STAT 231 Assignment 3

The purpose of this assignment is to use the software R to calculate confidence intervals, examine the behaviour of confidence intervals, and examine the sampling distribution of the likelihood ratio statistic. **The code for this assignment is posted both as a text file called RCodeAssignment3.txt and an R file called RCodeAssignment3R.R which are posted in the Assignment 3 folder in the Assignments folder under Content on Learn.**

Problem 1: Run the following R code.

```
######################################################################################
# Run this code only once
library(MASS)    # truehist is in the library MASS
######################################################################################

######################################################################################
# Problem 1: Binomial confidence intervals and sampling distribution of likelihood ratio statistic
id<-20456458
set.seed(id)
#   generate a random value of theta
theta<-rbeta(1, max(1,id-10*trunc(id/10)), max(1,trunc(id/10)-10*trunc(id/100)))
if (theta<0.1) {theta<-theta+0.1}   # avoid small values of theta
if (theta>0.9) {theta<-theta-0.1}   # avoid large values of theta
n<-30
cat("n = ",n," theta = ",theta)   # display values
# vector of observations for 5000 simulations from a Binomial(n,theta) distribution
yobs<- rbinom(5000,n,theta)
# corresponding vector of thetahat values
that<-yobs/n
# values used to construct an approximate 95% confidence interval  based on Gaussian approximation
pm<-1.96*sqrt(that*(1-that)/n)
# each approximate 95% confidence interval is stored in a row of matrix cibi
cibi<-matrix(c(that-pm,that+pm),nrow=5000,byrow=F)
cibi[1:10,1:2]        # Look at first 10 approximate 95% confidence intervals
# display proportion of approximate 95% confidence intervals which contain true value of theta
prop<- mean(abs(theta-that)<pm)
cat("proportion of approximate 95% confidence intervals which contain true value of theta = ",prop)
#
# create function to calculate Binomial relative likelihood function
BinRLF <- function(x) {dbinom(y,n,x)/dbinom(y,n,thetahat)}
li<-rep(0,2*5000)
li<- matrix(li,ncol=2,byrow=TRUE)              # initialize matrix to store likelihood intervals
```

