Solutions to Homework 4

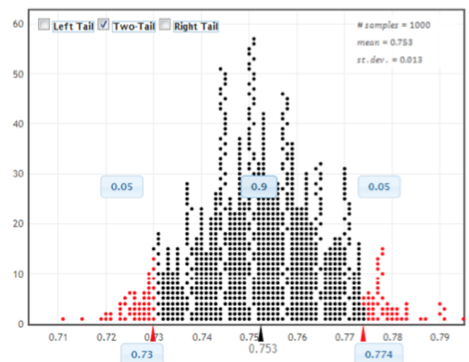
Statistics 302 Professor Larget

Textbook Exercises

3.102 How Important is Regular Exercise? In a recent poll of 1000 American adults, the number saying that exercise is an important part of daily life was 753. Use *StatKey* or other technology to find and interpret at 90% confidence interval for the proportion of American adults who think exercise is an important part of daily life.

<u>Solution</u>

Using StatKey or other technology, we produce a bootstrap distribution such as the figure shown below. For a 90% confidence interval, we find the 5%-tile and 95%-tile points in this distribution to be 0.730 and 0.774. We are 90% confident that the percent of American adults who think exercise is an important part of daily life is between 73.0% and 77.4%.



3.104 Comparing Methods for Having Dogs Identify Cancer in People Exercise 2.17 on page 55 describes a study in which scientists train dogs to smell cancer. Researchers collected breath and stool samples from patients with cancer as well as from healthy people. A trained dog was given five samples, randomly displayed, in each test, one from a patient with cancer and four from health volunteers. The results are displayed in the table below. Use *StatKey* or other technology to use a bootstrap distribution to find and interpret a 90% confidence interval for the difference in the proportion of time the dog correctly picks out the cancer sample between the two types of samples. Is it plausible that there is no difference in the effectiveness in the two types of methods (breath or stool)?

	Breath Test	Stool Test	Total
Dog selects cancer	33	37	70
Dog does not select cancer	3	1	4
Total	36	38	74

<u>Solution</u>

The dog got $\hat{p}_B = 33/36 = 0.917$ or 91.7% of the breath samples correct and $\hat{p}_S = 37/38 = 0.974$ or 97.4% of the stool samples correct. (A remarkably high percentage in both cases!) We create a bootstrap distribution for the difference in proportions using StatKey or other technology (as in the figure below) and then find the middle 90% of values. Using the figure, the 90% confidence interval for $p_B - p_S$ is -0.14 to 0.025. We are 90% confident that the difference between the proportion correct for breath samples and the proportion correct for stool samples for all similar tests we might give this dog is between -0.14 and 0.025. Since a difference of zero represents no difference, and zero is in the interval of plausible values, it is plausible that there is no difference in the effectiveness of breath vs stool samples in having this dog detect cancer.

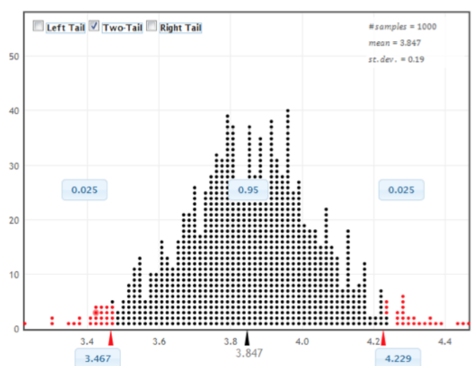
3.105 Average Tip for a Waitress Data 2.12 on page 119 describes information from a sample of 157 restaurant bills collected at the *First Crush* bistro. The data is available in **Restaurant Tips**. Create a bootstrap distribution using this data and find and interpret a 95% confidence interval for the average tip left at this restaurant. Find the confidence interval two ways: using the standard error and using percentiles. Compare your results.

Solution

Using one bootstrap distribution (as shown below), the standard error is SE $= 0.19$. The mean tip from the original sample is $\bar{x} = 3.85$, so a 95% confidence interval using the standard error is

$$\begin{aligned} \bar{x} &\pm 2SE \\ 3.85 &\pm 2(0.19) \\ 3.85 &\pm 0.38 \\ 3.47 &\text{ to } 4.23. \end{aligned}$$

For this bootstrap distribution, the 95% confidence interval using the 2.5%-tile and 97.5%-tile is 3.47 to 4.23. We see that the results (rounding to two decimal places) are the same. We are 95% confident that the average tip left at this restaurant is between \$3.47 and \$4.23.



