Mini Project 5

Yanjiao Jia (Contribution 50%)

Yuzhuang Feng (Contribution 50%)

Object

In the project, we finish three tasks, which is:

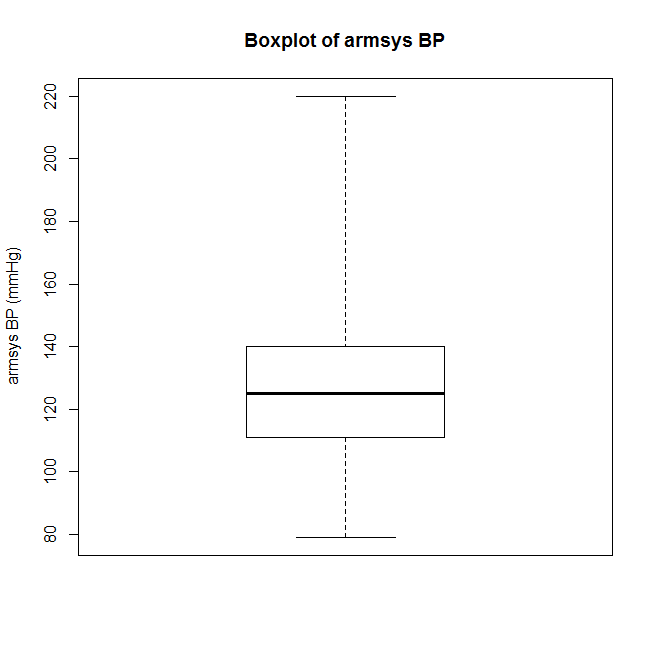
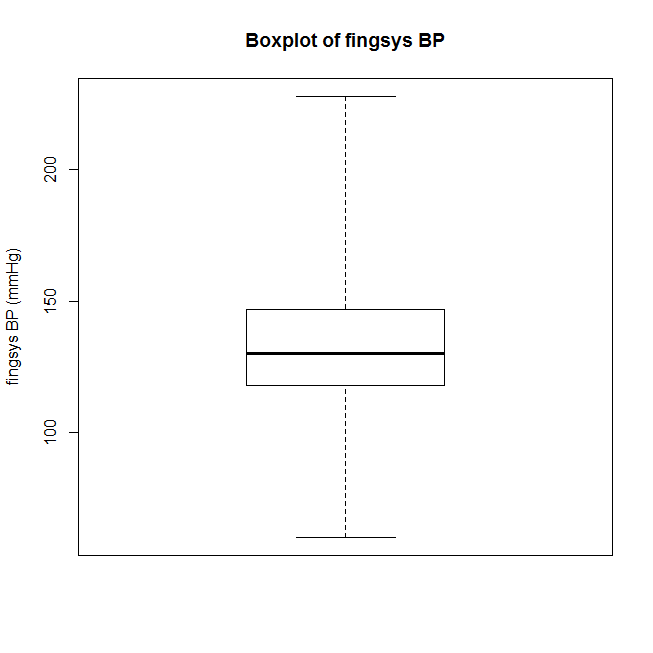
1. Exam if there is difference between the finger method and the arm method measuring systolic blood pressure (in mmHg).
2. Exam if the mean of a normal population is greater than 10.
3. Exam if credit limits on newly issued credit cards increased between January 2011 and May 2011.

In the exam process, we used boxplots, qqplots, histograms, 95% confidence interval and 5% level test.

