

# Mini Project # 5

Qingyu Lan, Lakshmi Priyanka Selvaraj

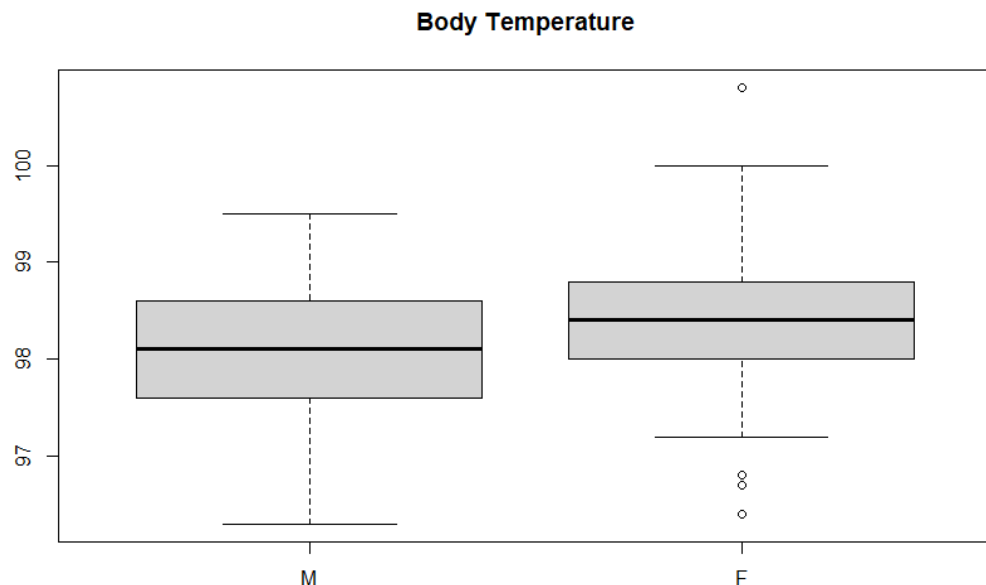
## Contribution of each group member :

Both members worked on the questions together.

## Section 1. Answers to the specific questions asked

### 1. Question 1

a.



From the exploratory data analysis it seems like there is a difference between male and female body temperature, since the range of the two have differences between them.

Null hypothesis: Difference between means of male and female = 0

Alternative hypothesis: Difference between means of male and female  $\neq$  0

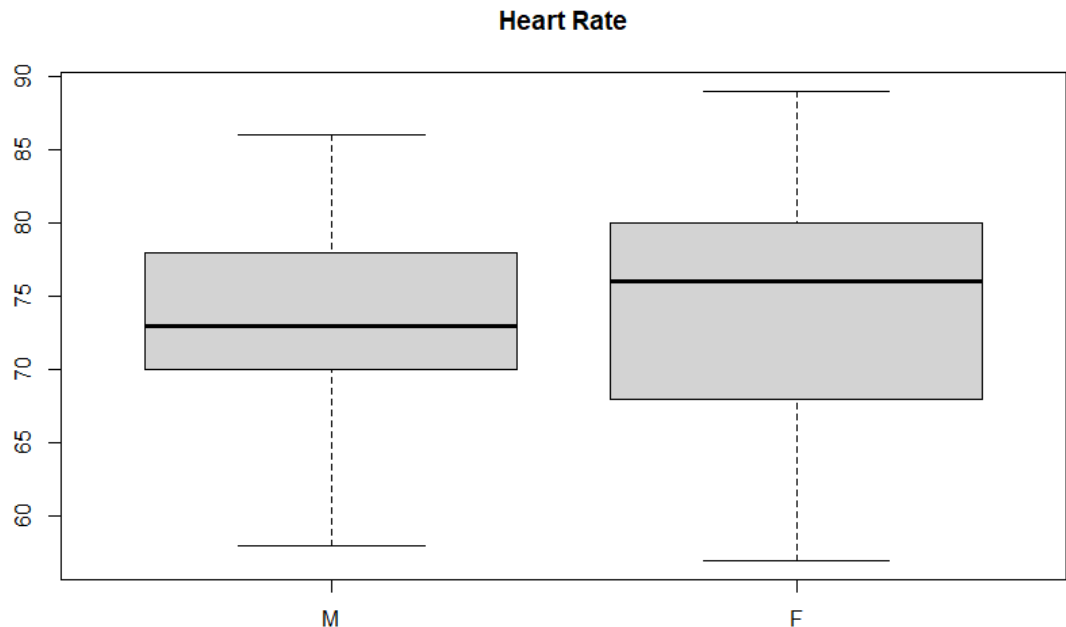
```
> t.test(male$body_temperature, female$body_temperature, alternative = 'two.sided', var.equal = F)

Welch Two Sample t-test

data: male$body_temperature and female$body_temperature
t = -2.2854, df = 127.51, p-value = 0.02394
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.53964856 -0.03881298
sample estimates:
mean of x mean of y
 98.10462  98.39385
```