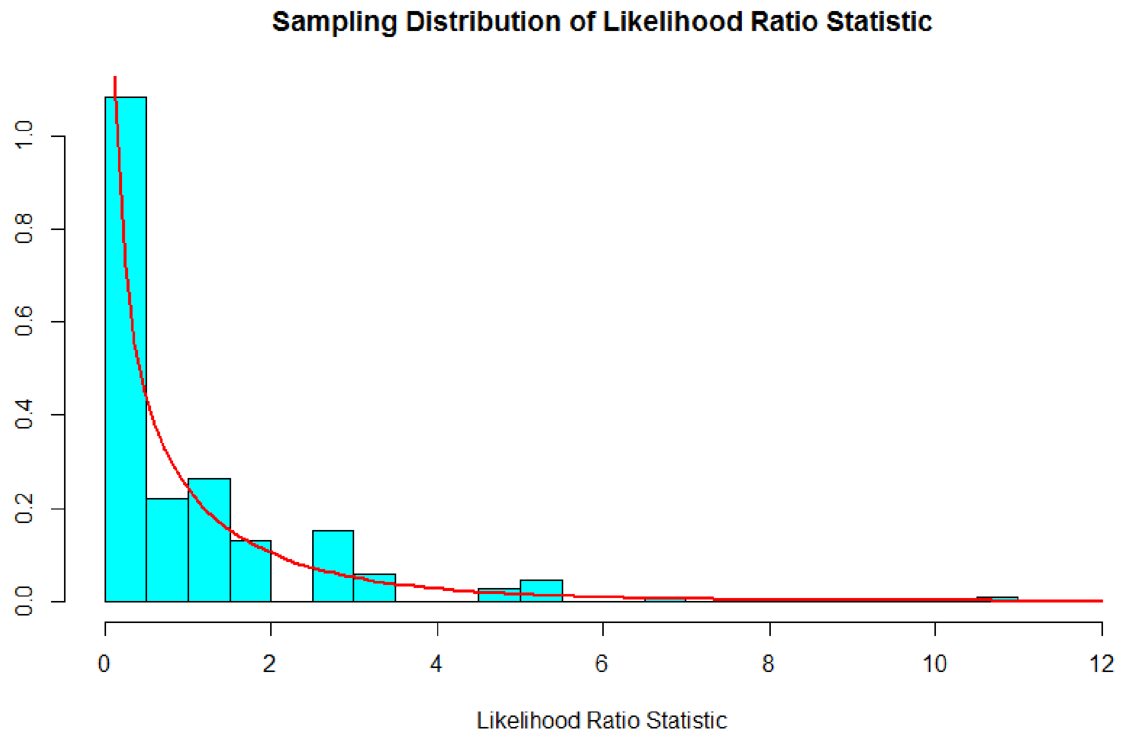
```
# For the 5000 simulations determine 15% likelihood intervals which are also
# approximate 95% likelihood intervals
for (i in 1:5000) {
y<-yobs[i]
thetahat<-that[i]
if (thetahat==0) { li[i,1]<-0}       # if thetahat=0 then likelihood interval has left endpoint = 0
else {result<-uniroot(function(x) BinRLF(x)-0.15,lower=0,upper=thetahat)
li[i,1]<-result$root}
if (thetahat==1) { li[i,2]<-1}      # if thetahat=1 then likelihood interval has right endpoint = 1
else {result<-uniroot(function(x) BinRLF(x)-0.15,lower=thetahat,upper=1)
li[i,2]<-result$root}
}
li[1:10,1:2]        # Look at first ten 15% likelihood intervals
# display proportion of 15% likelihood intervals which contain the true value of theta
prop<- mean(theta>=li[,1] & theta<=li[,2])
cat("proportion of 15% likelihood intervals which contain true value of theta = ",prop)
#
# calculate the likelihood ratio statistic for all 5000 simulations and plot a relative histogram of values
# the histogram approximates the sampling distribution of the likelihood ratio statistic
lambda<-(-2*log(dbinom(yobs,n,theta)/dbinom(yobs,n,that)))
truehist(lambda,h=0.5,xlab="Likelihood Ratio Statistic",main="Sampling Distribution of Likelihood Ratio Statistic")
curve(dchisq(x,1), from=0.001,to=12,add=TRUE,col="red",lwd=2) # superimpose Chi-squared (1) pdf
#
# use number of trials = 100 for the Binomial experiment
n<-100
cat("n = ",n," theta = ",theta)   # display values
# vector of observations for 5000 simulations from a Binomial(n,theta) distribution
yobs<- rbinom(5000,n,theta)
# corresponding vector of thetahat values
that<-yobs/n
# values used to construct an approximate 95% confidence interval  based on Gaussian approximation
pm<-1.96*sqrt(that*(1-that)/n)
# each approximate 95% confidence interval is stored in a row of matrix cibi
cibi<-matrix(c(that-pm,that+pm),nrow=5000,byrow=F)
cibi[1:10,1:2]        # Look at first 10 approximate 95% confidence intervals for theta
# display proportion of approximate 95% confidence intervals which contain true value of theta
prop<- mean(abs(theta-that)<pm)
cat("proportion of approximate 95% confidence intervals which contain true value of theta = ",prop)
#
# For the 5000 simulations determine 15% likelihood intervals which are also
# approximate 95% likelihood intervals
```

```
for (i in 1:5000) {
y<-yobs[i]
thetahat<-that[i]
if (thetahat==0) { li[i,1]<-0}      # if thetahat=0 then likelihood interval has left endpoint = 0
else {result<-uniroot(function(x) BinRLF(x)-0.15,lower=0,upper=thetahat)
li[i,1]<-result$root}
if (thetahat==1) { li[i,2]<-1}     # if thetahat=1 then likelihood interval has right endpoint = 1
else {result<-uniroot(function(x) BinRLF(x)-0.15,lower=thetahat,upper=1)
li[i,2]<-result$root}
}
li[1:10,1:2]        # Look at first ten 15% likelihood intervals
# display proportion of 15% likelihood intervals which contain true value of theta
prop<- mean(theta>=li[,1] & theta<=li[,2])
cat("proportion of 15% likelihood intervals which contain true value of theta = ",prop)
#
# calculate the likelihood ratio statistic for all 5000 simulations and plot a relative histogram of values
# the histogram approximates the sampling distribution of the likelihood ratio statistic
lambda<-(-2*log(dbinom(yobs,n,theta)/dbinom(yobs,n,that)))
truehist(lambda,h=0.5,xlab="Likelihood Ratio Statistic",main="Sampling Distribution of Likelihood Ratio
Statistic")
curve(dchisq(x,1), from=0.001,to=12,add=TRUE,col="red",lwd=2) # superimpose Chi-squared (1) pdf
#################################################################################
```

