

Mini Project: #3
Name: Ajay Vembu

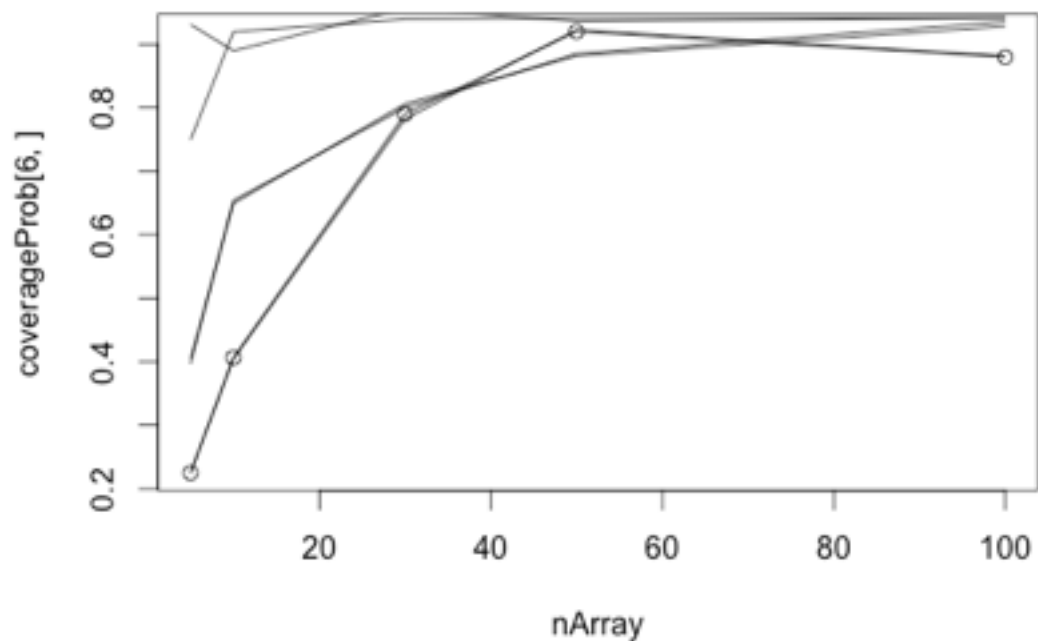
Excecise - 1:

The below is the coverage probability for the various N's and P's

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	0.2256	0.4042	0.7820	0.9226	0.8812
[2,]	0.3967	0.6541	0.7998	0.8851	0.9352
[3,]	0.7492	0.9192	0.9394	0.9408	0.9429
[4,]	0.9308	0.8896	0.9561	0.9362	0.9408
[5,]	0.4049	0.6491	0.8060	0.8813	0.9280
[6,]	0.2249	0.4065	0.7898	0.9198	0.8794

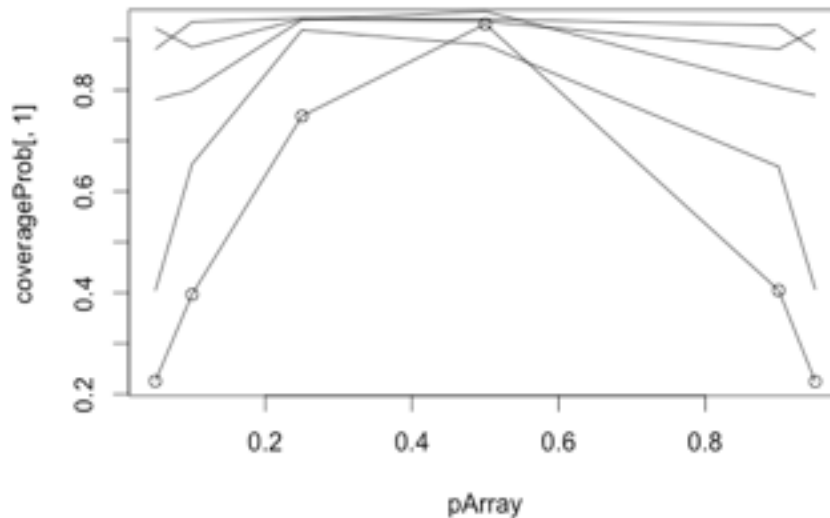
The rows are the p's with values 0.05,0.2,0.5,0.75,0.9,0.95 and columns are the n's with values 5,10,20,30,50,75,100,1000,10000 respectively.

