HW2

2023-04-19

```
source("bootsample.R")
source("bootstats.R")
```

Problem 1

(a)

Use the code from Homework 1 to simulate data from a bivariate normal distribution with mean vector mu = [0 0] and correlation matrix

1. Sample size n in {50, 200} and correlation coefficient r = 0

```
library(MASS)

set.seed(0)
r = 0
bivariate_0_Size50 <- mvrnorm(n = 50, mu = c(0,0),
                              Sigma = matrix(data = c(1,r,r,1),
                                             nrow= 2, byrow = TRUE))
bivariate_0_Size200 <- mvrnorm(n = 200, mu = c(0,0),
                               Sigma = matrix(data = c(1,r,r,1),
                                              nrow= 2, byrow = TRUE))
```

2. Sample size n in {50, 200} and correlation coefficient r = 0.5

```
set.seed(0)
r = 0.5
bivariate_5_Size50 <- mvrnorm(n = 50, mu = c(0,0),
                              Sigma = matrix(data = c(1,r,r,1), nrow= 2,
                              byrow = TRUE))
bivariate_5_Size200 <- mvrnorm(n = 200, mu = c(0,0),
                               Sigma = matrix(data = c(1,r,r,1), nrow= 2,
                               byrow = TRUE))
```

3. Sample size n in {50, 200} and correlation coefficient r = 0.85

```
set.seed(0)
r = 0.85
bivariate_85_Size50 <- mvrnorm(n = 50, mu = c(0,0),
                               Sigma = matrix(data = c(1,r,r,1),
                               nrow= 2, byrow = TRUE))
bivariate_85_Size200 <- mvrnorm(n = 200, mu = c(0,0),
```

```
                                    Sigma = matrix(data = c(1,r,r,1),
                                    nrow= 2, byrow = TRUE))
```

(b)

Obtain the following Bootstrap Confidence Intervals for the correlation coefficient r for the three cases above.

1. Normal Bootstrap CI

Case r = 0:

```
results <- character(0)

set.seed(0)

alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

#Size 50
B = 5000
boot.samples = bootsampling(bivariate_0_Size50,B)

cor_sampling_dist_r0_50=matrix(nrow = B)

for(B in 1:B){
  cor_sampling_dist_r0_50[B] <- cor(boot.samples[,,B][,1],boot.samples[,,B][,2])
}

se <- sd(cor_sampling_dist_r0_50)/sqrt(50)
results <- append(results, paste("The Normal Bootstrap CI with r = 0 n = 50 is: ",mean(cor_sampling_dis
      mean(cor_sampling_dist_r0_50)+zval*se))


#Size 200
B = 5000
boot.samples = bootsampling(bivariate_0_Size200,B)

cor_sampling_dist_r0_200=matrix(nrow = B)

for(B in 1:B){
  cor_sampling_dist_r0_200[B] <- cor(boot.samples[,,B][,1],boot.samples[,,B][,2])
}

se <- sd(cor_sampling_dist_r0_200)/sqrt(200)
results <- append(results, paste("The Normal Bootstrap CI with r = 0 n = 200 is: ",mean(cor_sampling_di
      mean(cor_sampling_dist_r0_200)+zval*se))
```

Case r = 0.5

```
set.seed(0)

alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

#Size 50
```

```
B = 5000
boot.samples = bootsampling(bivariate_5_Size50,B)

cor_sampling_dist_r5_50=matrix(nrow = B)

for(B in 1:B){
  cor_sampling_dist_r5_50[B] <- cor(boot.samples[,,B][,1],boot.samples[,,B][,2])
}

se <- sd(cor_sampling_dist_r5_50)/sqrt(50)
results <- append(results, paste("The Normal Bootstrap CI with r = 0.5 n = 50 is: ",mean(cor_sampling_di
      mean(cor_sampling_dist_r5_50)+zval*se))

#Size 200
B = 5000
boot.samples = bootsampling(bivariate_5_Size200,B)

cor_sampling_dist_r5_200=matrix(nrow = B)

for(B in 1:B){
  cor_sampling_dist_r5_200[B] <- cor(boot.samples[,,B][,1],boot.samples[,,B][,2])
}

se <- sd(cor_sampling_dist_r5_200)/sqrt(200)
results <- append(results, paste("The Normal Bootstrap CI with r = 0.5 n = 200 is: ",mean(cor_sampling_d
      mean(cor_sampling_dist_r5_200)+zval*se))
```