Procedure and Analysis

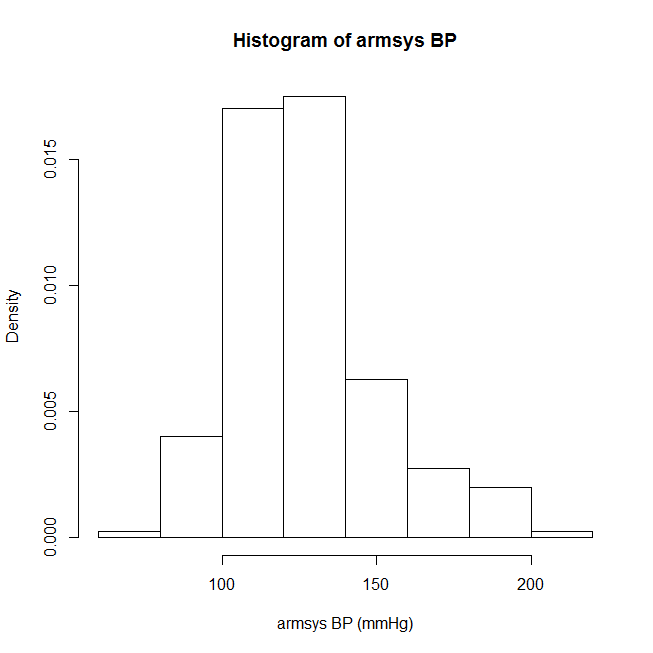
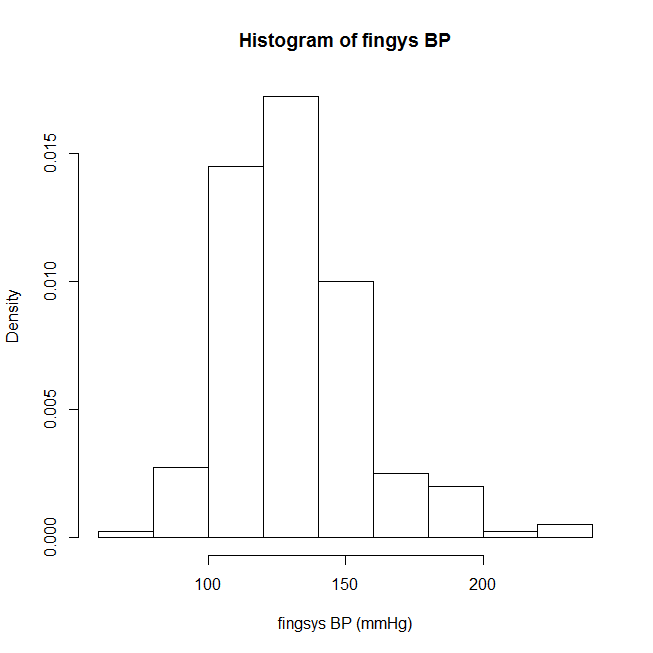
Exercise 1

