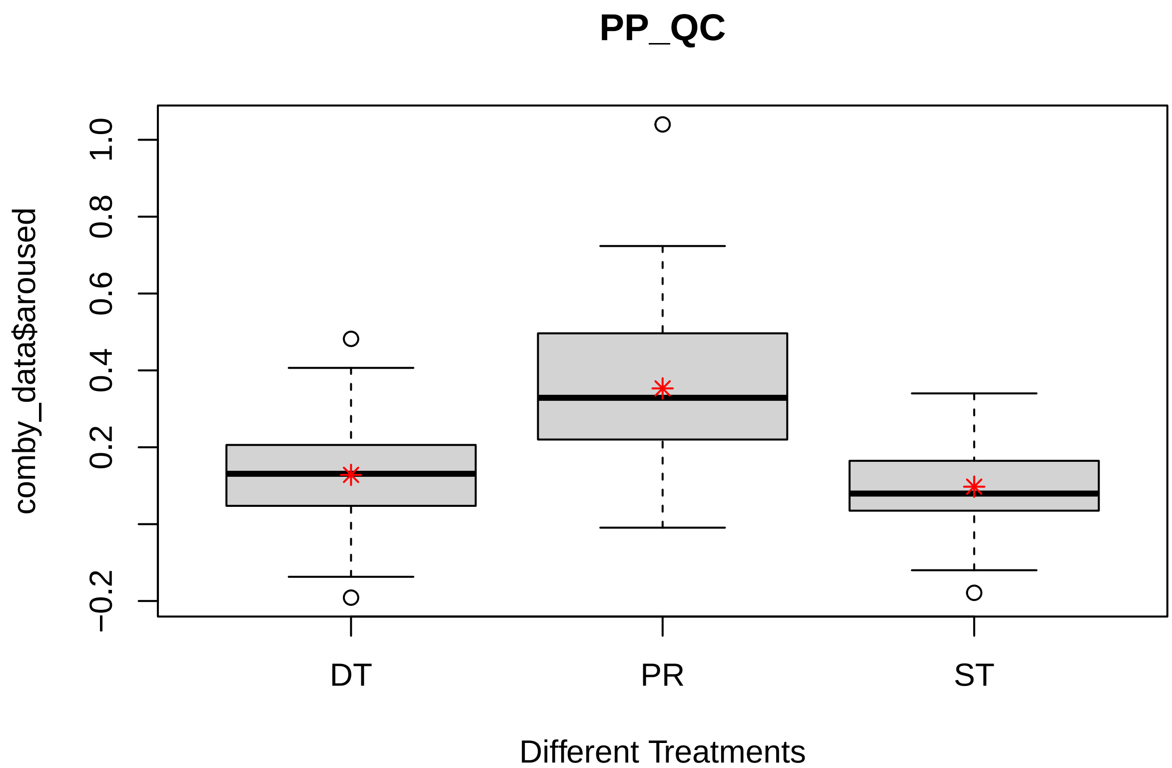
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = comby_data$aroused ~ treatment_factors, data = comby_data)
##
## $treatment_factors
##             diff         lwr         upr     p adj
## PR-DT  0.2249671  0.16107392  0.28886034 0.0000000
## ST-DT -0.0307125 -0.09408834  0.03266335 0.4876852
## ST-PR -0.2556796 -0.31957284 -0.19178642 0.0000000
```