From the t test, we have a low p-value of 0.02394, and the 95% confidence interval does not contain 0, therefore we have to reject our null hypothesis, and take the alternative hypothesis that there is a difference between the mean of male and female body temperature.

b.



From the exploratory analysis it seems male and female have around the same heart rate, although the median is different by a bit, the quartiles and minimum and maximum seems to center around the same area.

Null hypothesis: Difference between means of male and female = 0

Alternative hypothesis: Difference between means of male and female  $\neq$  0

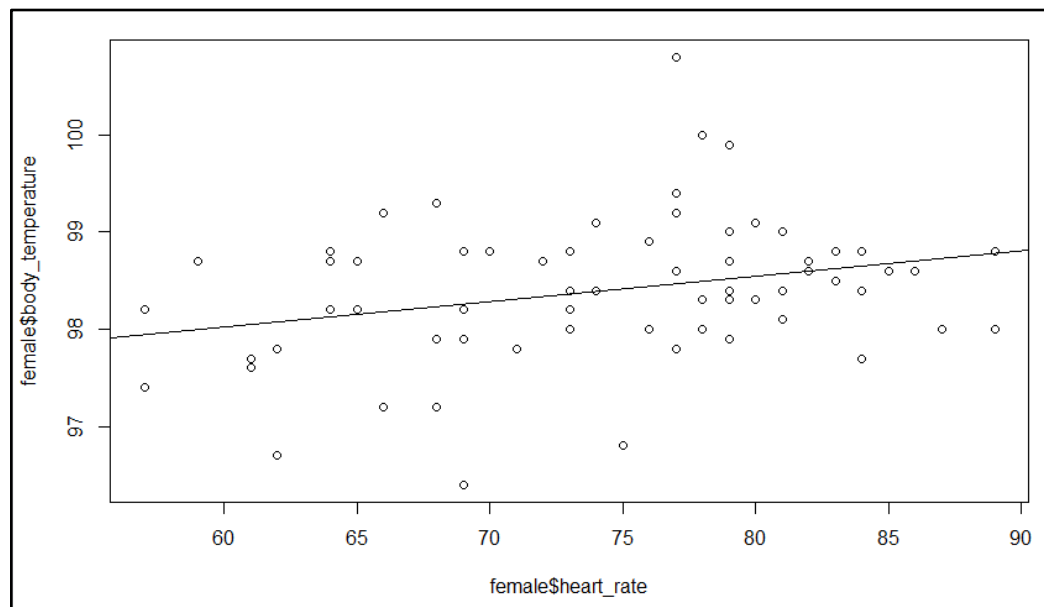
```
> t.test(male$heart_rate, female$heart_rate, alternative = "two.sided", var.equal = F)
```

