#######################################

Name: Raj Shah

PSID: 1499521

Email:rpshah6@uh.edu

#######################################

Prob1)

> library(readr)

> normal<-function(n,mean,sd){

+ dis<-rnorm(n,mean = mean,sd = sd)

+ mean.diff<-abs(mean-mean(dis))

+ var.diff<-abs(sd^2-var(dis))

+ med.diff<-abs(qnorm(.5,mean = mean,sd = sd)-median(dis))

+ quant.diff<-abs(quantile(dis,c(.99))-qnorm(.99,mean = mean,sd = sd))

+ prob<-pnorm(median(dis),mean = mean,sd = sd)

+ result<-matrix(c(mean.diff,var.diff,med.diff,quant.diff,prob),ncol = 5)

+ colnames(result)<-c("Mean.diff","Var.diff","Med.diff","Q99.dif.99%","Prob.Val")

+ result<-as.table(result)

+ return(result)

+ }

>

> normal(20,4,1)

Mean.diff Var.diff Med.diff Q99.dif.99% Prob.Val

A 0.2416719 0.5012475 0.1679463 0.1124304 0.5666872

> for (n in c(10^2,10^4,10^6))

+ print(normal(n,6,2))

Mean.diff Var.diff Med.diff Q99.dif.99% Prob.Val

A 0.19835460 0.35859701 0.29177784 0.07694758 0.55799546

Mean.diff Var.diff Med.diff Q99.dif.99% Prob.Val

A 0.004313575 0.085593107 0.005288361 0.059288380 0.501054874

Mean.diff Var.diff Med.diff Q99.dif.99% Prob.Val

A 0.001773266 0.014997628 0.002185673 0.006424910 0.499564021

As the number of samples increases,the differences for mean,variance,Median and 99% Quantile,decrease.

Also the probability value decreases with the increase in number of samples.

Prob2)

> my\_test <- function(vec,mu,ci\_lvl=.95){

+ t<-(mean(vec)-mu)/(sd(vec)/sqrt(length(vec)))

+ df<-length(vec)-1

+ p\_val<-2\*pt(-abs(t),df = df)

+ error<-qt((1-((1-ci\_lvl)/2)),df =df)\*sd(vec)/(sqrt(length(vec)))

+ CI\_low<-mean(vec)-error

+ CI\_high<-mean(vec)+error

+ sam\_est<-mean(vec)

+ result<-matrix(c(t,df,CI\_low,CI\_high,p\_val,sam\_est),ncol=6)

+ colnames(result)<-c("t.val","df.val","CI,low","CI.high","p.val","Sample.estimates")

+ result<-as.table(result)

+ return(result)

+ }

> my\_test(c(1:10),mu = 5,ci\_lvl = 0.95)

t.val df.val CI,low CI.high p.val Sample.estimates

A 0.5222330 9.0000000 3.3341494 7.6658506 0.6141173 5.5000000

> t.test(c(1:10),mu=5,conf.level = .95)

One Sample t-test

data: c(1:10)

t = 0.52223, df = 9, p-value = 0.6141

alternative hypothesis: true mean is not equal to 5

95 percent confidence interval:

3.334149 7.665851

sample estimates:

mean of x

5.5

> my\_test(c(10:100),mu=50,ci\_lvl = .99) #Confidence Level = 99%

t.val df.val CI,low CI.high p.val Sample.estimates

A 1.80578780 90.00000000 47.71352600 62.28647400 0.07429463 55.00000000

> t.test(c(10:100),mu=50,conf.level = .99)

One Sample t-test

data: c(10:100)

t = 1.8058, df = 90, p-value = 0.07429

alternative hypothesis: true mean is not equal to 50

99 percent confidence interval:

47.71353 62.28647

sample estimates:

mean of x

55

Prob3)

> library(readr)

>

> computers <-read.csv("C:/Users/Raj Shah/Downloads/Studies UH/MATH 6359 Stastical Computing/Homework/Answers/HW2/data2.csv",header=TRUE,sep=",")