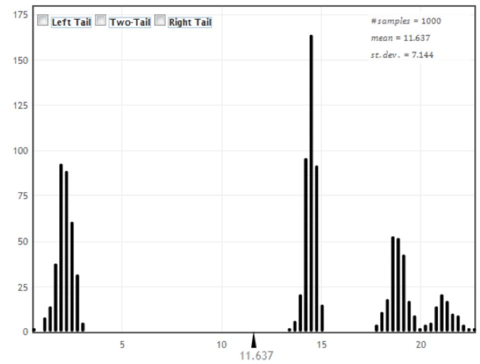
3.116 Small Sample Size and Outliers As we have seen, bootstrap distributions are generally symmetric and bell-shpaed and centered at the value of the original sample statistic. However, strange things can happen when the sample size is small and there is an outlier present. Use *StatKey* or other technology to create a bootstrap distribution for the standard deviation based on the following data:

8 10 7 12 13 8 10 50

Describe the shape of the distribution. Is it appropriate to construct a confidence interval from this distribution? Explain why the distribution might have the shape it does.

Solution

The bootstrap distribution for the standard deviations (shown below) has at least four completely separate clusters of dots. It is not at all symmetric and bell-shaped so it would not be appropriate to use this bootstrap distribution to find a confidence interval for the standard deviation. The clusters of dots represent the number of times the outlier is included in the bootstrap sample (with the cluster on the left containing statistics from samples in which the outlier was not included, the next one containing statistics from samples that included the outlier once, the next one containing statistics from samples that included the outlier twice, and so on.)



4.17 Beer and Mosquitoes Does consuming beer attract mosquitoes? A study done in Burkino Faso, Africa, about the spread of malaria investigated the connection between beer consumption and mosquito attraction. In the experiment, 25 volunteers consumed a liter of beer while 18 volunteers consumed a liter of water. The volunteers were assigned to the two groups randomly. The attractiveness to mosquitos of each volunteer was tested twice: before the beer or water and after. Mosquitoes were released and caught in traps as they approached the volunteers. For the beer group, the total number of mosquitoes caught in the traps before consumption was 434 and the total was 590 after consumption. For the water group, the total was 337 before and 345 after.

(a) Define the relevant parameter(s) and state the null and alternative hypotheses for a test to see if, after consumption, the average number of mosquitoes is higher for the volunteers who drank beer.

(b) Compute the average number of mosquitoes per volunteer before consumption for each group and compare the results. Are the two sample means different? Do you expect that this difference is just the result of random chance?

(c) Compute the average number of mosquitoes per volunteer after consumption for each group and compare the results. Are the two sample means different? Do you expect that this difference is just the result of random chance?

(d) If the difference in part (c) is unlikely to happen by random chance, what can we conclude about beer consumption and mosquitoes?

(e) If the difference in part (c) is statistically significant, do we have evidence that beer consumption increases mosquito attraction? Why or why not?

Solution

(a) We define μ_b to be the mean number of mosquitoes attracted after drinking beer and μ_w to be the mean number of mosquitoes attracted after drinking water. The hypotheses are:

$$H_0 : \mu_b = \mu_w$$

$$H_a : \mu_b > \mu_w$$

(b) The sample mean number of mosquitoes attracted per participant before consumption for the beer group is $434/25 = 17.36$ and is $337/18 = 18.72$ for the water group. These sample means are slightly different, but the small difference could be attributed to random chance.

(c) The sample mean number of mosquitoes attracted per participant after consumption is 590/25 = 23.60 for the beer group and is 345/18 = 19.17 for the water group. This difference is larger than the difference in means before consumption. It is less likely to be due just to random chance.

(d) The mean number of mosquitoes attracted when drinking beer is higher than when drinking water.

(e) Since this was an experiment, a statistically significant difference would provide evidence that beer consumption increases mosquito attraction.

4.18 Guilty Verdicts in Court Cases A reporter on cnn.com stated in July 2010 that 95% of all court cases that go to trial result in a guilty verdict. To test the accuracy of this claim, we collect a random sample of 2000 court cases that went to trial and record the proportion that resulted in a guilty verdict.

