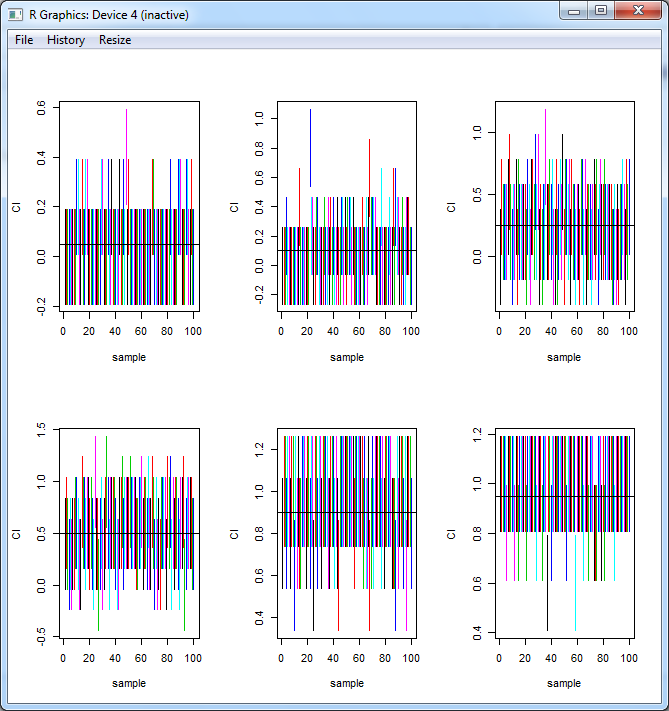
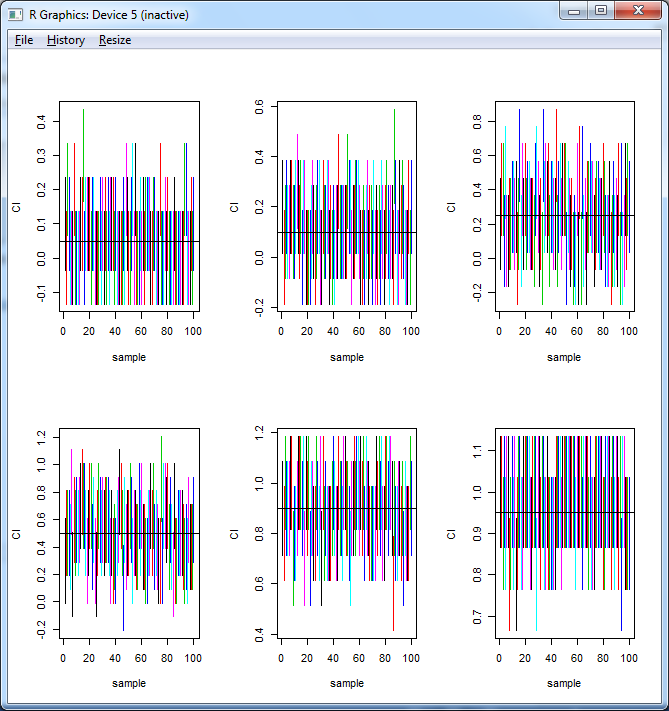
Problem 1:

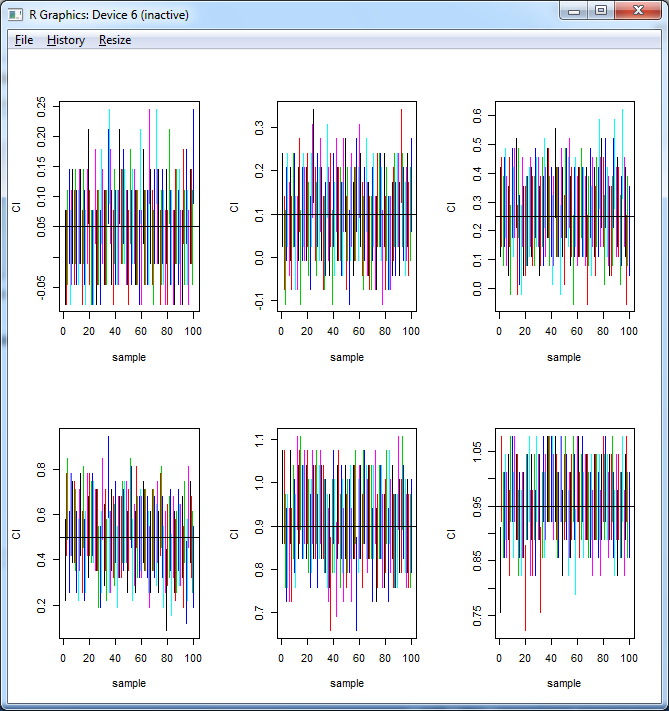
For n = 5



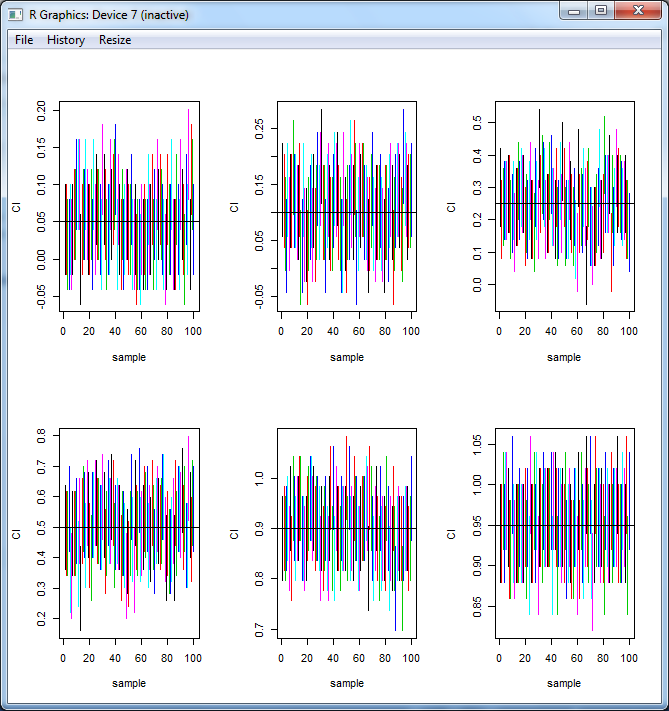
For N =10



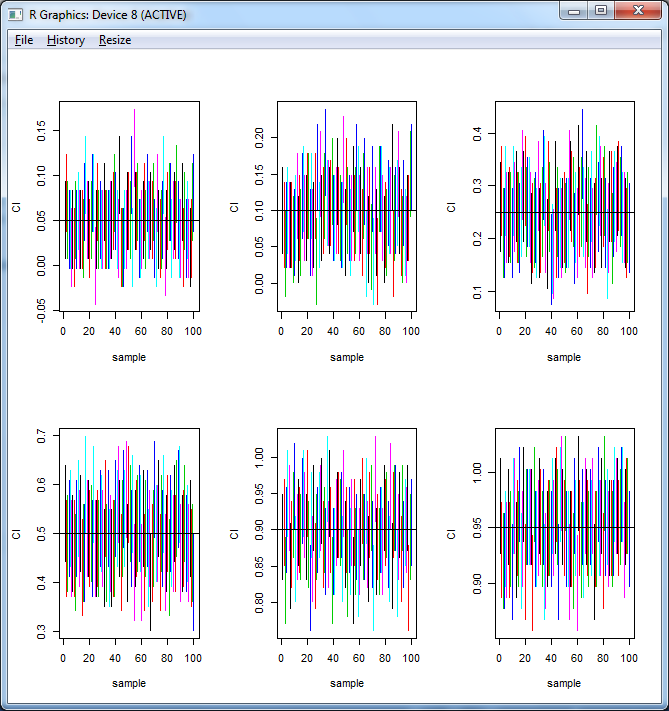
For N=30



For N=50



For N=100



As we see from the different graphs obtained above , For same probability value if we increase the value of N , the width of confidence interval decreases which actually shows that the accuracy increases with the increase in the number of N. For example for probability value : 0.95 the width varies as ,

|  |  |
| --- | --- |
| N | Width |
| 5 | 0.10 |
| 10 | 0.30 |
| 30 | 0.20 |
| 50 | 0.10 |
| 100 | 0.05 |

The same pattern is show as we vary N for different value of p. So for good value for this confidence interval is 100.

With change in value of P if we see one graph then width is less as the probability value as it closer to 0 or 1 , but for all the graphs the width is high when probability value is 0.5 .This is because for the probability value 0.5 the value p\*(1-p) is maximum which is 0.25 and we need More samples here to get good CI for same error , using the formula of CI in monte carlo.

Problem 2:

1. Does it seem reasonable to assume that each sample comes from a normal distribution? Draw Q-Q plots to answer this question

Command to draw Q-Q plots:

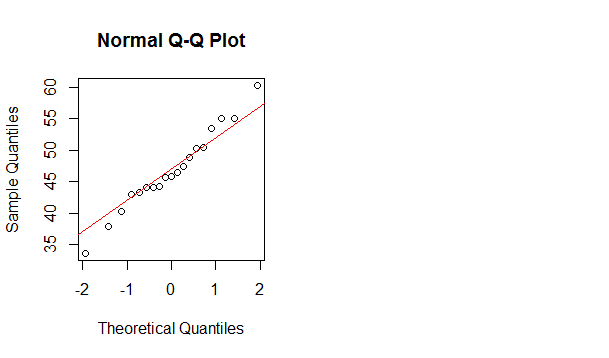
qqnorm(child)

qqline(child, col = "red")

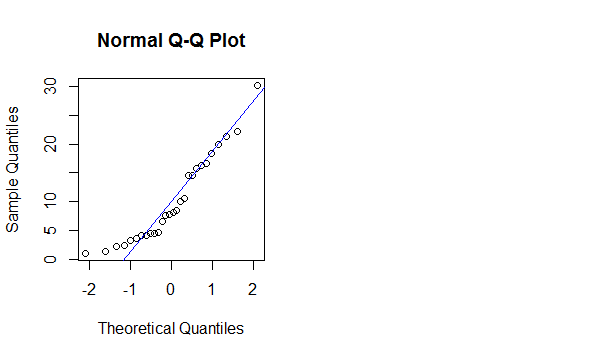
qqnorm(adult)

qqline(adult,col = "blue")

For child Cereals we get in QQPLOT the points are very close to line so we can assume that this is a normal distribution.



