Statistical Methods – COSC 6323 - HomeWork-3

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 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 (PR).

This report purely concentrates on:
i) Chest heart rate and Wrist heart rate are in agreement or not,
ii) Determining the appropriate test for the above.
iii) Checking the optimal method to use in the given data organizations methods.
iv) Determining the participants are under stress with perinasal perspiration as physiological measure

GETTING STARTED WITH THE HOMEWORK

Step-1 Importing physiological data and removing NAs from both Chest and wrist columns.

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

psy_data <- psy_data_ori[!is.na(psy_data_ori$Chest_HR_QC),]
psy_data <- psy_data[!is.na(psy_data$Wrist_HR_QC),]
```

Step-2 Aggregating by Participant ID and Treatment. Later performing pooled t.test

```
chest_HR <- aggregate(psy_data$Chest_HR_QC, by = list(psy_data$Participant_ID,
                                                      psy_data$Treatment),
                      FUN = mean)

wrist_HR <- aggregate(psy_data$Wrist_HR_QC, by = list(psy_data$Participant_ID,
                                                      psy_data$Treatment),
                      FUN = mean)


t.test(chest_HR$x, wrist_HR$x, alternative = 'two.sided')
```

```
##
##  Welch Two Sample t-test
##
## data:  chest_HR$x and wrist_HR$x
## t = 1.2461, df = 315.05, p-value = 0.2136
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7518689  3.3493476
## sample estimates:
## mean of x mean of y
##  80.72465  79.42591
```

Step-3 Performing pooled t.test of the measurement taken every second.

```
t.test(psy_data$Chest_HR_QC, psy_data$Wrist_HR_QC)
```

```
##
```

```
##  Welch Two Sample t-test
##
## data:  psy_data$Chest_HR_QC and psy_data$Wrist_HR_QC
## t = 24.704, df = 257230, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.9258472 1.0854193
## sample estimates:
## mean of x mean of y
##  79.65946  78.65383
```

Step-5 Performing paired t.test on grouped mean data.

```
t.test(chest_HR$x, wrist_HR$x, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  chest_HR$x and wrist_HR$x
## t = 3.3063, df = 160, p-value = 0.001167
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.5229857 2.0744931
## sample estimates:
## mean of the differences
##                1.298739
```

Step-6 performing paired t.test on the measurements taken every second.

```
t.test(psy_data$Chest_HR_QC, psy_data$Wrist_HR_QC, paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  psy_data$Chest_HR_QC and psy_data$Wrist_HR_QC
## t = 36.379, df = 129280, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.9514535 1.0598130
## sample estimates:
## mean of the differences
##                1.005633
```

Step-7 Sub setting based on treatments and removing NAs from each subset.

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

PR <- subset(psy_data_ori, psy_data_ori$Treatment == "PR")
PR <- PR[!is.na(PR$PP_QC),]
```

Step-8 Aggregating by ID and treatment. Inner joining the data to take common participants from both RB Experiment and PR Experiment.

```
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)

combined_data <- inner_join(RB_pp, PR_pp, by = 'Group.1')
```

```
per_second_data <- inner_join(RB, PR, by = 'Participant_ID')
```

Step-9 Performing t.test taken on measurement at every second.

```
t.test(combined_data$x.x, combined_data$x.y, paired = TRUE)
```