Case r = 0.85

```
set.seed(0)

alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

#Size 50
B = 5000
boot.samples = bootsampling(bivariate_85_Size50,B)

cor_sampling_dist_r85_50=matrix(nrow = B)

for(B in 1:B){
  cor_sampling_dist_r85_50[B] <- cor(boot.samples[,,B][,1],boot.samples[,,B][,2])
}

se <- sd(cor_sampling_dist_r85_50)/sqrt(50)
results <- append(results, paste("The Normal Bootstrap CI with r = 0.85 n = 50 is: ",mean(cor_sampling_d
      mean(cor_sampling_dist_r85_50)+zval*se))

#Size 200
B = 5000
boot.samples = bootsampling(bivariate_85_Size200,B)

cor_sampling_dist_r85_200=matrix(nrow = B)
```

```r
for(B in 1:B){
  cor_sampling_dist_r85_200[B] <- cor(boot.samples[,,B][,1],boot.samples[,,B][,2])
}

se <- sd(cor_sampling_dist_r85_200)/sqrt(200)
results <- append(results,paste("The Normal Bootstrap CI with r = 0.85 n = 200 is: ",mean(cor_sampling_d
      mean(cor_sampling_dist_r85_200)+zval*se))
```

```r
results
```

```
## [1] "The Normal Bootstrap CI with r = 0 n = 50 is:  -0.201802096370368 ,  -0.114918564710983"
## [2] "The Normal Bootstrap CI with r = 0 n = 200 is:  -0.0169395226245891 ,  0.00264186542574477"
## [3] "The Normal Bootstrap CI with r = 0.5 n = 50 is:  0.529329770743826 ,  0.580493392731537"
## [4] "The Normal Bootstrap CI with r = 0.5 n = 200 is:  0.481312870210314 ,  0.495475302979146"
## [5] "The Normal Bootstrap CI with r = 0.85 n = 50 is:  0.861750969287959 ,  0.879577567035012"
## [6] "The Normal Bootstrap CI with r = 0.85 n = 200 is:  0.842816051531635 ,  0.848146612829333"
```

Results Discussion

When observing the Normal Bootstrap Confidence intervals for n=50 versus n=200, we can observe that the lengths of the bootstraps with n=200 are generally smaller. For example, the length of the interval for sample size 50 versus 200 is 0.08688353 vs. 0.01958139 for r=0, 0.05116362 vs 0.01416243 for r = 0.5, and 0.0178266 vs 0.005330561 for r = 0.85. As we can see, the interval is always smaller for larger n. In addition, we can observe that as the value or r increases, the interval decreases. The interval for r=0.85 and n=200 is the smallest interval observed with a length of only 0.005330561.

2. Basic Bootstrap CI

Case r = 0:

```r
#n = 50
results <- character(0)
se <- sd(cor_sampling_dist_r0_50)/sqrt(50)
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results,paste("The Basic Bootstrap CI for r = 0, n= 50: ",
      2*mean(cor_sampling_dist_r0_50)-quantile(cor_sampling_dist_r0_50,probs=(1-alpha/2)), ", ", 2*mean


#n = 200
se <- sd(cor_sampling_dist_r0_200)/sqrt(200)
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Basic Bootstrap CI for r = 0, n= 200: ",
      2*mean(cor_sampling_dist_r0_200)-quantile(cor_sampling_dist_r0_200,probs=(1-alpha/2)), ", ", 2*mea
```