```
posthoc_chest <- TukeyHSD(x = comb_res_chest, "treatment_factors_chest", conf.level=0.95)
posthoc_chest
```

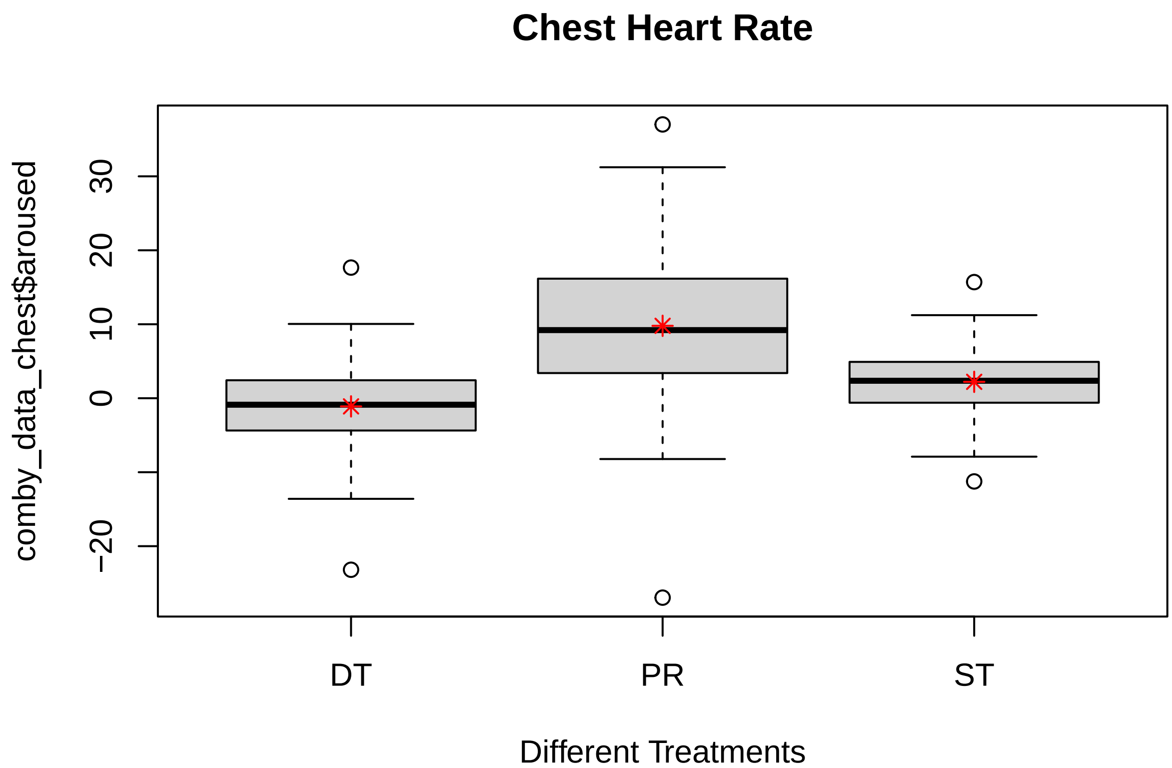
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = comby_data_chest$aroused ~ treatment_factors_chest, data = comby_data_chest)
##
## $treatment_factors_chest
##            diff         lwr       upr     p adj
## PR-DT 10.896505   7.1476246 14.645385 0.0000000
## ST-DT  3.321375  -0.3553653  6.998115 0.0856341
## ST-PR -7.575130 -11.3071215 -3.843138 0.0000109
```

Step-9 Plotting boxplots for both the aov results.

```
boxplot(comby_data$aroused ~ treatment_factors, main = 'PP_QC',
        xlab = 'Different Treatments')
means <- tapply(comby_data$aroused, treatment_factors, mean)
points( means, pch=8, col="red")
```

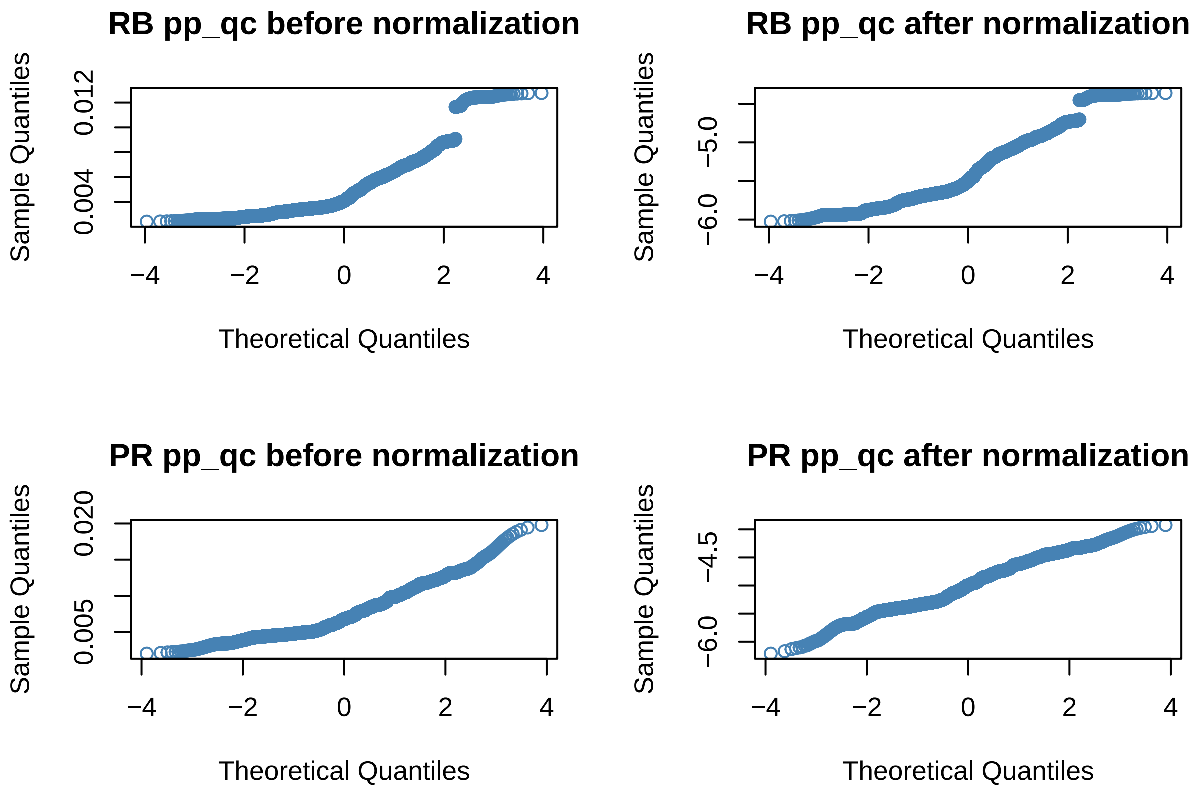



```r
boxplot(comby_data_chest$aroused ~ treatment_factors_chest,
        main = 'Chest Heart Rate', xlab = 'Different Treatments')
means_chest <- tapply(comby_data_chest$aroused, treatment_factors_chest, mean)
points( means_chest, pch=8, col="red")
```