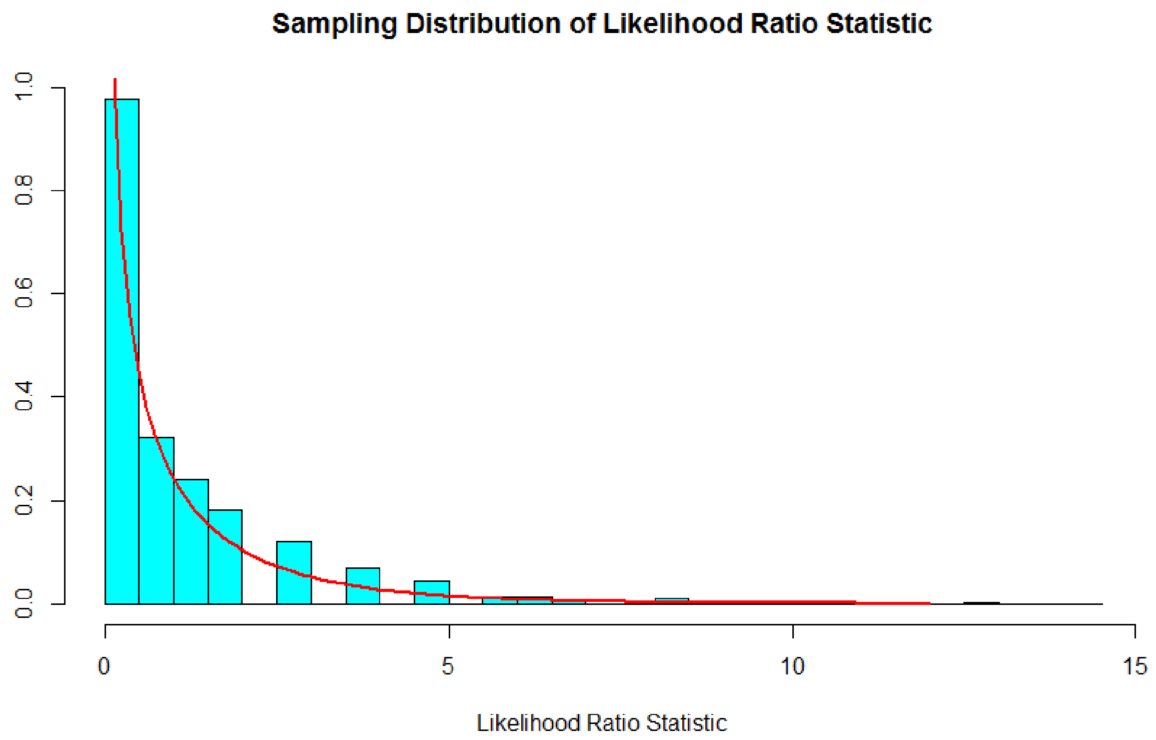
Verify that you obtain the following output:

```
> cat("n = ",n," theta = ",theta)   # display values
n =  30  theta =  0.8336096

> cibi[1:10,1:2]     # Look at first 10 approximate 95% confidence intervals
           [,1]      [,2]
 [1,] 0.6999722 0.9666944
 [2,] 0.6568618 0.9431382
 [3,] 0.4979767 0.8353566
 [4,] 0.7926464 1.0073536
 [5,] 0.6568618 0.9431382
 [6,] 0.7450226 0.9883107
 [7,] 0.7926464 1.0073536
 [8,] 0.6999722 0.9666944
 [9,] 0.7450226 0.9883107
[10,] 0.6153150 0.9180183

> prop<- mean(abs(theta-that)<pm)
> cat("proportion of approximate 95% confidence intervals which contain
true value of theta = ",prop)
proportion of approximate 95% confidence intervals which contain true value
of theta =  0.8914
```

```
> li[1:10,1:2]          # Look at first ten 15% likelihood intervals
           [,1]      [,2]
 [1,] 0.6764578 0.9363108
 [2,] 0.6367535 0.9146738
 [3,] 0.4903719 0.8159170
 [4,] 0.7618115 0.9738560
 [5,] 0.6367535 0.9146738
 [6,] 0.7179737 0.9562090
 [7,] 0.7618115 0.9738560
 [8,] 0.6764578 0.9363108
 [9,] 0.7179737 0.9562090
[10,] 0.5984359 0.8916837

> prop<- mean(that>=li[,1] & that<=li[,2])
> cat("proportion of 15% likelihood intervals which contain true value of
theta = ",prop)
proportion of 15% likelihood intervals which contain true value of theta
=  0.9544
```



```
> cat("n = ",n," theta = ",theta)   # display values
n =  100  theta =  0.8336096

> cibi[1:10,1:2]     # Look at first 10 approximate 95% confidence intervals
           [,1]      [,2]
 [1,] 0.7446993 0.8953007
 [2,] 0.8163075 0.9436925
 [3,] 0.7919905 0.9280095
 [4,] 0.7563760 0.9036240
 [5,] 0.7563760 0.9036240
 [6,] 0.6988077 0.8611923
 [7,] 0.7331090 0.8868910
 [8,] 0.8163075 0.9436925
 [9,] 0.7446993 0.8953007
[10,] 0.7331090 0.8868910


> prop<- mean(abs(theta-that)<pm)
> cat("proportion of approximate 95% confidence intervals which contain true
value of theta = ",prop)
proportion of approximate 95% confidence intervals which contain true value
of theta =  0.9436

> li[1:10,1:2]        # Look at first ten 15% likelihood intervals
           [,1]      [,2]
 [1,] 0.7376807 0.8863831
 [2,] 0.8074470 0.9335610
 [3,] 0.7837396 0.9182707
 [4,] 0.7490357 0.8944542
 [5,] 0.7490357 0.8944542
 [6,] 0.6929403 0.8530908
 [7,] 0.7263601 0.8781662
 [8,] 0.8074470 0.9335610
 [9,] 0.7376807 0.8863831
[10,] 0.7263601 0.8781662

> prop<- mean(that>=li[,1] & that<=li[,2])
> cat("proportion of 15% likelihood intervals which contain true value of
theta = ",prop)
proportion of 15% likelihood intervals which contain true value of theta
= 0.9538
```