(a) What is/are the relevant parameter(s)? What sample statistic(s) is/are used to conduct the test?
(b) Stat the null and alternative hypotheses.
(c) We assess evidence by considering how likely our sample results are when H_0 is true. What does that mean in this case?

Solution
(a) The parameter is p, the proportion of all court cases going to trial that end in a guilty verdict. The sample statistic is $\hat{p}$, the proportion of guilty verdicts in the sample of 2000 cases.

(b) The hypotheses are:

$$H_0 : p = 0.95$$

$$H_A : p \neq 0.95$$

(c) How likely is the observed sample proportion when we select a sample of size 2000 from a population with $p = 0.95$?

For exercises 4.21 to 4.25, describe tests we might conduct based on Data 2.3, introduced on page 66. This dataset, stored in **ICUAdmissions**, contains information about a sample of patients admitted to a hospital Intensive Care Unit (ICU). For each of the research questions below, define any relevant parameters and state appropriate null and alternative hypotheses.

4.21 Is there evidence that mean heart rate is higher in male ICU patients than in female ICU patients?

Solution
We define μ_m to be mean heart rate for males being admitted to an ICU and μ_f to be mean heart rate for females being admitted to an ICU. The hypotheses are:

$$H_0 : \mu_m = \mu_f$$

$$H_A : \mu_m > \mu_f$$

4.22 Is there a difference in the proportion who receive CPR based on whether the patient's race is white or black?

<u>Solution</u>
We define p_w to be the proportion of white ICU patients who receive CPR and p_b to be the proportion of black ICU patients who receive CPR. The hypotheses are:

$$H_0 : p_w = p_b$$

$$H_A : p_w \neq p_b$$

4.23 Is there a positive linear association between systolic blood pressure and heart rate?

<u>Solution</u>
We define ρ to be the correlation between systolic blood pressure and heart rate for patients admitted to an ICU. The hypotheses are:

$$H_0 : \rho = 0$$

$$H_A : \rho > 0$$

Note: The hypotheses could also be written in terms of β, the slope of a regression line to predict one of these variables using the other.

4.24 Is either gender over-representative in patients to the ICU or is the gender breakdown about equal?

<u>Solution</u>
Notice that this is a test for a single proportion. We define p to be the proportion of ICU patients who are female. (We could also have defined p to be the proportion who are male. The test will work fine either way.) The hypotheses are:

$$H_0 : p = 0.5$$

$$H_A : p \neq 0.5$$

Also accepted: We define p_m to be the proportion of ICU patients who are male and p_f to be the proportion of ICU patients who are female. The hypotheses are:

$$H_0 : p_m = p_f$$

$$H_A : p_m \neq p_f$$

4.25 Is the average age of ICU patients at this hospital greater than 50?

<u>Solution</u>
We define μ to be the mean age of ICU patients. The hypotheses are:

$$H_0 : \mu = 50$$

$$H_A : \mu > 50$$

For exercises 4.30, 4.32, and 4.36, indicate whether the analysis involves a statistical test. If it does involve a statistical test, state the population parameter(s) of interest and null hypotheses.

4.30 Polling 1000 people in a large community to determine the average number of hours a day people watch television

Solution

This analysis does not involve a test because there is no claim of interest. We would likely use a confidence interval to estimate the average.

4.32 Utilizing the census of a community, which includes information about all residents of the community, to determine if there is evidence for the claim that the percentage of people in the community living in a mobile home is greater than 10%.

Solution

This analysis does not include a test because from the information in a census, we can find exactly the true population proportion.

4.36 Using the complete voting record of a county to see if there is evidence that more than 50% of the eligible voters in the county voted in the last election.

Solution

This analysis does not include a statistical test. Since we have all the information for the population, we can compute the proportion who voted exactly and see if it is greater than 50%.

4.40 Euchre One of the authors and some statistician friends have an ongoing series of Euchre games that will stop when one of the two teams is deemed to be *statistically significantly* better than the other team. Euchre is a card game and each game results in a win for one team and a loss for the other. Only two teams are competing in this series, which we'll call Team A and Team B.

(a) Define the parameter(s) of interest.

(b) What are the null and alternative hypotheses if the goal is to determine if either team is statistically significantly better than the other at winning Euchre?