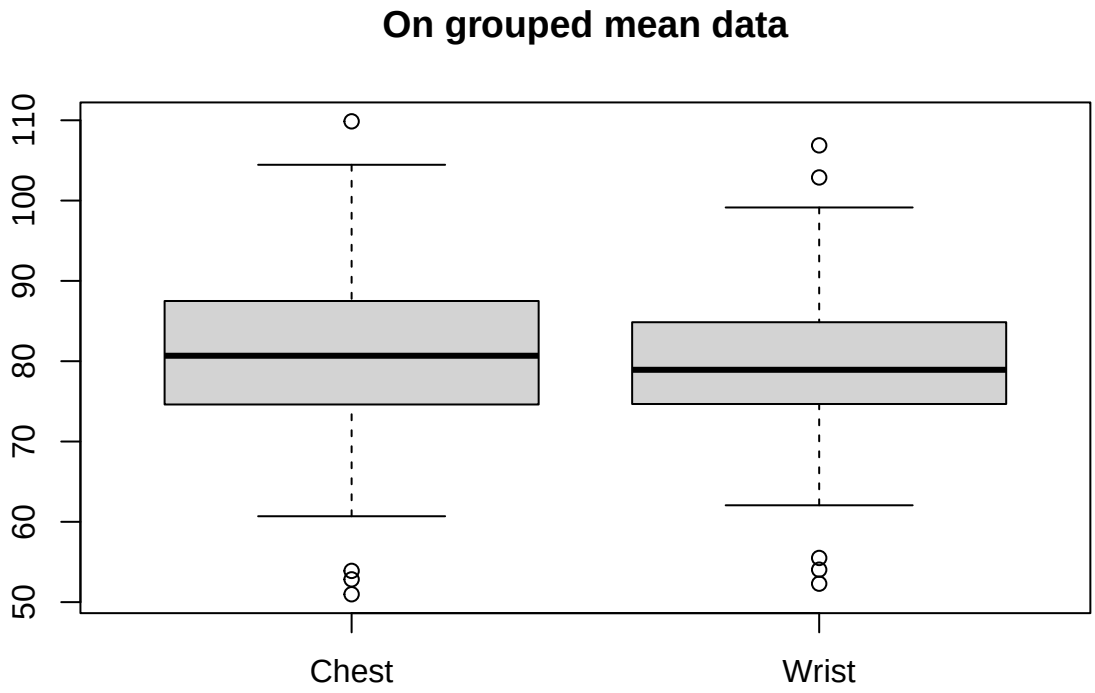
```
##
##  Paired t-test
##
## data:  combined_data$x.x and combined_data$x.y
## t = -11.514, df = 60, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.002499969 -0.001759904
## sample estimates:
## mean of the differences
##            -0.002129937
```

```
t.test(per_second_data$PP_QC.x, per_second_data$PP_QC.y, paired = TRUE )
```

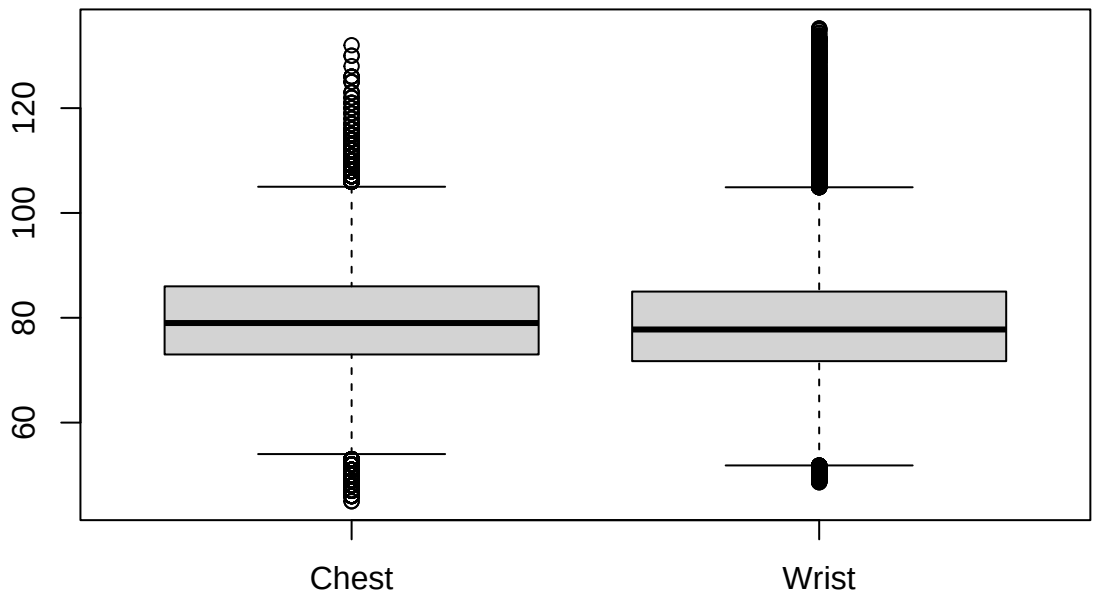
```
##
##  Paired t-test
##
## data:  per_second_data$PP_QC.x and per_second_data$PP_QC.y
## t = -1897.6, df = 2172501, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.001907502 -0.001903565
## sample estimates:
## mean of the differences
##            -0.001905534
```

Step-10) Plotting the graphs.