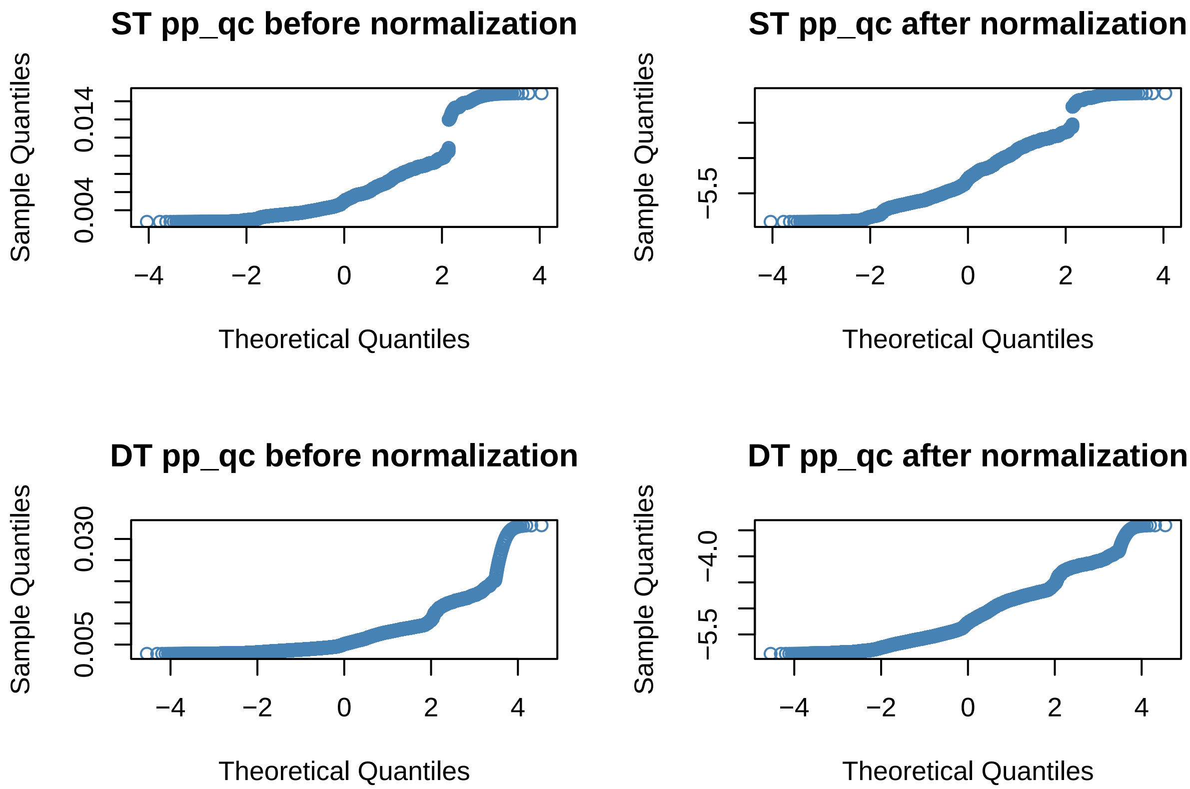


Step-10 qqplots for comparing before and after normalization of pp qc values

```
par(mfrow=c(2,2))
qqnorm(RB_before, main = "RB pp_qc before normalization", col = "steelblue")
qqnorm(RB$PP_QC, main = "RB pp_qc after normalization",col = "steelblue")
qqnorm(PR_before, main = "PR pp_qc before normalization", col = "steelblue")
qqnorm(PR$PP_QC, main = "PR pp_qc after normalization",col = "steelblue")
```