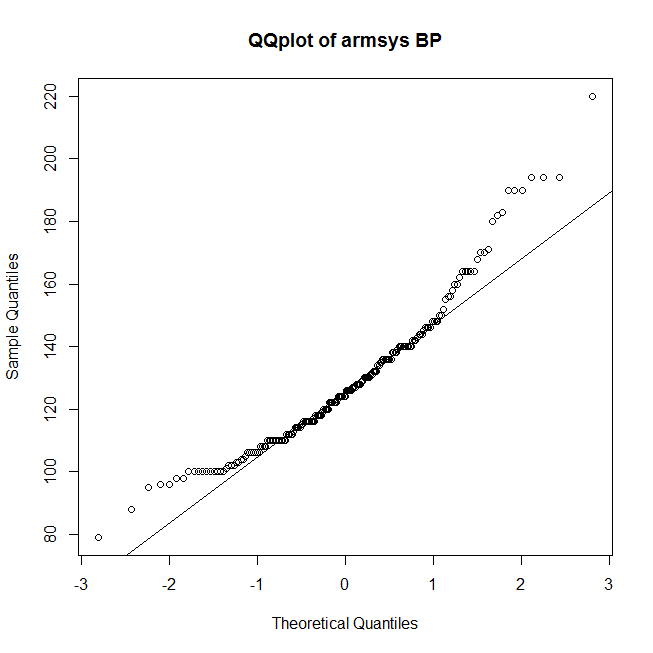
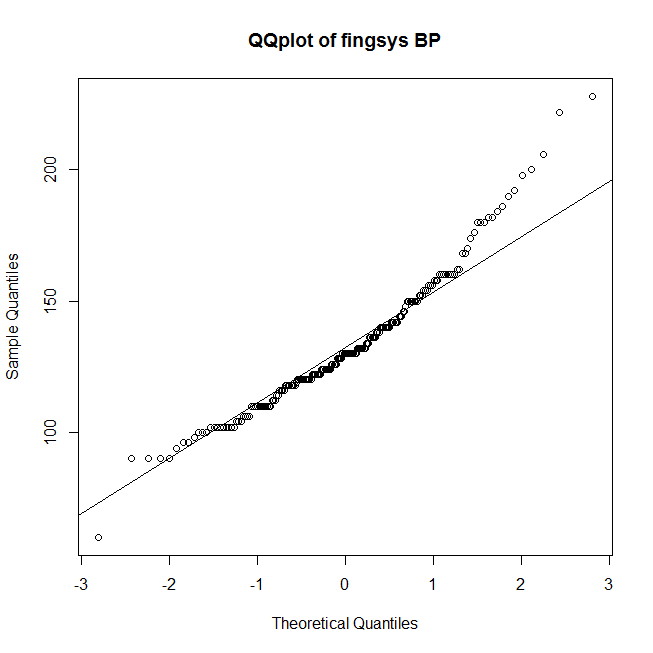
a. We made boxplots to examine the distributions of the measurements from two methods—a finger method and an arm method.

From above two boxplots , we can see these two distribution are similar, but they are still different. The first one for armsy BP is a little bit right-screwed and the boxplot of fingsys BP is almost symmetric.

b. Then we made histograms and qqplots of these two datas.

From histograms and qqplots , we can see their normality seem reasonable. The first one seems like right-screwed and the second one seems symmetric.

c. Then we constructed an appropriate 95% confidence interval for the difference in the means of the two method which is: [-6.316529, -2.273471]*.*

So Their means are not identical. The data of the armsys BP is smaller than the data of fingsys BP. We assumed that n is large enough, because n is 200.

d. Then we performed an 5% level t-test to see if there is any difference in the means of the two methods. We made H0 be armsys-fingsys equal to 0 and H1 armsys-fingsys is not equal to 0. And the result is:

***One Sample t-test***

***data: deltabp***

***t = -4.1642, df = 199, p-value = 4.652e-05***

***alternative hypothesis: true mean is not equal to 0***

***95 percent confidence interval:***

***-6.328898 -2.261102***

***sample estimates:***

***mean of x***

***-4.295***

Since p=4.652e-05 is much less than α=0.05 ,we rejected H0 which means not equal.

e. The result from c and d seems consistent, both shows the data got from the two methods are different.

Exercise 2

We need to test the null hypothesis that the mean of a normal population is 10 against the alternative that it is greater than 10. And we knew the size of a random sample and its mean and its sample standard deviation.

a. First, we need to set up the null and alternative hypotheses. We made H0 be mean equal to 10 and H1 be mean not equal to 10.

b. We chose to use t-test. The test statistic is the 20 examples. And the null distribution of the test statistic is t(20-1) distribution.

c. We computed the observed value of the test statistic using “*tobs<-(mean-10)/(s/sqrt(20*))”, the observed value of the test statistic is: -1.974186

d. We computed the p-value of the test using the usual way *“pvalue<-1-pt(tobs,19)”* , the p-value of the test using the usual way is: 0.9684606

e. We estimated the p-value of the test using Monte Carlo simulation. Use rt function to make random draws from the t(20-1) distribution, then calculate the percentage of these draws is greater than tobs, which is the estimation of the p-value. The p-value of the test using Monte Carlo simulation is: 0.963964. The result of (d) and € are very close.

f. Here, p= 0.968 is larger than α=0.05. So we accepted H0 which means the mean of a normal population is 10.