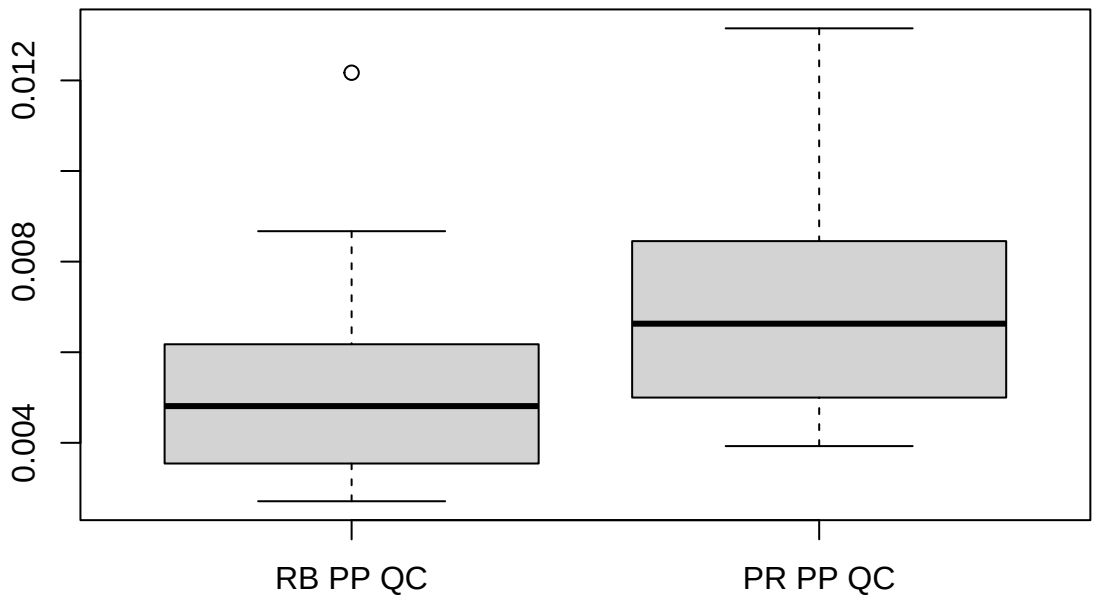
```
#par(mfrow=c(2,2))
boxplot(chest_HR$x, wrist_HR$x, main = "On grouped mean data",
        names=c('Chest', 'Wrist'))
```



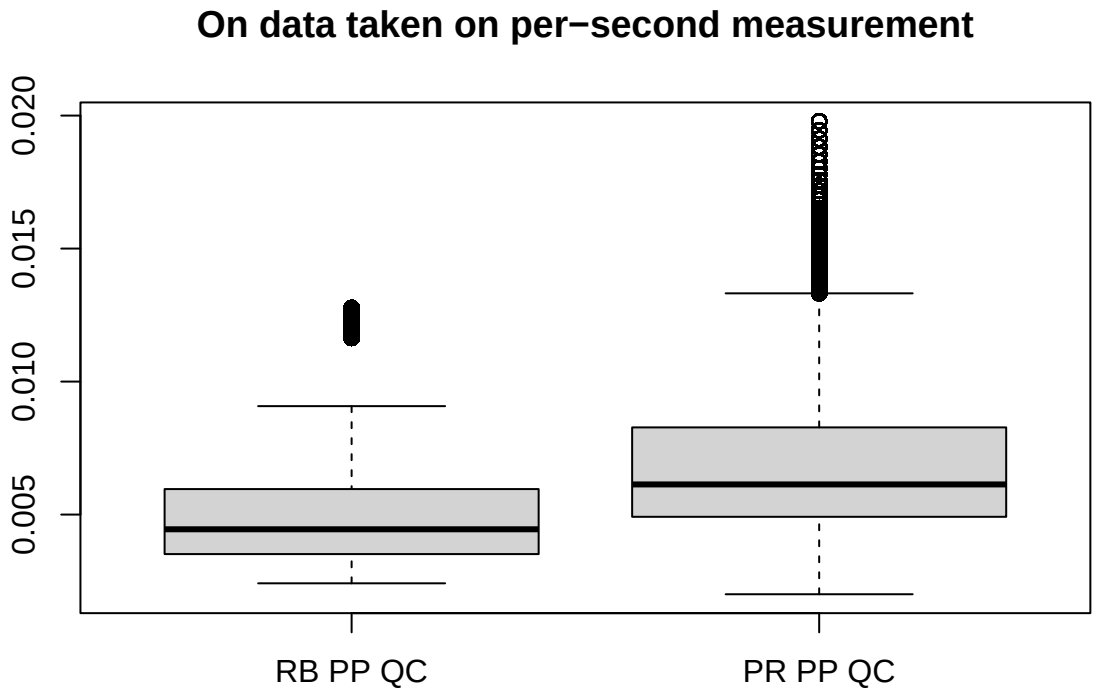

```
boxplot(psy_data$Chest_HR_QC, psy_data$Wrist_HR_QC,
        main = "On data taken on per-second.", names = c('Chest', 'Wrist') )
```




```
boxplot(combined_data$x.x, combined_data$x.y, main = "On Grouped data",
        names = c('RB PP QC', 'PR PP QC'))
```




```
boxplot(per_second_data$PP_QC.x, per_second_data$PP_QC.y,
        main = "On data taken on per-second measurement",
        names = c('RB PP QC', 'PR PP QC'))
```