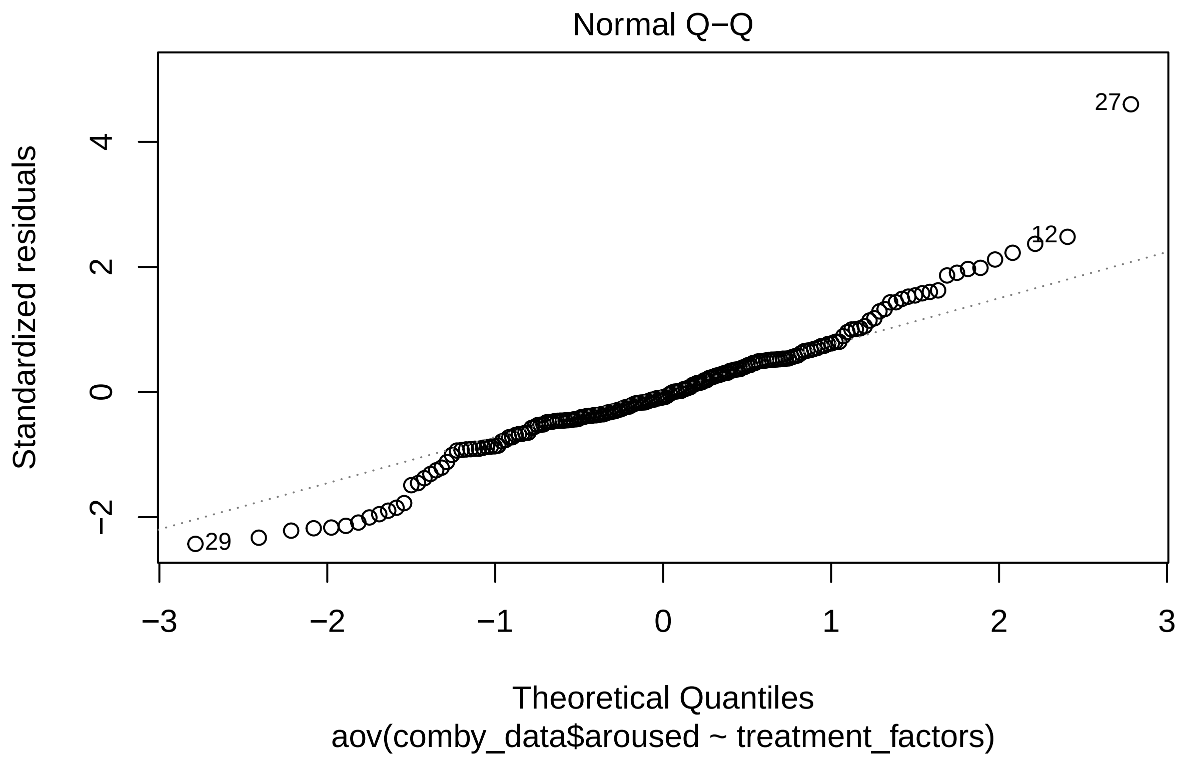



```
par(mfrow=c(2,2))
qqnorm(ST_before, main = "ST pp_qc before normalization",col = "steelblue")
qqnorm(ST$PP_QC, main = "ST pp_qc after normalization",col = "steelblue")
qqnorm(DT_before, main = "DT pp_qc before normalization",col = "steelblue")
qqnorm(DT$PP_QC, main = "DT pp_qc after normalization",col = "steelblue")
```