Case r = 0.5

```r
#n = 50
se <- sd(cor_sampling_dist_r5_50)/sqrt(50)
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Basic Bootstrap CI for r = 0.5, n= 50: ",
```

```r
      2*mean(cor_sampling_dist_r5_50)-quantile(cor_sampling_dist_r5_50,probs=(1-alpha/2)), ", ", 2*mean


#n = 200
se <- sd(cor_sampling_dist_r5_200)/sqrt(200)
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Basic Bootstrap CI for r = 0.5, n= 200: ",
      2*mean(cor_sampling_dist_r5_200)-quantile(cor_sampling_dist_r5_200,probs=(1-alpha/2)), ", ", 2*mea
```

Case r = 0.85

```r
#n = 50
se <- sd(cor_sampling_dist_r85_50)/sqrt(50)
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Basic Bootstrap CI for r = 0.85, n= 50: ",
      2*mean(cor_sampling_dist_r85_50)-quantile(cor_sampling_dist_r85_50,probs=(1-alpha/2)), ", ", 2*mea


#n = 200
se <- sd(cor_sampling_dist_r85_200)/sqrt(200)
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Basic Bootstrap CI for r = 0.85, n= 200: ",
      2*mean(cor_sampling_dist_r85_200)-quantile(cor_sampling_dist_r85_200,probs=(1-alpha/2)), ", ", 2*m
```

```r
results
```

```
## [1] "The Basic Bootstrap CI for r = 0, n= 50:  -0.478184598499841 ,  0.135034237701975"
## [2] "The Basic Bootstrap CI for r = 0, n= 200:  -0.144526566300063 ,  0.134776793950893"
## [3] "The Basic Bootstrap CI for r = 0.5, n= 50:  0.393191321795091 ,  0.746848107882896"
## [4] "The Basic Bootstrap CI for r = 0.5, n= 200:  0.39467735177745 ,  0.591560052599396"
## [5] "The Basic Bootstrap CI for r = 0.85, n= 50:  0.817698702154826 ,  0.94151820454376"
## [6] "The Basic Bootstrap CI for r = 0.85, n= 200:  0.811470856042606 ,  0.885514345733031"
```

Results Discussion

The length of the Basic Bootstrap confidence interval is generally lower for the larger sample size of 200. The lengths of the intervals for n=50 and n=200 is 0.6132188 and 0.2793034 for r=0, 0.3536568 and 0.1968827 for r=0.5, and 0.12381950 and 0.07404349 for r=0.85. We again see that for larger values of r, the interval length decreases.

3. Percentile Bootstrap CI

Case r = 0:

```r
results <- character(0)
#n = 50
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Percentile Bootstrap CI for r = 0, n= 50: ",
```

```r
      quantile(cor_sampling_dist_r0_50 , probs=(alpha/2)),
      ", ",quantile(cor_sampling_dist_r0_50,probs=(1-alpha/2))))


#n = 200
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Percentile Bootstrap CI for r = 0, n= 50: ",
      quantile(cor_sampling_dist_r0_200 , probs=(alpha/2)),
      ", ",quantile(cor_sampling_dist_r0_200,probs=(1-alpha/2))))
```

Case r = 0.5:

```r
#n = 50
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Percentile Bootstrap CI for r = 0, n= 50: ",
      quantile(cor_sampling_dist_r5_50 , probs=(alpha/2)),
      ", ",quantile(cor_sampling_dist_r5_50,probs=(1-alpha/2))))


#n = 200
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Percentile Bootstrap CI for r = 0, n= 50: ",
      quantile(cor_sampling_dist_r5_200 , probs=(alpha/2)),
      ", ",quantile(cor_sampling_dist_r5_200,probs=(1-alpha/2))))
```

Case r = 0.85:

```r
#n = 50
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Percentile Bootstrap CI for r = 0, n= 50: ",
      quantile(cor_sampling_dist_r85_50 , probs=(alpha/2)),
      ", ",quantile(cor_sampling_dist_r85_50,probs=(1-alpha/2))))


#n = 200
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Percentile Bootstrap CI for r = 0, n= 50: ",
      quantile(cor_sampling_dist_r85_200 , probs=(alpha/2)),
      ", ",quantile(cor_sampling_dist_r85_200,probs=(1-alpha/2))))
```

```r
results
```

```
## [1] "The Percentile Bootstrap CI for r = 0, n= 50:  -0.451754898783326 ,  0.161463937418489"
## [2] "The Percentile Bootstrap CI for r = 0, n= 50:  -0.149074451149737 ,  0.130228909101219"
## [3] "The Percentile Bootstrap CI for r = 0, n= 50:  0.362975055592467 ,  0.716631841680271"
## [4] "The Percentile Bootstrap CI for r = 0, n= 50:  0.385228120590065 ,  0.58211082141201"
```

```
## [5] "The Percentile Bootstrap CI for r = 0, n= 50:  0.799810331779211 ,  0.923629834168145"
## [6] "The Percentile Bootstrap CI for r = 0, n= 50:  0.805448318627937 ,  0.879491808318362"
```