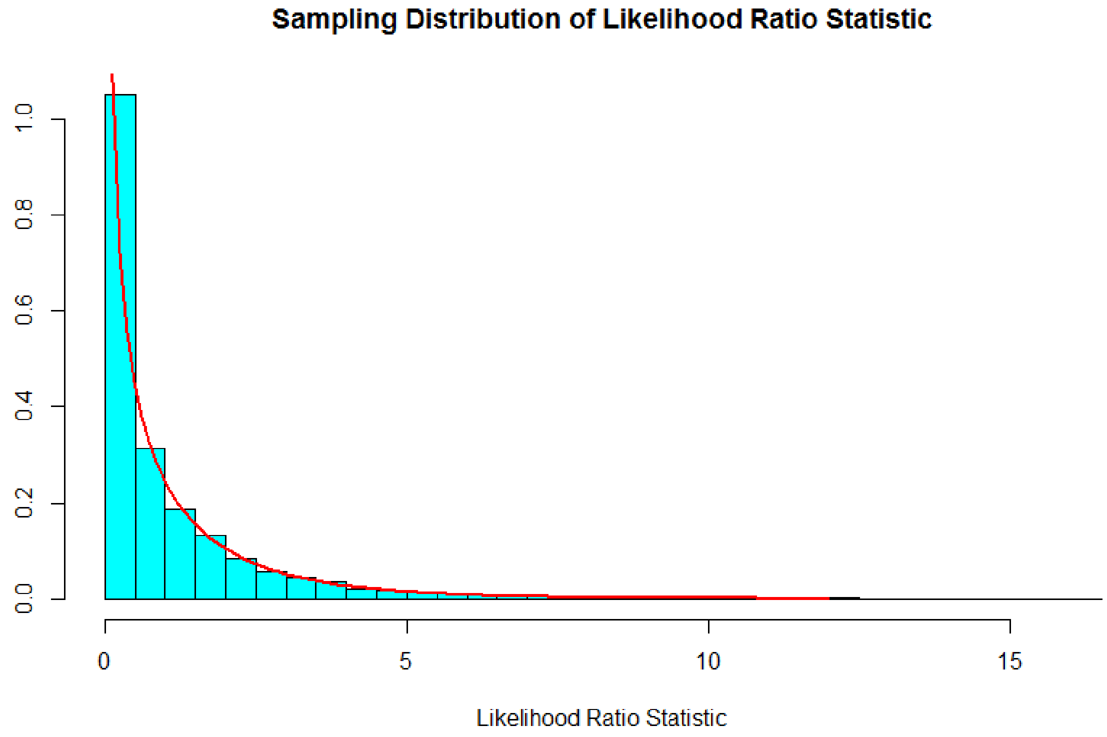
Problem 2: Run the following R code.

```
#########################################################################################
# Problem 2: Exponential confidence intervals and sampling distribution of likelihood ratio statistic
set.seed(id)
theta<-max(1,id-10*trunc(id/10))              # theta = last digit of ID unless it is zero
n<-20
cat("n = ",n," theta = ",theta)   # display values
ye<-rexp(5000*n,1/theta)
# each of the 5000 rows of the matrix ye contains n independent observations from
# Exponential(theta) distribution
ye<-matrix(ye,ncol=n,byrow=TRUE)
that<-apply(ye,1,mean)                    # vector of 5000 means
pm<-1.96*that/sqrt(n)                    # used to get approximate 95% confidence interval
# each approximate 95% confidence interval is stored in a row of matrix ciexp
ciexp<-matrix(c(that-pm,that+pm),nrow=5000,byrow=F)
ciexp[1:10,1:2]        # Look at first 10 approximate 95% confidence intervals
# display proportion of approximate 95% confidence intervals which contain true value of theta
prop<- mean(abs(theta-that)<pm)
cat("proportion of approximate 95% confidence intervals which contain true value of theta = ",prop)
#
# create function to calculate Exponential relative likelihood function
```

```
ExpRLF<-function(x) {(thetahat/x)^n*exp(n*(1-thetahat/x))}
li<-rep(0,2*5000)
li<- matrix(li,ncol=2,byrow=TRUE)                # initialize matrix to store likelihood intervals
# For the 5000 simulations determine 15% likelihood intervals which are also
# approximate 95% likelihood intervals
for (i in 1:5000) {
thetahat<-that[i]
result<-uniroot(function(x) ExpRLF(x)-0.15,lower=max(0,thetahat-4*thetahat/(n^0.5)),upper=thetahat)
li[i,1]<-result$root
result<-uniroot(function(x) ExpRLF(x)-0.15,lower=thetahat,upper= thetahat+4*thetahat/(n^0.5))
li[i,2]<-result$root
}
li[1:10,1:2]       # Look at first ten 15% likelihood intervals
# display proportion of 15% likelihood intervals which contain the value of theta
prop<- mean(theta>=li[,1] & theta<=li[,2])
cat("proportion of 15% likelihood intervals which contain true value of theta = ",prop)
#
# calculate the likelihood ratio statistic for all 5000 simulations and plot a relative histogram of values
# the histogram approximates the sampling distribution of the likelihood ratio statistic
lambda<- -2*log((that/theta)^n*exp(n*(1-that/theta)))
truehist(lambda,h=0.5,xlab="Likelihood Ratio Statistic",main="Sampling Distribution of Likelihood Ratio Statistic")
curve(dchisq(x,1), from=0.001,to=12,add=TRUE,col="red",lwd=2) # superimpose Chi-squared (1) pdf
##########################################################################################
```