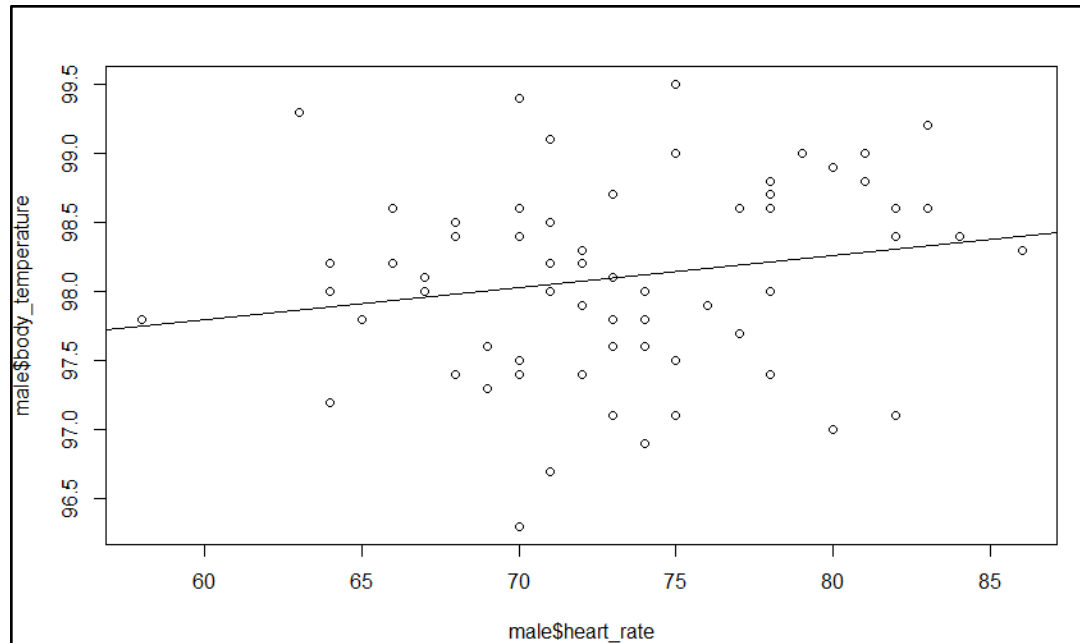
Welch Two Sample t-test

```
data: male$heart_rate and female$heart_rate
t = -0.63191, df = 116.7, p-value = 0.5287
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.243732  1.674501
sample estimates:
mean of x mean of y
 73.36923  74.15385
```

From the t test, we have a p-value of 0.5287, and the 95% confidence interval does contain 0, therefore we keep our null hypothesis, there is no difference between male and female heart rate

c.

