Mini Project #4

Kainan Xu

2021229632

**Exercise 1**

**(a)**

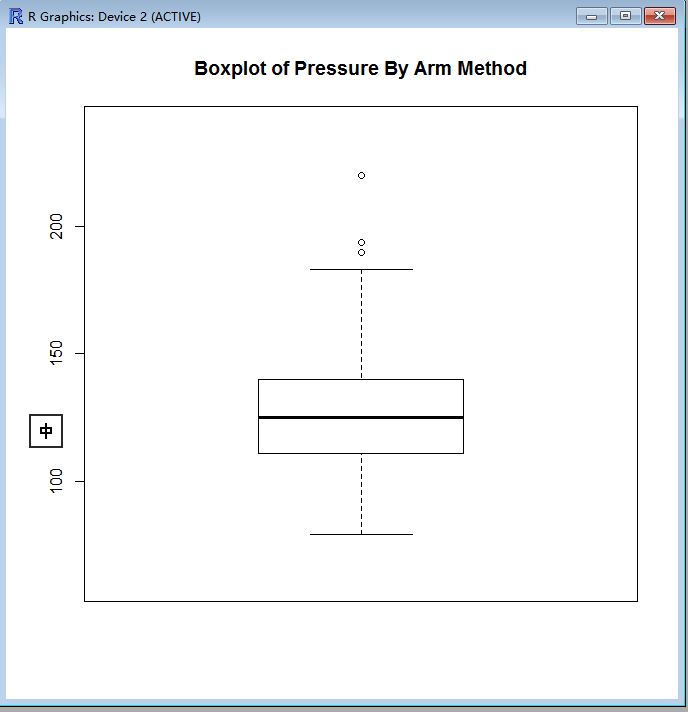


Fig 1.Boxplot of pressure by arm method

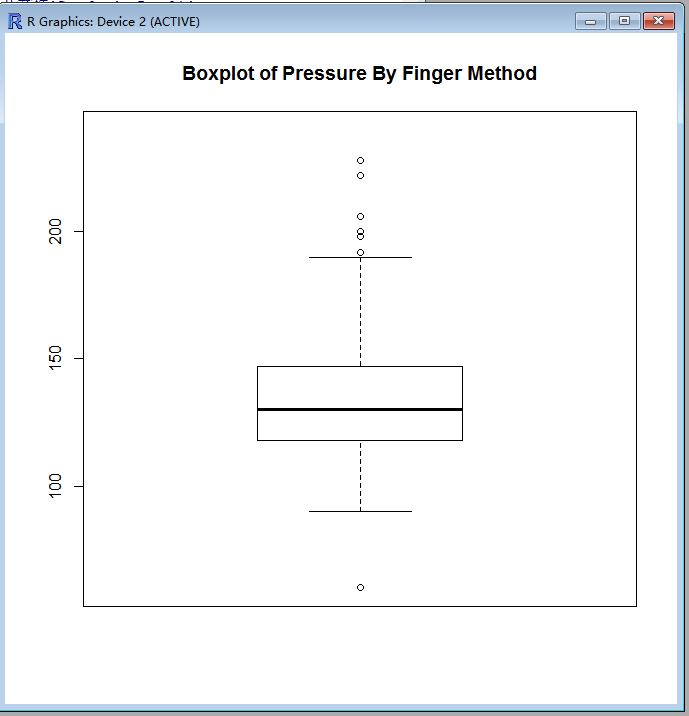


Fig 2.Boxplot of pressure by finger method

Fig1 and Fig2 show the boxplots of two methods. From above we can see that the two figures looks very similar, but there is some miner differences, for example, the finger method looks larger a bit than arm method. We can see the five number summaries;

Arm:

Number sum for arm method: 79 111 125 140 220

Finger:

Number sum for fin method: 60 118 130 147 228

From these two we see, in the arm method, nearly every number is larger than the finger method, that’s the difference showed in the figures above.