Verify that you obtain the following output:

```
> cat("n = ",n," theta = ",theta)   # display values
n =  20  theta =  8

> ciexp[1:10,1:2]    # Look at first 10 approximate 95% confidence intervals
          [,1]      [,2]
 [1,] 5.566600 14.252862
 [2,] 5.280761 13.520992
 [3,] 2.667126  6.828977
 [4,] 3.874224  9.919661
 [5,] 5.992533 15.343433
 [6,] 3.120172  7.988966
 [7,] 4.192294 10.734054
 [8,] 4.614491 11.815057
 [9,] 4.196023 10.743603
[10,] 4.966030 12.715147

> prop<- mean(abs(theta-that)<pm)
> cat("proportion of approximate 95% confidence intervals which contain true
value of theta = ",prop)
proportion of approximate 95% confidence intervals which contain the true
value of theta =  0.9246



> li[1:10,1:2]         # Look at first ten 15% likelihood intervals
          [,1]      [,2]
 [1,] 6.602221 15.849327
 [2,] 6.263204 15.035480
 [3,] 3.163324  7.593868
 [4,] 4.594993 11.030757
 [5,] 7.107457 17.062053
 [6,] 3.700655  8.883794
 [7,] 4.972236 11.936374
 [8,] 5.472980 13.138463
 [9,] 4.976660 11.946993
[10,] 5.889920 14.139372

> prop<- mean(that>=li[,1] & that<=li[,2])
> cat("proportion of 15% likelihood intervals which contain true value of
theta = ",prop)
proportion of 15% likelihood intervals which contain true value of theta
=  0.9442
```