The below are the plots between n and coverage probability for all the probabilities,



This shows that as n increases the proportion of the confidence interval covering the true value increases and approaches 0.95. Based on the experiment conducted I would recommend an n value of 50 as the coverage probability for all the probabilities converge here.

Below is the plot between the p values and coverage probabilities for the constant n values,



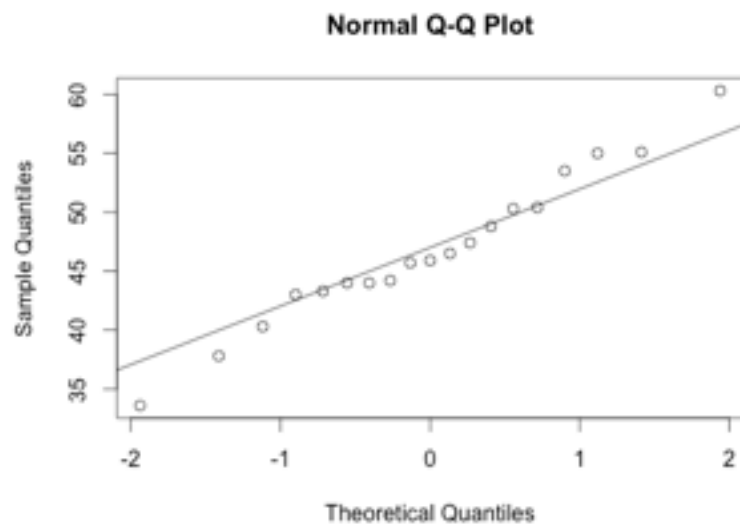
Bases on the experiment conducted as the p approaches 0.5 all the coverage probability reaches the maximum that is the 0.95. That is as p increases coverage probability also increases up until 0.5 and thereafter decreases. This shows that the coverage probability is the maximum at $p = 0.5$.

With the results of the above experiments in hand I would recommend an **$N = 50$** and this also depends on **P which should be 0.5 to get a maximum coverage probability**.

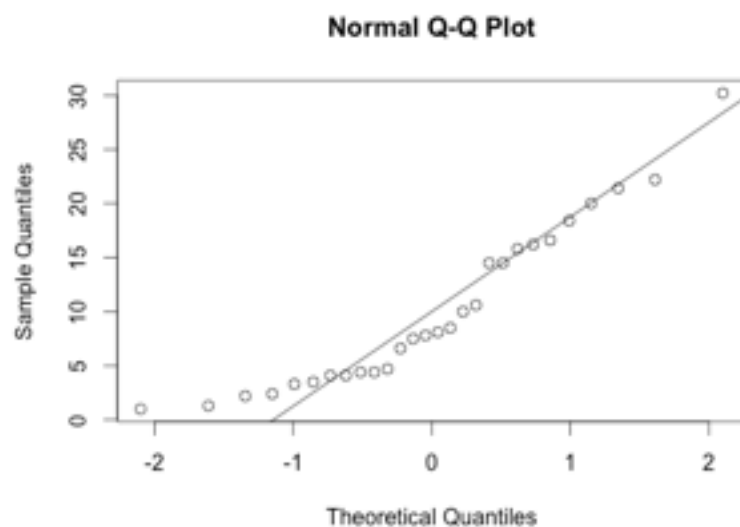
Exercise - 2:

Part - a:

The below is the qqplot for the children data,



Below is the qqplot for the adult,



The above two plots shows that the distribution of the two samples are from a normal distribution as the quantiles of the standard normal distribution is in a linear relationship with the quantiles of the samples.

Part - b:

The confidence interval of the ratio of the two sample variance has been constructed using the F distribution and found to be

confidenceInterval = (0.2649826, 2.1005499)

This shows that equal variance can be assumed as the interval contains 1 in it i.e. $S_x = S_y$.

Part - c:

The confidence interval for the difference in mean is given by the following,

confidenceInterval = (5.961163 67.321168)

The assumption made is equal variance and hence pooled variance and degree of freedom as $N_x + N_y - 2$ has been used to compute the confidence interval.

Part - d:

From the above answer we can conclude that the cereals of the children has more sugar than the cereals of the adult, i.e as there is no negative values in the confidence interval.