**(b)**

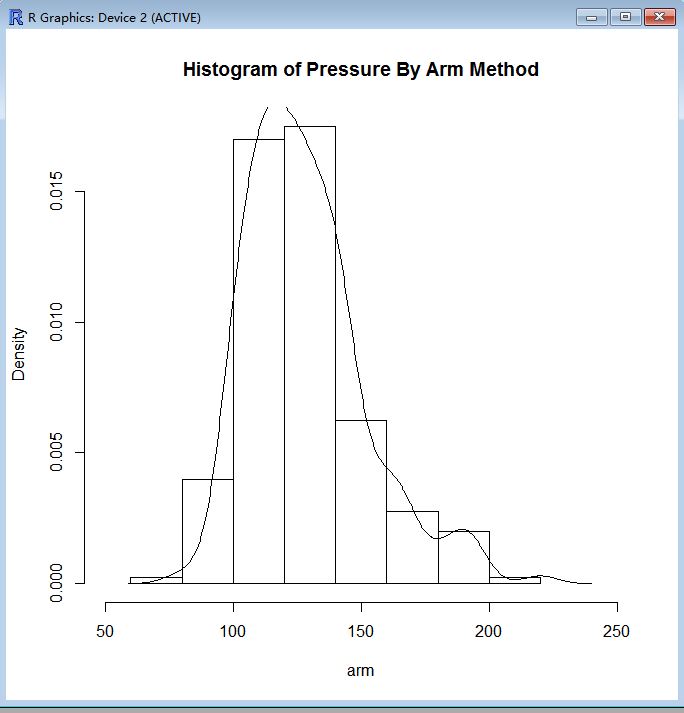


Fig.3 Histogram of Blood Pressure by arm method

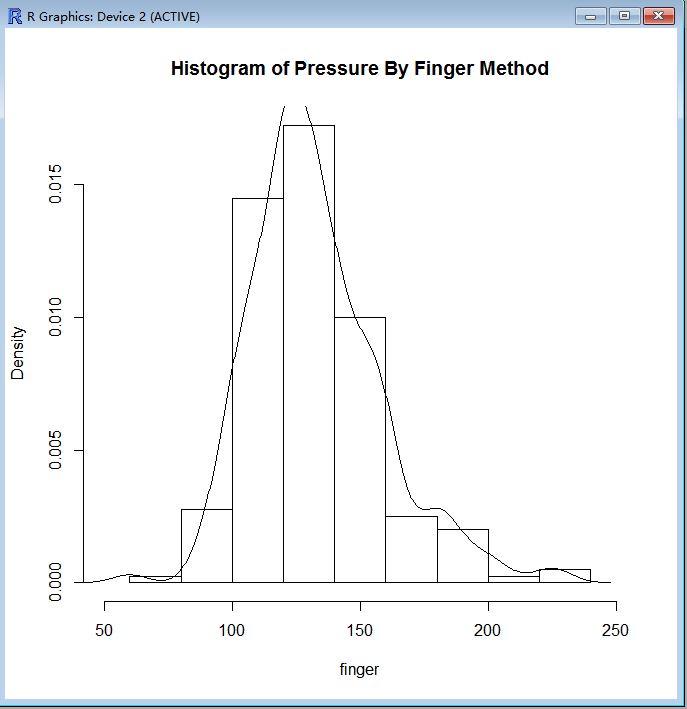


Fig.4 Histogram of Blood Pressure by finger method

Figure 3 and 4 show the histogram of two blood pressure measurements. The curve show the density. We can see that, very same with the box plots, the histogram between these two method are very same. In the fig 3, the curve is more left than fig.4, it can be explained by the data shown in (a) that finger method is larger than the arm method.