> par(mfrow=c(2,2))

>

> attach(computers)

> Harddisk<-computers$hd

> Price<-computers$price

>

> hist(Harddisk,breaks = 20,main = "Hard Disk for PC",col="red",labels = TRUE)

> hist(Price,breaks = 20,main = "Price for PC",col="green",labels = TRUE)

>

> qqnorm(Harddisk,main="Q-Q plot of Hard Disk for PC")

> qqnorm(Price,main="Q-Q plot of Price for PC")

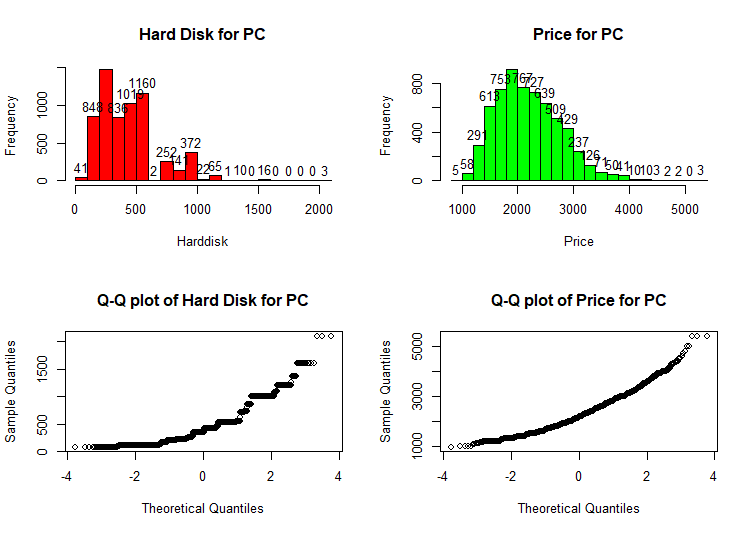


Fig1. Histogram and Q-Q plot for the data

> par(mfrow=c(1,1))

> boxplot(Harddisk,Price,ylab="Personal Computers",col=2:3,ylim=c(100,4000),names = c("Hard Disk","Price"))

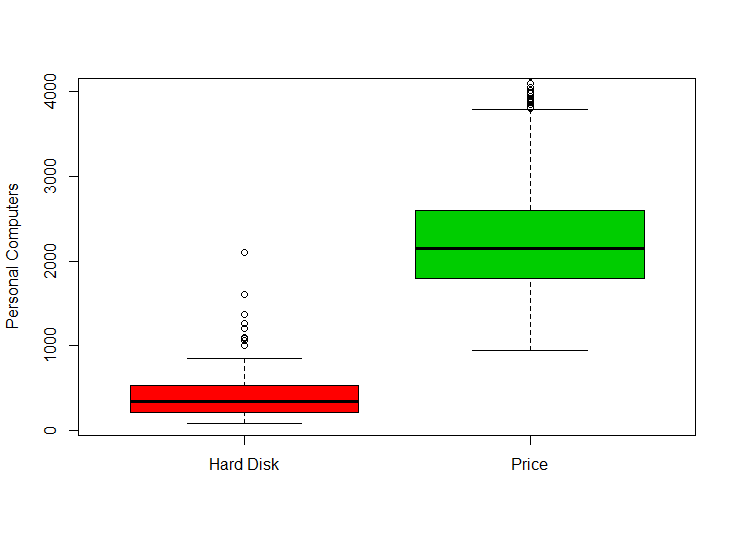


Fig 2. Box plot for the data set.

> t.test(Harddisk,Price)

Welch Two Sample t-test

data: Harddisk and Price

t = -224.36, df = 8644.5, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-1818.727 -1787.223

sample estimates:

mean of x mean of y

416.6017 2219.5766

> wilcox.test(Harddisk,Price)

Wilcoxon rank sum test with continuity correction

data: Harddisk and Price

W = 32897, p-value < 2.2e-16

alternative hypothesis: true location shift is not equal to 0

**Summary**