(c) What sample statistic(s) would they need to measure as the games go on?

(d) Could the winner be determined after one or two games? Why or why not?

Solution

(a) The population of interest is all Euchre games that could be played between these two teams. The parameter of interest is the proportion of games that a certain team would win, say $p =$the proportion of all possible games that team A wins. (We could also just as easily have used team B.)

(b) We are testing to see whether this proportion is either significantly higher or lower than 0.5. The hypotheses are:

$$H_0 : p = 0.5$$

$$H_A : p \neq 0.5$$

(c) The sample statistic is the proportion of games played so far that team A has won. We could choose to look at the proportion of wins for either team, but must be consistent defining the population parameter and calculating the sample statistic. We also need to keep track of the sample size (number of games played).

(d) No. Even if the two teams are equal ($p = 0.5$), it is quite possible that one team could win the first two games just by random chance. Therefore, even if one team wins the first two games, we would not have conclusive evidence that that team is better.

4.52 Arsenic in Chicken Data 4.5 on page 228 discusses a test to determine if the mean level of arsenic in chicken meat is about 80 ppb. if a restaurant chain finds significant evidence that the mean arsenic level is above 80, the chain will stop using that supplier of chicken meat. They hypotheses are

$$H_0 : \mu = 80$$

$$H_A : \mu > 80$$

where μ represent the mean arsenic level in all chicken meat from that supplier. Samples from two different suppliers are analyzed, and the resulting p-values are given:

Sample from Supplier A: p-value is 0.00003

Sample from Supplier B: p-value is 0.3500

(a) Interpret each p-value in terms of the probability of the results happening by random chance.
(b) Which p-value shows stronger evidence for the alternative hypothesis? What does this mean in terms of arsenic and chickens?
(c) Which supplier, A or B, should the chain get chickens from in order to avoid too high a level of arsenic?

Solution

(a) If the mean arsenic level is really 80 ppb, the chance of seeing a sample mean as high (or higher) than was observed in the sample from supplier A by random chance is only 0.0003. For supplier B, the corresponding probability (seeing a sample mean as high as B's when $\mu = 80$) is 0.35.

(b) The smaller p-value for Supplier A provides stronger evidence against the null hypothesis and in favor of the alternative that the mean arsenic level is higher than 80 ppb. Since it is very rare for the mean to be that large when $\mu = 80$, we have stronger evidence that there is too much arsenic in Supplier A's chickens.

(c) The chain should get chickens from Supplier B, since there is strong evidence that Supplier A's chicken have a mean arsenic level above 80 ppb which is unacceptable.

4.60 Smiles and Leniency Data 4.2 on page 223 describes and experiment to study the effects of smiling on leniency in judging students accused of cheating. The full data are in **Smiles**. in Example 4.2 we consider hypotheses $H_0 : \mu_s = \mu_n$ vs $H_A : \mu_s > \mu_n$ to test if the data provide evidence that average leniency score is higher for smiling students (μ_s) than for students with a neutral expression (μ_n). A dot plot for the difference in sample means based on 1000 random assignments of leniency scores from the original sample to smile and neutral groups is shown in the book.

(a) The difference in sample means for the original sample is $D = \bar{x}_s - \bar{x}_n = 4.91 - 4.12 = 0.79$ (as shown in Figure 4.20). What is the p-value for the one-tailed test? Hint: There are 27 dots in the tail beyond 0.79.
(b) In Example 4.3 on page 223 we consider the test with a two-tailed alternative, $H_0 : \mu_s = \mu_n$

vs $H_A : \mu_s \neq \mu_n$, where we make no assumption in advance on whether smiling helps or discourages leniency. How would the randomization distribution in the figure change for this test? How would the p-value change?

Solution
(a) This is an upper tail test, so the p-value is the proportion of randomization samples with differences more than the observed $D = 0.79$. There are 27 dots to the right of 0.79 in the plot, so the p-value is $27/1000 = 0.027$.

(b) The randomization distribution depends only on H_0 so it would not change for $H_A : \mu_s \neq \mu_n$. For a two-tailed alternative we need to double the proportion in one tail, so the p-value is 2(0.027) = 0.054.