Results Discussion

The lengths of the Percentail Bootstrap Confidence intervals for n=50 and n=200 are 0.6132188 and 0.2793034 for r=0, 0.3536568 and 0.1968827 for r=0.5, and 0.12381950 and 0.07404349 for r=085. The lengths of the interval decrease for larger sample size and larger r values.

4. Bootstrap-t (Studentized) Bootstrap CI

Case r=0:

```r
results <- character(0)

## n = 50
set.seed(0)
B= 5000
boot.samples = bootsampling(bivariate_0_Size50,B)
boot.samples.se <- matrix(ncol = 1)

for(i in 1:B){
  boot.samples.se[i]<- sd(boot.samples[,,i])/sqrt(nrow(boot.samples[,,i]))
}
#T statistic
t_statistic = (cor_sampling_dist_r0_50-cor(bivariate_0_Size50[,1], bivariate_0_Size50[,2])) / boot.sampl
qval = quantile(t_statistic,probs=c(alpha/2,1-alpha/2))

se <- sd(cor_sampling_dist_r0_50)/sqrt(50)
# construct the studentized.CI
results <- append(results, paste("The Bootstrap-t (Studentized) Bootstrap CI with r=0 n=50:",
                                 mean(cor_sampling_dist_r0_50)-qval[2]*se ,
                                 mean(cor_sampling_dist_r0_50)-qval[1]*se))

## n = 200
set.seed(0)
B= 5000
boot.samples = bootsampling(bivariate_0_Size200,B)

boot.samples.se <- matrix(ncol = 1)

for(i in 1:B){
  boot.samples.se[i]<- sd(boot.samples[,,i])/sqrt(nrow(boot.samples[,,i]))
}
#T statistic
t_statistic = (cor_sampling_dist_r0_200-cor(bivariate_0_Size200[,1], bivariate_0_Size200[,2])) / boot.sa
qval = quantile(t_statistic,probs=c(alpha/2,1-alpha/2))

se <- sd(cor_sampling_dist_r0_200)/sqrt(50)
# construct the studentized.CI
results <- append(results,
                  paste("The Bootstrap-t (Studentized) Bootstrap CI with r=0 n=200:",
      mean(cor_sampling_dist_r0_200)-qval[2]*se ,
      mean(cor_sampling_dist_r0_200)-qval[1]*se))
```

Case r=0.5:

```
## n = 50
set.seed(0)
B= 5000
boot.samples = bootsampling(bivariate_5_Size50,B)

boot.samples.se <- matrix(ncol = 1)

for(i in 1:B){
  boot.samples.se[i]<- sd(boot.samples[,,i])/sqrt(nrow(boot.samples[,,i]))
}
#T statistic
t_statistic = (cor_sampling_dist_r5_50-cor(bivariate_5_Size50[,1], bivariate_5_Size50[,2])) / boot.sampl
qval = quantile(t_statistic,probs=c(alpha/2,1-alpha/2))

se <- sd(cor_sampling_dist_r5_50)/sqrt(50)
# construct the studentized.CI
results <- append(results,
                  paste("The Bootstrap-t (Studentized) Bootstrap CI with r=0.5 n=50:",
      mean(cor_sampling_dist_r5_50)-qval[2]*se ,
      mean(cor_sampling_dist_r5_50)-qval[1]*se))

## n = 200
set.seed(0)
B= 5000
boot.samples = bootsampling(bivariate_5_Size200,B)

boot.samples.se <- matrix(ncol = 1)

for(i in 1:B){
  boot.samples.se[i]<- sd(boot.samples[,,i])/sqrt(nrow(boot.samples[,,i]))
}
#T statistic
t_statistic = (cor_sampling_dist_r5_200-cor(bivariate_5_Size200[,1], bivariate_5_Size200[,2])) / boot.sa
qval = quantile(t_statistic,probs=c(alpha/2,1-alpha/2))

se <- sd(cor_sampling_dist_r5_200)/sqrt(50)
# construct the studentized.CI
results <- append(results,
                  paste("The Bootstrap-t (Studentized) Bootstrap CI with r=0.5 n=200:",
      mean(cor_sampling_dist_r5_200)-qval[2]*se ,
      mean(cor_sampling_dist_r5_200)-qval[1]*se))
```