ANSWERS TO THE QUESTIONS

A-1) According to my observation, pooled t.test when grouped is not in agreement with paired t test of grouped data. Pooled t test of grouped says there is no significance difference while paired t test of grouped data says there is significance difference between the means of chest data and wrist data. While pooled t test on grouped data, paired t test on grouped data and paired t test taken on every second measurement are in agreement. Except pooled t test on grouped data, rest all the t test says there is a significance difference in the means of the two heart rates.

As we are comparing two independent groups (Chest heart rate and Wrist heart rate), and are not dependent on each other, Pooled t test is appropriate to use here.

Taking the mean of each participant in each treatment is optimal method than taking the data of measurements every second because if we take a mean, we get a single value to compare or perform tests while in the other case, number of observations per participant differs. T003 in Rb have 279 observations while T005 in RB have 285.

```
nrow(subset(psy_data_ori, psy_data_ori$Participant_ID == 'T003' & psy_data_ori$Treatment == 'RB'))
```

```
## [1] 279
```

```
nrow(subset(psy_data_ori, psy_data_ori$Participant_ID == 'T005' & psy_data_ori$Treatment == 'RB'))
```

```
## [1] 285
```

Analysis of Pooled t test on grouped data:
p-value here is 0.2136, which is greater than significance level of 0.05. So, we can accept the null hypothesis and conclude that there is no significant difference.
Analysis of Pooled t test taken on every second measurement

p-value here is <2.2e-16, which is less than 0.05. So, we can reject the null hypothesis and conclude that there is significant difference.
Analysis of Paired t test on grouped data
p-value here is 0.001167, which is less than 0.05. So, we can reject the null hypothesis and conclude that there is significant difference.
Analysis of Paired t test taken on every second measurement
p-value here is <2.2e-16, which is less than 0.05. So, we can reject the null hypothesis and conclude that there is significant difference.

A-2) As p value is less than the significance level of 0.05, we can reject the null and conclude that there is significant difference between PR PP_QC and RB PP_QC. We can also infer and conclude that participants are in stress in PR experiment.

Paired t test is appropriate to use here because there is a relationship between two groups. Relationships are:
1. Same exact group is tested twice
2. Group is tested before presentation session (can be called RB experiment) and after presentation session.

Optimal method here is to use mean of each participant in each treatment. Because, when observed, each participant didn't have same number of readings. So, taking mean of the available data is accurate to perform t test and plot visualizations. Same explanation from above answser also suits here.
Analysis of t test on grouped mean data
p value here is <2.2e-16, which is less than 0.05. So, we can reject the null hypothesis and conclude that there is significant difference. That means participants in PR experiment are in stress. This is told based on the perinasal perspiration values of PR experiment and RB experiment.
Analysis of t test taken on every second measurement
p value here is <2.2e-16, which is less than 0.05. So, we can reject the null hypothesis and conclude that there is significant difference. That means participants in PR experiment are in stress. This is told based on the perinasal perspiration values of PR experiment and RB experiment.
To verify: Let's put alternative as greater.

```
t.test( combined_data$x.y,combined_data$x.x, alternative = 'greater', paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  combined_data$x.y and combined_data$x.x
## t = 11.514, df = 60, p-value < 2.2e-16
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.001820885         Inf
## sample estimates:
## mean of the differences
##             0.002129937
```

As p-value is 1, which is greater than 0.05, we can reject the null hypothesis and conclude mean of PR perinasal perspiration is greater than mean RB perinasal perspiration.