4.62 Classroom Games Two professors at the University of Arizona were interested in whether having students actually play a game would help them analyze theoretical properties of the game. The professors perfumed an experiment in which students played one of two games before coming to class where both games were discussed. Students were randomly assignment to which of the two games they played, which we'll call Game 1 and Game 2. On a later exam, students were asked to solve problems involving both games, with Question 1 referring to Game 1 and Question 2 referring to Game 2. When comparing the performance of the two groups on the exam question related to Game 1, they suspected that the mean for students who had plead Game 1 (μ_1) would be higher than the mean for the other students μ_1, so they considered the hypotheses $H_0 : \mu_1 = \mu_2$ vs $H_A : \mu_1 > \mu_2$.

(a) The paper states: "the test of difference in means results in a p-value of 0.7619". Do you think this provides sufficient evidence to conclude that playing Game 1 helped student performance on that exam question? Explain.

(b) If they were to repeat this experiment 1000 times, and there really is no effect from playing the game, roughly how many times would you expect the results to be as extreme as those observed in the actual study?

(c) When testing a difference in mean performance between the two groups on exam Question 2 related to Game 2 (so now the alternative is reversed to be $H_A : \mu_1 < \mu_2$ where μ_1 and μ_2 represent the mean on Question 2 for the respective groups), they computed a p-value of 0.5490. Explain what it means (in the context of this problem) for both p-values to be greater than 0.5.

Solution
(a) A p-value of 0.7619 is not at all small, so the difference in means between the two groups is not significant. Thus there is insufficient evidence to conclude that playing Game 1 helped those students on the exam question about Game 1.

(b) The p-value (0.7619) measures the chance of seeing results so extreme when H_0 is true, so we would expect about 762 out of 1000 experiments to be this extreme if there is no effect.

(c) If both p-values for the one tail tests are greater than 0.5, it means the differences in the sample means were in the opposite direction of H_A in both cases. So for both questions the students who did not play the game in class actually had the higher mean score on the exam question related to the game. This was a very surprising result!

4.63 Classroom Games: Is One Question Harder? Exercise 4.62 describes an experiment involving playing games in class. One concern in the experiment is that the exam question related to Game 1 might be a lot easier or harder than the question for Game 2. In fact, when they compared the mean performance of all students on Question 1 to Question 2 (using a two-tailed test for a difference in means), they report a p-value equal to 0.0012.

(a) If you were to repeat this experiment 1000 times, and there really is no difference in the difficult of the questions, how often would you expect the means to be as different as observed in the actual study?

(b) Do you think this p-value indicates that there is a difference in the average difficult of the two questions? Why or why not?

(c) Based on the information given, can you tell which (if either) of the two questions is easier?

Solution

(a) For the small p-value of 0.0012, we expect about $0.0012 \times 1000 = 1.2$ or about one time out of every 1000 to be as extreme as the difference observed, if the questions are equally difficult.

(b) The p-value is very small, so seeing this large a difference would be very unusual if the two questions really were equally difficult. Thus we conclude that there is a difference in the average difficulty of the two questions.

(c) There is nothing in the information given that indicates which question had the higher mean, so we can't tell which of the two questions is the easier one.

Computer Exercises

For each R problem, turn in answers to questions with the written portion of the homework. Send the R code for the problem to Katherine Goode. The answers to questions in the written part should be well written, clear, and organized. The R code should be commented and well formatted.

R problem 1 Consider again the data in 3.116. Using 1,000,000 bootstrap replicates, use the boot() function from the boot library to find 95% bootstrap confidence intervals for the population mean. Here is sample code to do this.

```
library(boot)
foo = c(8,10,7,12,13,8,10,50)
my.mean = function(x,indices) {
  return( mean(x[indices]) )
}
boot.out = boot(foo,my.mean,1000000)
boot.ci(boot.out)
```

This will take a few seconds to calculate.

1. Describe the differences among the four intervals. Where are they centered? How wide are they?

Solution

Below is the output from R showing the four confidence intervals.