Exercise 3

a. We constructed an appropriate 95% confidence interval for the difference in mean credit limits of all credit cards issued in January 2011 and in May 2011 using the average credit limits and the sample standard deviations of these two samples. The 95% confidence interval for the difference is: [-302.8289, -201.1711]. This means the mean credit limit of all credit cards is increased.

b. We performed an appropriate 5% level test to see if the mean credit limit of all credit cards issued in May 2011 is greater than the same in January 2011. And we made H0 be Jan-May=0 be and H1 be Jan-May<0. The p-value of the test is: 1.395927e-21

Since p is much less than α=0.05, we rejected H0, which means the mean credit limit of all credit cards issued in May 2011 is greater.

Code

#Exercise 1

cat("Exercise 1\n")

#read the bp.txt file

bp<-read.table("bp.txt", fill = FALSE, sep = "",header = T)

#a.make boxplots of these two data

boxplot(bp$armsys,ylab='armsys BP (mmHg)',main='Boxplot of armsys BP',range = 0)

boxplot(bp$fingsys,ylab='fingsys BP (mmHg)',main='Boxplot of fingsys BP',range = 0)

#b.make histograms of these two data

hist(bp$armsys,xlab='armsys BP (mmHg)',main='Histogram of armsys BP',freq = F)

hist(bp$fingsys,xlab='fingsys BP (mmHg)',main='Histogram of fingys BP',freq = F)

# make qqplot of these two data

qqnorm(bp$armsys,main='QQplot of armsys BP');qqline(bp$armsys)

qqnorm(bp$fingsys,main='QQplot of fingsys BP');qqline(bp$fingsys)

#c.Construct an appropriate 95% confidence interval for the difference

# in the means of these two data

deltabp<-bp$armsys-bp$fingsys

sd<-sd(deltabp)

mean<-mean(deltabp)

CI<-(mean+c(-1,1)\*qnorm(1-0.025)\*sd/sqrt(length(deltabp)))

cat("95% confidence interval for the difference is: ",CI,"\n")

#d.Perform an appropriate 5% level test to see if there is any difference

# in the means of these two data

#H0:armsys=fingsys

#H1:armsys!=fingsys

cat("5% level test is:\n")

print(t.test(deltabp, alternative = "two.sided", mu = 0, conf.level = (1 - 0.05)))

#Exercise 2

cat("\nExercise 2\n")

#a.set up the null and alternative hypotheses

#H0:mean=10

#H1:mean>10

#b.choose t-test

#c.compute the observed value of the test statistic

mean<-9.02

s<-2.22

tobs<-(mean-10)/(s/sqrt(20))

cat("the observed value of the test statistic is: ",tobs,"\n")

#d.compute the p-value of the test using the usual way

pvalue<-1-pt(tobs,19)

cat("the p-value of the test using the usual way is: ",pvalue,"\n")

#e.estimate the p-value of the test using Monte Carlo simulation

MC<-rt(999,19)

pvalueofMC<-length(which(MC>tobs))/999

cat("the p-value of the test using Monte Carlo simulation is: ",pvalueofMC,"\n")

#Exercise 3

cat("\n\nExercise 3\n")

#a.construct an appropriate 95% confidence interval for the difference

# in mean credit limits

CI<-((2635-2887)+c(-1,1)\*(qnorm(1-0.025)\*sqrt(365^2/400+412^2/500)))

cat("95% confidence interval for the difference is: ",CI,"\n")

#b.Perform an appropriate 5% level test to see if the mean credit limit

# of all credit cards issued in May 2011 is greater than the same in January 2011.

#delta=Jan-May

#H0: delta=0

#H1: delta<0

tobs<-(2635-2887)/sqrt(365^2/400+412^2/500)

dof<-(365^2/400+412^2/500)^2/(365^4/(400^2\*(400-1))+412^4/(500^2\*(500-1)))

pvalue<-pt(tobs,dof)

cat("the p-value of the test is: ",pvalue,"\n")