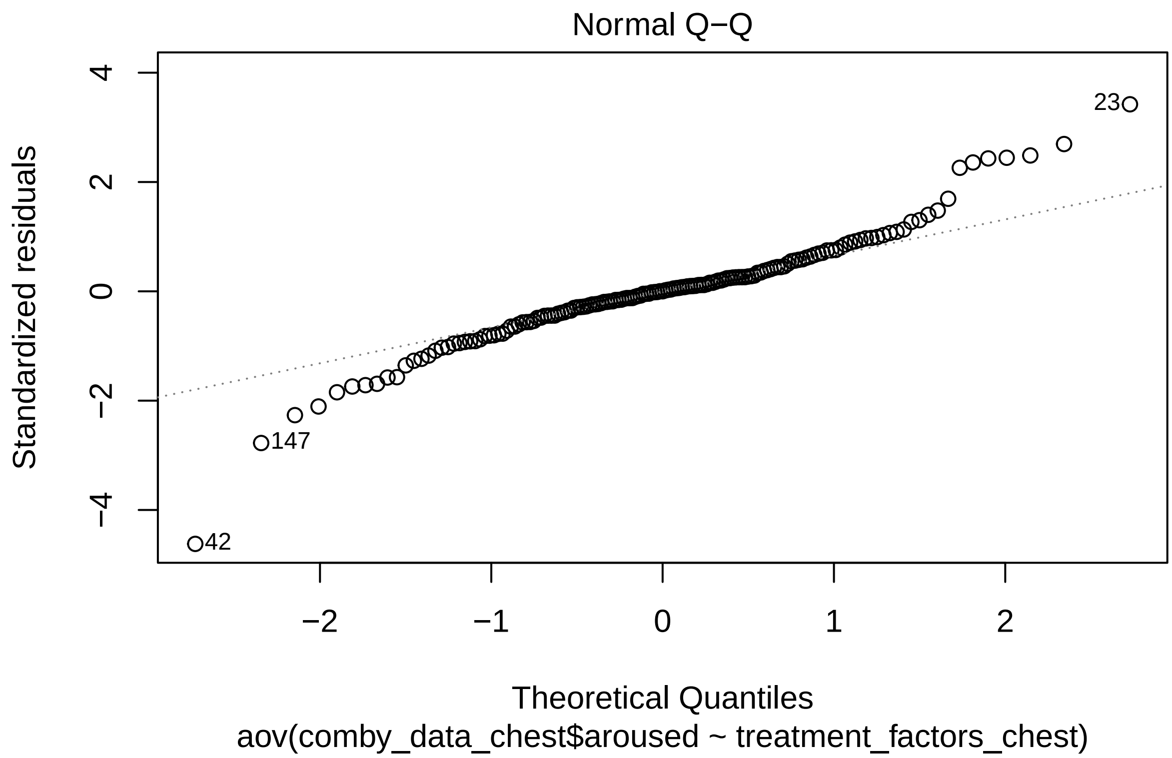


Step-11

```
plot(comb_res,2)
```




```
plot(comb_res_chest,2)
```



ANALYSIS OF THE RESULTS

Question-1

1) From the summary of aov results of different treatments' pp_qc values, the p-value is less than 0.05 (<2e-16), so that we can reject the null hypothesis and can conclude that the difference of means of the treatments is significantly different.

2) From the results of TuckeyHSD, we can conclude that PR is significantly different from both ST and DT. Because the pvalue of PR-DT is 0.000 and p-value of ST-PR is 0.000 which are less than 0.05 and can also reject null hypothesis.

3) From the boxplot, we can clearly see that PR is most stressful treatment. This shows that social stressor is the most stressor when compared to time pressure stressor and cognitive effort stressor.

4) Also, When we observe the Tukey results, difference of PR-DT is positive (0.2249671) and ST-PR difference is negative (-0.2556796), This shows that PR mean is greater than DT and ST. It indicates that PR treatment is the most stressful treatment than DT and ST. This concludes that, according to normalized pp_qc values, social stressor is the most stressful than cognitive effort stressor and time pressure stressor.

Question-2

5) From the summary of aov results of different treatments' chest heart rate values, as the p-value is less than 0.05 (4.89e-10), so that we can reject null hypothesis and we can conclude that difference of means of all three treatments is significantly different.

6) From TuckeyHSD results, we can conclude that PR is significantly different from both DT and ST. Because the p-values of PR-DT is 0.000 and ST-PR is 0.0000109 which are less than 0.05 and can also reject null hypothesis.

7) From the boxplot, we can also conclude PR is most stressful treatment than DT and ST. This tells that social stressor is most stressor than cognitive effort stressor and time pressure stressor.

8)Also, When we observe the Tukey results, difference of PR-DT is positive (10.896505) and ST-PR difference is negative (-7.575130), This shows that PR mean is greater than DT and ST. It indicates that PR treatment is the most stressful treatment than DT and ST. This concludes that, according to chest heart rate values, social stressor is the most stressful than cognitive effort stressor and time pressure stressor.

9) Both the methods indicates most stressful treatment is PR.