For Adult Cereals we get in QQPLOT the points are far from the line so we can’t assume this distribution to be normal.



1. Can the variances of the two distributions be assumed to be equal? Justify your answer.

Using 95% of confidence we apply F-distribution for ratio of variances.

f.l.crit <- qf(alpha/2, n.x - 1, n.y - 1)

f.u.crit <- qf(1 - (alpha/2), n.x - 1, n.y - 1)

((sd(child)/sd(adult))^2) \* c(1/f.u.crit, 1/f.l.crit)

The confidence interval for variance child / variance Adult : [**0.3102977 1.7564758**]

Since the ratio contains the value 1 in the interval so we can assume that the variances of both the distribution can be equal.

1. Since we assume that variance of both distributions are equal we can use pooled variance to calculate the confidence interval of meanChild – meanAdult.

mean(child) - mean(adult) + c(-1, 1) \* qt(1 - (alpha/2), df = n.x + n.y -2 ) \* sqrt(pooled\_variance\*((1/n.x) + (1/n.y)))

Confident Interval of **(meanChild – meanAdult)** = [32.35553 40.92680]

1. SO the answer for confidence interval for meanChild – meanAdult clearly states the fact that child cereals have more sugar than Adults Cereals.

And child cereals have sugar almost greater than 35 units by adult cereals.

3. We calculate the confidence interval for px-py using 95% CI and we get the interval as :

**[-0.04652425 0.04580106]**

The code is in Appendix Section.

1. Since the Confidence interval contains the value 0 in the interval so we can say that there is no difference in abused children for single parent household and two parent household.
2. We assumed that the distribution is normal for both the distribution. So the difference of these two distributions also follows normal distribution.

Appendix:

1. Program:

conf.init<-function(pcap,n,se.pcap){

pcap = rbinom(n,1,pcap)

ci<- mean(pcap) + c(-1, 1) \* qnorm(1 - (alpha/2)) \* se.pcap

return(ci)

}

pc = c(0.05, 0.1, 0.25, 0.5, 0.9, 0.95)

#pc = c(0.05)

n = c(5, 10, 30, 50, 100)

for (ni in n){

dev.new()

par(mfrow=c(2,3))

for(pi in pc) {

n=ni; pcap=pi

alpha = 0.05

se.pcap = sqrt(pcap\*(1-pcap)/n) # compute SE

conf.int(pcap,n,se.pcap)

nsim<-100

ci.mat<-replicate(nsim,conf.int(pcap,n,se.pcap))

mean(pcap)

matplot(rbind(1:100,1:100),type="l",lty=1,ci.mat[,1:100],xlab="sample",ylab="CI")

abline(h=pcap)

mean((pcap>=ci.mat[1,])\*(pcap<=ci.mat[2,]))

}

}

1. Program:

child<-c(40.3, 55, 45.7, 43.3, 50.3, 45.9, 53.5, 43, 44.2, 44, 47.4, 44, 33.6, 55.1,

48.8, 50.4, 37.8, 60.3, 46.5)

adult<-c(20, 30.2, 2.2, 7.5, 4.4, 22.2, 16.6, 14.5, 21.4, 3.3, 6.6, 7.8, 10.6, 16.2,

14.5, 4.1, 15.8, 4.1, 2.4, 3.5, 8.5, 10, 1, 4.4, 1.3, 8.1, 4.7, 18.4)

par(mfrow=c(1,2))

qqnorm(child)

qqline(child, col = "red")

qqnorm(adult)

qqline(adult,col = "blue")

alpha<-.05

mean(child)

mean(adult)

n.x<-length(child)

n.y<-length(adult)

f.l.crit <- qf(alpha/2, n.x - 1, n.y - 1)

f.u.crit <- qf(1 - (alpha/2), n.x - 1, n.y - 1)

((sd(child)/sd(adult))^2) \* c(1/f.u.crit, 1/f.l.crit)

pooled\_variance<-((n.x-1)\* var(child) + (n.y-1)\*var(adult))/(n.x+n.y-2)

mean(child) - mean(adult) + c(-1, 1) \* qt(1 - (alpha/2), df = n.x + n.y -2 ) \* sqrt(pooled\_variance\*((1/n.x) + (1/n.y)))

1. Program:

px <- 61/414

py <- 74/501

m <- 414

n <- 501

phat <- px - py

alpha <- 0.05

SE <- sqrt( ((px\*(1-px))/m) + ((py\*(1-py))/n))

CI <- phat + qnorm(1-(alpha/2))\*SE\*c(-1,1)

CI