This amount of sugar for children is more by the range (5.961163 67.321168)

Exercise - 3:

Part - a

The confidence interval for the two proportions is given as,

confidenceInterval = (-0.04652425,0.04580106)

As 0 is contained in the range, it shows that there is no difference between the single parent and two parent household when it comes to the child abuse.

Part - b:

The assumption comes from the central limit theorem that the proportions belong to the normal distribution, and hence applied Z - critical point to compute the confidence interval. This assumption seems reasonable as the n value for both the proportions are large.

R - code:

Ex- 1

```
pArray <- c(0.05,0.1,0.25,0.5,0.9,0.95)
nArray <- c(5,10,30,50,100)
alpha <- 1 - 0.95
# to maintain all the coverage prob
coverageProb <- matrix(nrow = length(pArray),ncol = length(nArray))

for (i in 1:length(pArray)){
  for (j in 1:length(nArray)){
```

```

# monte carlo simulation for 10000 different cases

totalResult <- replicate(10000,{
  allElements <- rbinom(n = nArray[j],size = 1,prob = pArray[i])
  # prob for proportions
  phat <- sum(allElements) / length(allElements)
  # confidence interval of proportions
  ci.prop <- phat + c(-1,+1) * qnorm(1 - alpha / 2) * sqrt((phat * (1 - phat)) / nArray[j])

  if (pArray[i] >= ci.prop[1] && pArray[i] <= ci.prop[2]){
    return(1)
  }
  else{
    return(0)
  }
})
# coverage prob
coverageProb[i,j] <- sum(totalResult) / length(totalResult)
cat("For P = ",pArray[i]," and N = ",nArray[j]," Coverage probability = ",coverageProb[i,j],"\\n")
}
}

# Plots for N and coverage prob
plot(nArray,coverageProb[1,])
lines(nArray,coverageProb[1,])
lines(nArray,coverageProb[2,])
lines(nArray,coverageProb[3,])
lines(nArray,coverageProb[4,])
lines(nArray,coverageProb[5,])
lines(nArray,coverageProb[6,])

# Plots for P and coverage prob

plot(pArray,coverageProb[,1])
lines(pArray,coverageProb[,1])
lines(pArray,coverageProb[,2])
lines(pArray,coverageProb[,3])
lines(pArray,coverageProb[,4])
lines(pArray,coverageProb[,5])

```

Ex - 2

Part - a

```

# to find if the sample is normal
#children
qqnorm (children); qqline(children)
# adult
qqnorm (adult); qqline(adult)

```

Part - b:

to find the sample variances are equal

```
children <- 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)
alpha <- 1 - 0.98
children.var <- var(children)
adult.var <- var(adult)
children.n <- length(children)
adult.n <- length(adult)
```

```
# upper critical
f.u.crit <- qf(1 - alpha / 2, children.n - 1, adult.n - 1)
# lower critical
f.l.crit <- qf(alpha / 2, children.n - 1, adult.n - 1)
```

```
#confidence interval for the ratio of variance
var.ci <- (children.var / adult.var) * (1 / c(f.u.crit, f.l.crit) )
```

```
print(var.ci)
```

Part - c:

```
children <- 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)
alpha <- 1 - 0.95
```

```
children.mean <- mean(children)
adult.mean <- mean(adult)
children.sd <- sd(children)
adult.sd <- sd(adult)
children.n <- length(children)
adult.n <- length(adult)
```

```
# degree of freedom
v <- children.n + adult.n - 2
```

```
# pooled variance
sp <- ( (children.n-1) * children.sd^2 + (adult.n - 1) * adult.sd^2 ) / v
```

```
# confidence interval for the diff in mean
ci.diff.mean <- (children.mean - adult.mean) + c(-1,+1) * qt (1 - alpha / 2, v) * sp * sqrt(1 / children.n + 1 / adult.n)
```

```
print(ci.diff.mean)
```

Ex - 3:

Part - a:

```
P1 <- 61 / 414
```

```
P2 <- 74 / 501
```

```
N1 <- 414
```

```
N2 <- 501
```

```
alpha <- 1 - 0.95
```

```
# CI for the two proportions
```

```
ci.prop <- (P1 - P2) + c(-1,+1) * qnorm(1 - alpha / 2) * sqrt( ((P1 * (1 - P1)) / N1 ) + ((P2 * (1 - P2)) / N2 ))
```

```
print(ci.prop)
```