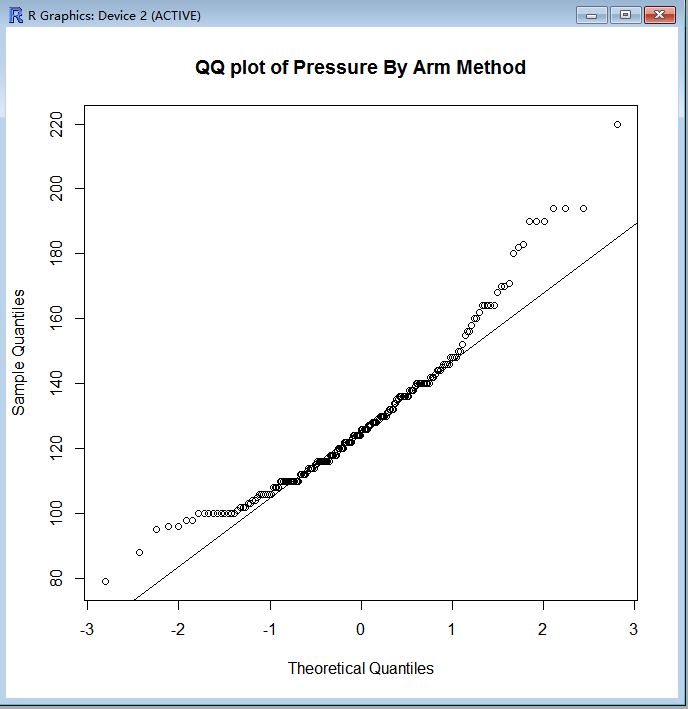


Fig.5 QQ plots of blood Pressure by arm method

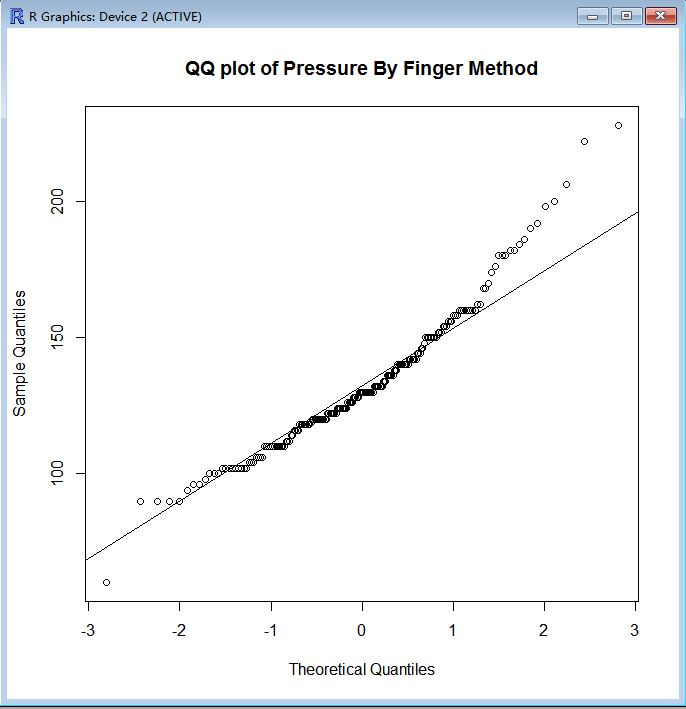


Fig.6 Q-Q plots of blood Pressure by finger method

Fig 5 and 6 show the Q-Q plots of two methods. From the figure we can see that in some part of it fits the normal line but in the tail and start, there are some difference, so we can’t say that this sample fits a normal distribution.

#inport data

>bp.data=read.table("G:\\bp.csv",header=T,sep=",")

#draw

> arm=bp.data$armsys

> finger=bp.data$fingsys

> boxplot(arm,ylim=c(60,240),main="Boxplot of Pressure By Arm Method")

> boxplot(finger,ylim=c(60,240),main="Boxplot of Pressure By Finger Method")

> cat("Number sum for arm method:",fivenum(arm),"\n")

Number sum for arm method: 79 111 125 140 220

> cat("Number sum for fin method:",fivenum(finger),"\n")

Number sum for fin method: 60 118 130 147 228

>

> hist(arm,,freq=FALSE,xlim=c(50,250),main="Histogram of Pressure By Arm Method")

> hist(finger,,freq=FALSE,xlim=c(50,250),main="Histogram of Pressure By Finger Method")

> hist(arm,,freq=FALSE,xlim=c(50,250),main="Histogram of Pressure By Arm Method")

> lines(density(arm))

> lines(density(finger))

> hist(finger,,freq=FALSE,xlim=c(50,250),main="Histogram of Pressure By Finger Method")

> lines(density(finger))

> qqnorm(arm,main="QQ plot of Pressure By Arm Method")

> qqline(arm)

> qqnorm(finger,main="QQ plot of Pressure By Finger Method")

> qqline(finger)

**(c)**

(c)95% confidence interval is [-9.096194, 0.5061939], which covers 0. We can conclude that the two methods have identical means. The assumption I made is that the size of two samples are large enough for the interval to be accurate because the size, 200 is large.