Case r=0.85:

```
## n = 50
set.seed(0)
B= 5000
boot.samples = bootsampling(bivariate_85_Size50,B)

boot.samples.se <- matrix(ncol = 1)

for(i in 1:B){
  boot.samples.se[i]<- sd(boot.samples[,,i])/sqrt(nrow(boot.samples[,,i]))
```

```
}
#T statistic
t_statistic = (cor_sampling_dist_r85_50-cor(bivariate_85_Size50[,1],bivariate_85_Size50[,2])) / boot.sam
qval = quantile(t_statistic,probs=c(alpha/2,1-alpha/2))

se <- sd(cor_sampling_dist_r85_50)/sqrt(50)
# construct the studentized.CI
results <- append(results,
                  paste("The Bootstrap-t (Studentized) Bootstrap CI with r=0.85 n=50:",
      mean(cor_sampling_dist_r85_50)-qval[2]*se ,
      mean(cor_sampling_dist_r85_50)-qval[1]*se))

## n = 200
set.seed(0)
B= 5000
boot.samples = bootsampling(bivariate_85_Size200,B)

boot.samples.se <- matrix(ncol = 1)

for(i in 1:B){
  boot.samples.se[i]<- sd(boot.samples[,,i])/sqrt(nrow(boot.samples[,,i]))
}
#T statistic
t_statistic = (cor_sampling_dist_r85_200-cor(bivariate_85_Size200[,1],bivariate_85_Size200[,2])) / boot
qval = quantile(t_statistic,probs=c(alpha/2,1-alpha/2))

se <- sd(cor_sampling_dist_r85_200)/sqrt(50)
# construct the studentized.CI
results <- append(results,
                  paste("The Bootstrap-t (Studentized) Bootstrap CI with r=0.85 n=200:",
      mean(cor_sampling_dist_r85_200)-qval[2]*se ,
      mean(cor_sampling_dist_r85_200)-qval[1]*se))
```

```
results
```

```
## [1] "The Bootstrap-t (Studentized) Bootstrap CI with r=0 n=50: -0.217763388332143 -0.106357560739116
## [2] "The Bootstrap-t (Studentized) Bootstrap CI with r=0 n=200: -0.0264166970534254 0.012622217385618
## [3] "The Bootstrap-t (Studentized) Bootstrap CI with r=0.5 n=50: 0.538203171719579 0.575724428430876
## [4] "The Bootstrap-t (Studentized) Bootstrap CI with r=0.5 n=200: 0.478716829448487 0.49884400820029
## [5] "The Bootstrap-t (Studentized) Bootstrap CI with r=0.85 n=50: 0.868845108593625 0.87340971450276
## [6] "The Bootstrap-t (Studentized) Bootstrap CI with r=0.85 n=200: 0.844166809581764 0.84701707522250
```

Results Discussion

The lengths of the Bootstrap-t (Studentized) Bootstrap Confidence Intervals for n=50 and n=200 are 0.11140583 and 0.03903891 for r=0, 0.03752126 and 0.02012718 for n= 200, and 0.004564606 and 0.002850266 for r = 0.85.

Problem 1 Discussion

In every type of bootstrap, a larger sample size of n=200 results in a smaller confidence interval length. In addition, for larger values of r, we observe a narrower confidence interval.