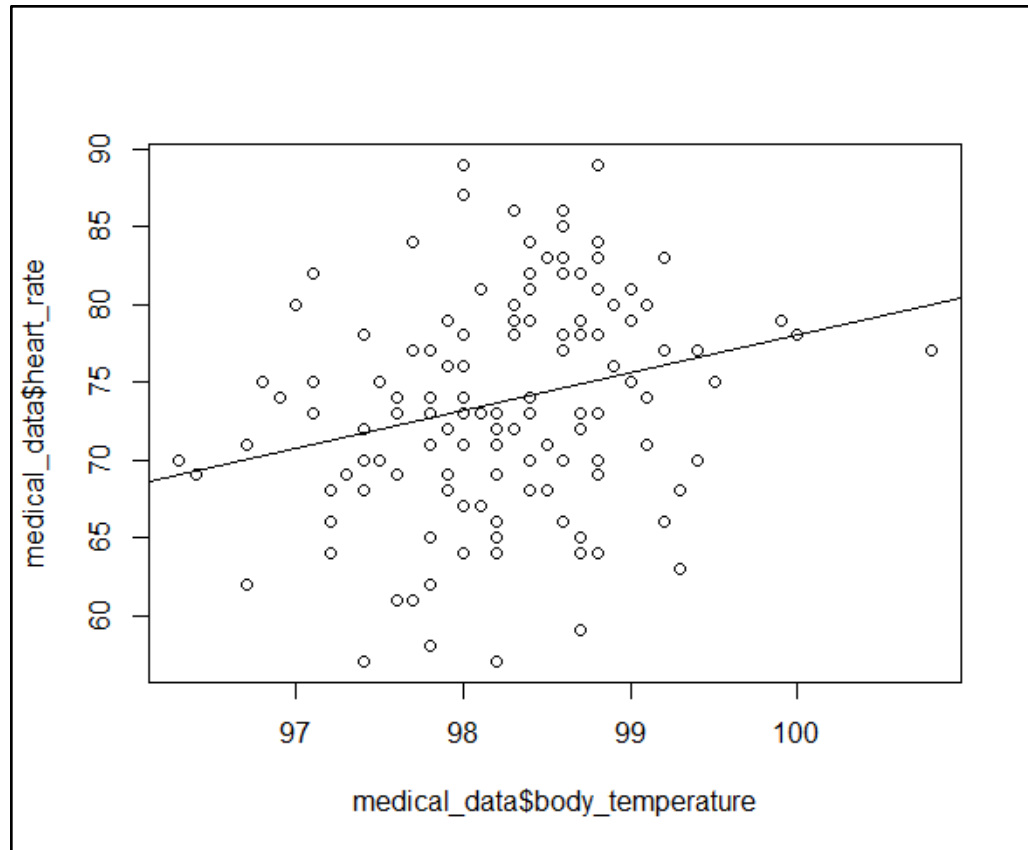




From the exploratory analysis, we can see that both male and female have a positive relationship between body temperature and heart rate.

```
> cor(male$body_temperature, male$heart_rate)
[1] 0.1955894
> cor(female$body_temperature, female$heart_rate)
[1] 0.2869312
```

From the correlation results, we can see that there is a weak positive relationship between body temperature and heart rate for both male and female, with the female having a bit stronger correlation than male.



On comparing the correlation between the heart rate and body temperature, irrespective of the gender, we find there is a very weak positive correlation of 0.2536.

```
> cor(medical_data$body_temperature,medical_data$heart_rate)
[1] 0.2536564
```

## 2. Question 2

a.

```

/
> temp = coverageProb(0.01,5,0.05)
> cp1 <- c(cp1,temp[1])
> cp2 <- c(cp2,temp[2])
> print(temp)
[1] 0.8130 0.7902

```

Our first monte carlo run to find the coverage probability of the confidence intervals using  $\lambda = 0.01$ ,  $n = 5$

b.

Coverage probability for confidence interval using the z-interval method:

	n = 5	n = 10	n = 30	n = 100
$\lambda = 0.01$	0.8086	0.8594	0.9166	0.9382
$\lambda = 0.1$	0.8062	0.8688	0.9126	0.9338
$\lambda = 1$	0.8006	0.8678	0.9138	0.9370
$\lambda = 10$	0.8086	0.8684	0.9222	0.9374

Coverage probability for confidence interval using the bootstrap percentile interval method:

	n = 5	n = 10	n = 30	n = 100
$\lambda = 0.01$	0.7814	0.8530	0.9182	0.9380
$\lambda = 0.1$	0.7816	0.8650	0.9110	0.9356
$\lambda = 1$	0.7772	0.8624	0.9136	0.9374
$\lambda = 10$	0.7818	0.8624	0.9242	0.9380

c.

From our result, the larger the  $n$  value is, the more accurate is the intervals for both the z-interval method and the percentile method. At  $n = 100$ , both the z-interval and bootstrap percentile method have around 0.93 coverage probability. The answer does not seem to be dependent on  $\lambda$  as the result is fairly similar with different  $\lambda$ s but the same  $n$  value for both types of intervals. Although the z-interval method is more accurate when  $n = 5$ , the accuracy between the two becomes almost the same for  $n = 10$  and above. I would recommend the z-interval method because it performs better at  $n = 5$ , but there is no significant difference at higher  $n$  values.

d.

No, it does not depend on the specific values of  $\lambda$  that were fixed in advance, since the result improves across all values of  $\lambda$  as  $n$  increases.

## Section 2: R code.

R code for question 1

```

# make sure to have the file bodytemp-heartrate.csv in the working directory
# we did it by having it pasted in the working directory
getwd()