> mean(x[['armsys']])-mean(x[['fingsys']])-qnorm(0.975)\*sqrt(var(x[['armsys']])/200+var (x[['fingsys']])/200)

[1] -9.096194

> mean(x[['armsys']])-mean(x[['fingsys']])+qnorm(0.975)\*sqrt(var(x[['armsys']])/200+var (x[['fingsys']])/200)

[1] 0.5061939

**(d)**

1. The hypotheses would be:

P-value is 0.0795 which is bigger than 0.05. Therefore, at 5% level of significance, there is no statistically significant difference in the two means. The assumption I made is that the size of two samples are large enough for the interval to be accurate because the size, 200 is large.

> thetastar = mean(x[['armsys']])-mean(x[['fingsys']])

> varstar = var(x[['armsys']])/200+var(x[['fingsys']])/200

> z = thetastar/sqrt(varstar)

> z

[1] -1.753323

> pval = 2\*(1-pnorm(abs(z)))

> pval

[1] 0.07954652

**(e)**

Results from (c) and (d) are consistent. From (c), confidence interval is computed by using . From (d), by comparing and p-value, we get conclusion. Since the value of is the same, we can draw the same conclusion.

**Exercise 2**

**(a)**

The hypotheses would be:

**(b)**

I use t-test. Test statistic is .

**(c)**

-1.974186

> mean = 9.02

> std = 2.22

> x0 = 10

> t = (mean - x0)/(std/sqrt(20))

> t

[1] -1.974186

**(d)**

0.9684606

> pval =1 - pt(t,20-1)

> pval

[1] 0.9684606

**(e)**

The simulation result is 0.96816, which is almost the same as the result of (d).

> count = 0

> for(i in 1:100000){

+ x = rt(1,20-1)

+ if(x>t)

+ count = count+1

+ }

> count/100000

[1] 0.96816

**(f)**

P-value is 0.9684606. 95% confidence interval is [7.981008, 10.058992]. Thus, there is no statistically significant evidence to claim that the mean is greater than 10.

> mean + c(-1,1)\*qt(1-0.05/2,20-1)\*std/sqrt(20)

[1] 7.981008 10.058992

**Exercise 3**

**(a)**

The hypotheses would be:

We need to do t-test for these hypotheses. But we need to decide whether or not make the equal variance assumption. For this, we test the hypotheses:

> n.x = 400

> n.y = 500

> mean.x = 2635

> mean.y = 2887

> std.x = 365

> std.y = 412

> stat <- std.x^2/std.y^2

> stat

[1] 0.7848584

> pval <- 2\*min(pf(stat, n.x - 1, n.y - 1), 1 - pf(stat, n.x - 1, n.y - 1))

> pval

[1] 0.01136492

The small p-value for the F-test suggests that the two population variances cannot be assumed to be equal. Therefore, we will use Satterthwaite's approximation for the test of the means.

> tstat <- (mean.x - mean.y)/sqrt( (std.x^2/n.x) + (std.y^2/n.y))

> tstat

[1] -9.717132

> df.satterth.approx <- function(n.x, n.y, s.x, s.y) {

+ num <- ((s.x^2/n.x) + (s.y^2/n.y))^2

+ denom <- (s.x^4/((n.x^2 \* (n.x - 1)))) + (s.y^4/(n.y^2 \* (n.y - 1)))

+ return(num/denom)

+ }

> df.est <- df.satterth.approx(n.x, n.y, std.x, std.y)

> df.est

[1] 888.674

> pval <- 2\*(1-pt(abs(tstat), df.est))

> pval

[1] 0

The small p-value suggest that there is statistically significant difference in the means.

**(b)**

The hypotheses would be:

We need to do t-test for these hypotheses. The small p-value for the F-test suggests that the two population variances cannot be assumed to be equal. Therefore, we will use Satterthwaite's approximation for the test of the means, as (a).

> pval2 <- pt(tstat, df.est)

> pval2

[1] 1.395927e-21

The small p-value suggest that there is statistically significant evidence that the mean credit limit of all credit cards issued in May 2011 is greater than the same in January 2011.