I decided to use B=5000 number of bootstrap replicates. As seen in confidence intervals for the population median in Lecture2-2 and Lecture3-1, the Bootstrap confidence interval for greater values of B result in negligable differences.

Problem 2

```
library(MASS)
data(cats)
summary(cats)
```

```
##  Sex         Bwt             Hwt
##  F:47   Min.   :2.000   Min.   : 6.30
##  M:97   1st Qu.:2.300   1st Qu.: 8.95
##         Median :2.700   Median :10.10
##         Mean   :2.724   Mean   :10.63
##         3rd Qu.:3.025   3rd Qu.:12.12
##         Max.   :3.900   Max.   :20.50
```

Part a)

Construct the following bootstrap CI for the difference of the body weight means between female and male cats.

1. Normal Bootstrap CI

```
results <- character(0)
set.seed(1)
B = 5000
boot.sample_F <- bootsampling(cats$Bwt[cats$Sex == "F"], boot.replicates = B)
sampling_dist_F <- apply(boot.sample_F,2,  mean)
boot.sample_M <- bootsampling(cats$Bwt[cats$Sex == "M"], boot.replicates = B)
sampling_dist_M <- apply(boot.sample_M,2,  mean)
diffs<- sampling_dist_F-sampling_dist_M

se<- sd(diffs)/sqrt(length(cats$Bwt))

results <- append(results,paste("The Normal Bootstrap CI with B = 5000: ",mean(diffs)-zval*se,", ",
      mean(diffs)+zval*se))
```

2. Basic Bootstrap CI

```
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results,paste("The Basic Bootstrap CI for with B = 5000: ",2*mean(diffs)-quantile(dif
       2*mean(diffs)-quantile(diffs,probs=(alpha/2))))
```

3. Percentile Bootstrap CI

```
results <- append(results,paste("The Percentile Bootstrap CI with B = 5000: ",quantile(diffs , probs=(al
      ", ",quantile(diffs,probs=(1-alpha/2))))
```

4. Bootstrap-t (Studentized) Bootstrap CI

```
set.seed(0)
B= 5000
boot.sample_F <- bootsampling(cats$Bwt[cats$Sex == "F"], boot.replicates = B)
sampling_dist_F <- apply(boot.sample_F,2,  mean)
boot.sample_M <- bootsampling(cats$Bwt[cats$Sex == "M"], boot.replicates = B)
```

```r
sampling_dist_M <- apply(boot.sample_M,2,  mean)
diffs<- sampling_dist_F-sampling_dist_M

boot.samples.se <- matrix(ncol = 1)

for(i in 1:B){
  boot.samples.se[i]<-
    sqrt((sd(boot.sample_F[,i]))^2/
    length(boot.sample_F[,i]) +
    (sd(boot.sample_M[,i]))^2/
    length(boot.sample_M[,i]))
}

#T statistic
t_statistic = (diffs-(mean(cats$Bwt[cats$Sex == "F"])- mean(cats$Bwt[cats$Sex == "M"]))) / boot.samples
qval = quantile(t_statistic,probs=c(alpha/2,1-alpha/2))

se <- sd(diffs)/sqrt(length(cats$Bwt))
# construct the studentized.CI
results <- append(results, paste("Bootstrap-t (Studentized) Bootstrap CI",
                                 mean(diffs)-qval[2]*se ,", ",
                                 mean(diffs)-qval[1]*se))
```

```r
results
```

```
## [1] "The Normal Bootstrap CI with B = 5000:  -0.550755464054004 ,  -0.53065740718969"
## [2] "The Basic Bootstrap CI for with B = 5000:  -0.661564220223733 ,  -0.420521228339548"
## [3] "The Percentile Bootstrap CI with B = 5000:  -0.660891642904146 ,  -0.419848651019961"
## [4] "Bootstrap-t (Studentized) Bootstrap CI -0.550099598383429 ,  -0.530281391370451"
```

The length of the Bootstrap Confidence Intervals are as follows: Normal Bootstrap CI: 0.02009806, Basic Bootstrap CI: 0.24104299, Percentile Bootstrap CI: 0.24104299, Bootstrap-t (Studentized) Bootstrap CI: 0.01981821.