Problem 3: Run the following R code.

```
######################################################################################
# Problem 3: Gaussian confidence intervals
# The following R code runs a simulation in which 95% confidence intervals for the mean mu and the
# standard deviation sigma are  calculated for 5000 randomly generated Gaussian data sets
set.seed(id)
mu<-id-10*trunc(id/10)                          # mu = last digit of ID
sig<-max(1,trunc(id/10)-10*trunc(id/100))    # sig = second last digit of ID unless last digit is zero
cat("mu = ", mu, ", sigma = ", sig)      #display values of mu and sigma
yn<-rnorm(5000*25,mu,sig)            # generate G(mu,sig) observations
# each of the 5000 rows of the matrix yn contains 25 independent observations from a
# G(mu,sig) distribution
yn<-matrix(yn,ncol=25,byrow=TRUE)
ybar<-apply(yn,1,mean)                  # vector of 5000 means
s<-apply(yn,1,sd)                             # vector of 5000 sample standard deviations
a<-qt(0.975,24)              # value from t tables for 95% confidence interval for mu
pm<-a*s/sqrt(25)          # used to get 95% confidence interval for mu
# each confidence interval for mu is stored in a row of matrix cimu
cimu<-matrix(c(ybar-pm,ybar+pm),nrow=5000,byrow=F)
cimu[1:10,1:2]      #Look at first ten 95% confidence intervals for mu
# proportion of 95% confidence intervals which contain the true value of mu
prop<- mean(abs(mu-ybar)<pm)
cat("proportion of 95% confidence intervals which contain true value of mu = ",prop)
# values from Chi-square distribution for 95% confidence interval for sigma
a<-qchisq(0.025,24)
b<-qchisq(0.975,24)
# each confidence interval for sigma is stored in a row of matrix cisig
cisig<-matrix(c(sqrt(24*s^2/b), sqrt(24*s^2/a)),nrow=5000,byrow=F)
cisig[1:10,1:2]       #Look at first ten 95% confidence intervals for sigma
# proportion of 95% confidence intervals which contain the true value of sigma
prop<-mean(sig>=cisig[,1] & sig<=cisig[,2])
cat("proportion of 95% confidence intervals which contain true value of sigma = ",prop)
######################################################################################
```

Verify that you obtain the following output:

```
> cat("mu = ", mu, ", sigma = ", sig)       #display values of mu and sigma
mu =  8 , sigma =  5

> cimu[1:10,1:2]      #Look at first ten 95% confidence intervals for mu
          [,1]      [,2]
 [1,] 6.244030 10.164982
 [2,] 6.697515  9.783502
 [3,] 6.212403  9.020471
 [4,] 4.624907  9.944456
 [5,] 7.704320 11.294890
 [6,] 6.578151 11.437126
 [7,] 5.346232  9.403112
 [8,] 5.838531  9.540677
 [9,] 5.704073  9.729931
[10,] 5.110502  9.524473

> prop<- mean(abs(mu-ybar)<pm)
> cat("proportion of 95% confidence intervals which contain true value of mu
= ",prop)
proportion of 95% confidence intervals which contain true value of mu
=  0.9454



> cisig[1:10,1:2]    #Look at first ten 95% confidence intervals for sigma
          [,1]     [,2]
 [1,] 3.708505 6.607207
 [2,] 2.918779 5.200203
 [3,] 2.655918 4.731881
 [4,] 5.031321 8.963985
 [5,] 3.396023 6.050478
 [6,] 4.595703 8.187871
 [7,] 3.837067 6.836258
 [8,] 3.501554 6.238495
 [9,] 3.807726 6.783983
[10,] 4.174810 7.437994

> prop<-mean(sig>=cisig[,1] & sig<=cisig[,2])
> cat("proportion of 95% confidence intervals which contain true value of
sigma = ",prop)
proportion of 95% confidence intervals which contain true value of sigma =
0.9492
```

Run the R code for the 3 problems above again except modify the line

"id<-20456458"

in Problem 1 by replacing the number 20456458 with your UWaterloo ID number.

When you run the R code with your ID number you will generate 3 new plots. Export these 3 plots as .png files using RStudio (See Introduction to R and RStudio Section 6).

Download the Assignment 3 Template which is posted as a Word document on Learn. Fill in the required information and plots based on the output for the data generated using your ID number. Your assignment must follow the template exactly. See Assignment 3 Example posted on Learn.

Create a .pdf file for the answer to EACH problem.

Upload your assignment to Crowdmark using the link which was emailed to you.