```

```

# read from the csv file
bodytemp_hearttrate = read.csv("bodytemp-hearttrate.csv", header = T)

# print the reading variable to check
print(bodytemp_hearttrate)

# divide the variables into male and female
male = subset(bodytemp_hearttrate, bodytemp_hearttrate$gender == 1)
female = subset(bodytemp_hearttrate, bodytemp_hearttrate$gender == 2)

# box plot for a and b
boxplot(male$body_temperature, female$body_temperature, names = c("M","F"), main = "Body
Temperature")
boxplot(male$heart_rate, female$heart_rate, names = c("M","F"), main = "Heart Rate")

# t test for question a and b
t.test(male$body_temperature, female$body_temperature, alternative = 'two.sided', var.equal =
F)
t.test(male$heart_rate, female$heart_rate, alternative = "two.sided", var.equal = F)

# correlation values
cor(male$body_temperature, male$heart_rate)
cor(female$body_temperature, female$heart_rate)

# scatter plot with regression line
plot(male$heart_rate, male$body_temperature, pch = 1)
abline(lm(male$body_temperature~male$heart_rate))

plot(female$heart_rate, female$body_temperature, pch = 1)
abline(lm(female$body_temperature~female$heart_rate))

#scatterplot for entire data

cor(bodytemp_hearttrate$body_temperature,bodytemp_hearttrate$heart_rate)
plot(bodytemp_hearttrate$body_temperature,bodytemp_hearttrate$heart_rate,pch=1)
abline(lm(bodytemp_hearttrate$heart_rate~bodytemp_hearttrate$body_temperature))

```

## R code for question 2

```

library(boot)

# make a function for bootstrap to use for finding the mean
mean.npar <- function(data, indices){
  return(mean(data[indices]))
}

```

```
}
```

```
# make a function to generate both confidence interval based on
# lambda, n, and alpha
conf.fun <- function(lambda, n, alpha){
  # generate the data
  x= rexp(n,lambda)
  # find ci with z
  ci= mean(x) + c(-1, 1) * qnorm(1-alpha/2) * sd(x)/sqrt(n)
  #print(ci)
  # find ci with boottraps
  mean.npar.boot = boot(x, mean.npar, R=999, sim="ordinary", stype = "i")
  ci2 = sort(mean.npar.boot$t)[c(25, 975)]
  #print(ci2)
  return(c(ci,ci2))
}
```

```
#test
#cic = conf.fun(0.01,5,0.05)
#print(cic)
#print(cic[3])
```

```
# make a function to do replication and find out the coverage probability
coverageProb <- function(lambda,n,alpha){
  ci_collection = replicate(10, conf.fun(lambda, n, alpha))
  # find the conditional probablity of both ci with z and ci with bootstrap
  p1 = mean((((1/lambda)>=ci_collection[1,])*((1/lambda)<=ci_collection[2,])))
  p2 = mean((((1/lambda)>=ci_collection[3,])*((1/lambda)<=ci_collection[4,])))
  return(c(p1,p2))
}
```

```
# make list to store coverage probability
cp = c(coverageProb(0.01,5,0.05),
  coverageProb(0.01,10,0.05),
  coverageProb(0.01,30,0.05),
  coverageProb(0.01,100,0.05),
  coverageProb(0.1,5,0.05),
  coverageProb(0.1,10,0.05),
  coverageProb(0.1,30,0.05),
  coverageProb(0.1,100,0.05),
  coverageProb(1,5,0.05),
  coverageProb(1,10,0.05),
```



```

coverageProb(1,30,0.05),
coverageProb(1,100,0.05),
coverageProb(10,5,0.05),
coverageProb(10,10,0.05),
coverageProb(10,30,0.05),
coverageProb(10,100,0.05)
)

```

```
print(cp)
```

```
# break apart the cp into cp for z and cp for percentage ci method
```

```

cp1 = c()
cp2 = c()
odd = T
for(i in cp){
  if (odd == T){
    odd = F
    cp1 = c(cp1,i)
  }else{
    odd = T
    cp2 = c(cp2,i)
  }
}

```

```
# test codes
```

```
#print(cp1)
```

```
#print(cp2)
```

```
# make table for z-interval
```

```

cptable <- matrix(cp1,ncol=4,byrow=TRUE)
colnames(cptable) <- c("n = 5","n = 10","n = 30", "n = 100")
rownames(cptable) <- c("lambda = 0.01","lambda = 0.1","lambda = 1", "lambda = 10")
cptable <- as.table(cptable)
cptable

```

```
# make table for percentile method
```

```

cptable1 <- matrix(cp2,ncol=4,byrow=TRUE)
colnames(cptable1) <- c("n = 5","n = 10","n = 30", "n = 100")
rownames(cptable1) <- c("lambda = 0.01","lambda = 0.1","lambda = 1", "lambda = 10")
cptable1 <- as.table(cptable1)
cptable1

```