I decided to use B=5000 number of bootstrap replicates. As seen on the slides, after 5000 replicates, increasing the amount of replicates does not lead to a much more accurate interval that justifies the increased computation.

We observe that the Normal Bootstrap and the Bootstrap-t Studentized Bootstrap CI are much more narrow than the other types of Intervals.

The results of our confidence intervals infer that there is difference of the body weight means between female and male cats. In each interval of Female bodyweight-male bodyweight, the center is around -0.54. None of the intervals contain 0 which provides significant evidence that there is a difference in the true means

Part b)

Using your code from Problem 1 above to: construct the following bootstrap CI for the correlation coefficient between the body weight and the heart weight of female cats.

1. Normal Bootstrap CI

```r
results <- character(0)
cats_F <- cbind(cats$Bwt[cats$Sex == "F"], cats$Hwt[cats$Sex == "F"])

set.seed(1)
cat_bootsamples <- bootsampling(cats_F, boot.replicates = B)
```

```r
cor_matrix <- matrix(nrow = 5000)
for(B in 1:5000){
  cor_matrix[B] <- cor(cat_bootsamples[,,B][,1],cat_bootsamples[,,B][,2])
}

alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

se <- sd(cor_matrix)/sqrt(47)
results <- append(results,paste("The Normal Bootstrap CI: ",mean(cor_matrix)-zval*se,", ",
      mean(cor_matrix)+zval*se))
```

2. Basic Bootstrap CI

```r
#n = 200
se <- sd(cor_matrix)/sqrt(nrow(cats_F))
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results, paste("The Basic Bootstrap CI: ",
      2*mean(cor_matrix)-quantile(cor_matrix,probs=(1-alpha/2)), ", ", 2*mean(cor_matrix)-quantile(cor_m
```

3. Percentile Bootstrap CI

```r
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results,paste("The Percentile Bootstrap CI: ",
      quantile(cor_matrix , probs=(alpha/2)),
      ", ",quantile(cor_matrix,probs=(1-alpha/2))))
```

4. Bootstrap-t (Studentized) Bootstrap CI

```r
B= 5000

boot.samples.se <- matrix(ncol = 1)

for(i in 1:B){
  boot.samples.se[i]<-
    sqrt(sd(cat_bootsamples[,,i][,1])^2/length(cat_bootsamples[,,i][,1]) +
    sd(cat_bootsamples[,,i][,2])^2/length(cat_bootsamples[,,i][,2]))
}


#T statistic
t_statistic = (cor_matrix-cor(cats$Bwt[cats$Sex == "F"], cats$Hwt[cats$Sex == "F"])) / boot.samples.se
qval = quantile(t_statistic,probs=c(alpha/2,1-alpha/2))

se <- sd(cor_matrix)/sqrt(length(cat_bootsamples[,,i][,2]))
# construct the studentized.CI
results <- append(results,paste("The Bootstrap-t (Studentized) Bootstrap CI: ", mean(cor_matrix)-qval[2
      mean(cor_matrix)-qval[1]*se))
```

```r
results
```

```
## [1] "The Normal Bootstrap CI:  0.492547253695381 ,  0.553624599601357"
```

```
## [2] "The Basic Bootstrap CI:  0.334113803040965 ,  0.754538820559025"
## [3] "The Percentile Bootstrap CI:  0.291633032737713 ,  0.712058050255774"
## [4] "The Bootstrap-t (Studentized) Bootstrap CI:  0.510178671709239 ,  0.543362541148539"
```

The length of the Bootstrap Confidence Intervals are as follows: Normal Bootstrap CI: 0.06107735, Basic Bootstrap CI: 0.42042502, Percentile Bootstrap CI: 0.42042502, Bootstrap-t (Studentized) Bootstrap CI: 0.03318387.

I decided to use B=5000 number of bootstrap replicates. As seen on the slides, after 5000 replicates, increasing the amount of replicates does not lead to a much more accurate interval that justifies the increased computation.

The Confidence Intervals of the correlation coefficient between the body weight and the heart weight of female cats contains significant evidence that the correlation is not equal to 0. The Basic Bootstrap CI and Percentile Bootstrap CI both have a much larger length than the Normal Bootstrap and Bootstrap-t Studentized Bootstrap, with the latter being the most narrow. All intervals center around 0.52, which infer that there is a correlation between body weight and heart weight of female cats.

Part c)

Using your code from Problem 1 above to: construct the following bootstrap CI for the correlation coefficient between the body weight and the heart weight of male cats.

1. Normal Bootstrap CI

```r
results <- character(0)
cats_M <- cbind(cats$Bwt[cats$Sex == "M"], cats$Hwt[cats$Sex == "M"])

set.seed(1)
cat_bootsamples <- bootsampling(cats_M, boot.replicates = B)

cor_matrix <- matrix(nrow = 5000)
for(B in 1:5000){
  cor_matrix[B] <- cor(cat_bootsamples[,,B][,1],cat_bootsamples[,,B][,2])
}

alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

se <- sd(cor_matrix)/sqrt(47)
results <- append(results, paste("The Normal Bootstrap CI: ",mean(cor_matrix)-zval*se,", ",
      mean(cor_matrix)+zval*se))
```

2. Basic Bootstrap CI

```r
#n = 200
se <- sd(cor_matrix)/sqrt(nrow(cats_F))
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)

results <- append(results,paste("The Basic Bootstrap CI for r = 0, n= 200: ",
      2*mean(cor_matrix)-quantile(cor_matrix,probs=(1-alpha/2)), ", ", 2*mean(cor_matrix)-quantile(cor_r
```

3. Percentile Bootstrap CI

```r
alpha=.05
zval=qnorm(p=alpha/2, lower.tail=FALSE)
```

```r
results <- append(results,paste("The Percentile Bootstrap CI: ",
      quantile(cor_matrix , probs=(alpha/2)),
      ", ",quantile(cor_matrix,probs=(1-alpha/2))))
```

4. Bootstrap-t (Studentized) Bootstrap CI

```r
B= 5000

boot.samples.se <- matrix(ncol = 1)

for(i in 1:B){
  boot.samples.se[i]<-
    sqrt(sd(cat_bootsamples[,,i][,1])^2/length(cat_bootsamples[,,i][,1]) +
    sd(cat_bootsamples[,,i][,2])^2/length(cat_bootsamples[,,i][,2]))
}


#T statistic
t_statistic = (cor_matrix-cor(cats$Bwt[cats$Sex == "M"], cats$Hwt[cats$Sex == "M"])) / boot.samples.se
qval = quantile(t_statistic,probs=c(alpha/2,1-alpha/2))

se <- sd(cor_matrix)/sqrt(length(cat_bootsamples[,,i][,2]))
# construct the studentized.CI
results <- append(results,paste("The Bootstrap-t (Studentized) Bootstrap CI",
                                mean(cor_matrix)-qval[2]*se ,", ",
                                mean(cor_matrix)-qval[1]*se))
```

```r
results
```

```
## [1] "The Normal Bootstrap CI:  0.783145019132168 ,  0.80299227121392"
## [2] "The Basic Bootstrap CI for r = 0, n= 200:  0.732731691829907 ,  0.867866386717462"
## [3] "The Percentile Bootstrap CI:  0.718270903628626 ,  0.853405598516181"
## [4] "The Bootstrap-t (Studentized) Bootstrap CI 0.792308535712384 ,  0.794198924972136"
```

The length of the Bootstrap Confidence Intervals are as follows: Normal Bootstrap CI: 0.01984725, Basic Bootstrap CI: 0.1351347, Percentile Bootstrap CI: 0.1351347, Bootstrap-t (Studentized) Bootstrap CI: 0.001890389

Again, I chose to use B=5000 replicates in order to get an optimal confidence interval that does not require too long of a computation process.

The Confidence Intervals inter that the correlation coefficient between the body weight and the heart weight of male cats is not equal to 0. The intervals all center around 0.79, therefore there does seem to be a correlation between body weight and heart weight of male cats, a higher correlation than female cats.