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
```

```
Based on 1000000 bootstrap replicates

CALL :
boot.ci(boot.out = boot.out)

Intervals :
Level      Normal              Basic
95%   ( 5.41, 24.07 )   ( 4.25, 20.88 )

Level     Percentile            BCa
95%   ( 8.62, 25.25 )   ( 9.12, 30.50 )
Calculations and Intervals on Original Scale
```

We know that basic uses the estimated standard error, percentile uses percentiles, and BCa also uses percentiles but is adjusted to account for bias and skewness.

Below is a table showing the centers and ranges of the confidence intervals.

CI	Center	Range
Normal	14.74	18.66
Basic	12.57	16.63
Percentile	16.94	16.63
BCa	19.81	21.38

We see that the center varies from the lowest value of 12.57 for the basic interval to the highest value of 19.81for the BCa interval. We also see that the basic and percentile intervals have the same width of 16.63, while the normal and BCa intervals are wider.

```
R Code
data <- c(8,10,7,12,13,8,10,50)
mean(data)
library(boot)
my.mean = function(x,indices) {
  return( mean(x[indices]) )
}
boot.out = boot(foo,my.mean,1000000)
boot.ci(boot.out)
```

2. The variable boot.out$t contains the sampled 1,000,000 means. Check to see that the "normal" bootstrap interval agrees with the mean plus or minus 1.96 times the estimated standard error. (Note the use of 1.96 instead of 2, which comes from a more precise calculation of the quantiles for the middle 95% of a perfect normal distribution.)

Solution

We obtain the following output from R.

```
Bootstrap Statistics :
    original      bias    std. error
t1*    14.75 0.00180325    4.756461
```

From it, we are able to calculate the normal bootstrap interval by hand.

$$\begin{array}{ccc} 14.75 & \pm & 1.96 \times 4.756461 \\ (5.427336 & , & 24.072664) \end{array}$$

We see that this agrees with the normal confidence interval calculated by R in part 1.

```
R Code
14.75+2*sd(boot.out$t)
c(mean(data)-1.96*4.756461,mean(data)+1.96*4.756461)
```

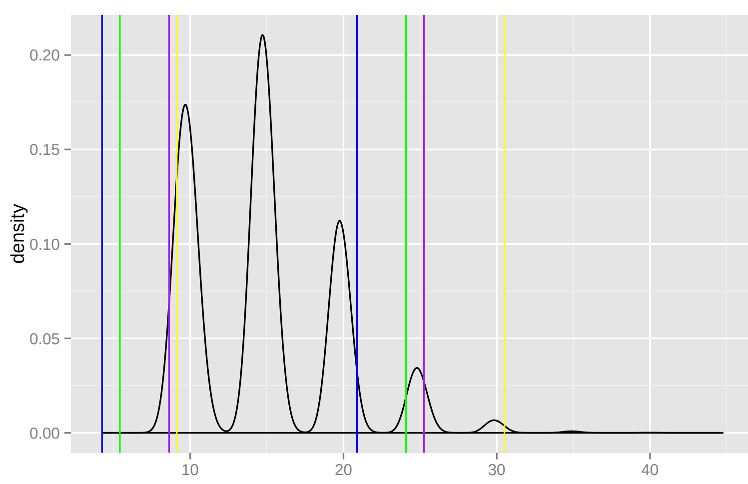
3. Check to see that the percentile bootstrap matches what you find using quantile() on the sampled means.

Solution
Using R, we find that the 2.5th percentile is 8.625, and the 9.75th percentile is 25.25. This agrees with the percentile confidence interval calculated by R in part 1.

```
R Code
quantile(boot.out$t,0.025)
quantile(boot.out$t,0.975)
```

4. Make a density plot of the sampled means. Describe the appearance of the bootstrap distribution. Add to this plot an indication of the endpoints of each confidence interval.

Solution
Below is density plot showing the bootstrap distribution. We see that it is not bell shaped or symmetric. In fact, it appears to be slightly skewed to the right, and it has several peaks.



The colored lines indicate the endpoints of the confidence intervals. The colors represent the four confidence intervals in the following manner:

Normal ∼ Green Basic ∼ Blue Percentile ∼ Purple BCa ∼ Yellow

```
R Code
ggplot(data.frame(x = boot.out$t), aes(x = x)) + geom_density()+
geom_vline(xintercept=c( 5.41, 24.07),color="green")+
geom_vline(xintercept=c( 4.25, 20.88),color="blue")+
geom_vline(xintercept=c(8.62, 25.25),color="purple")+
geom_vline(xintercept=c( 9.12, 30.50 ),color="yellow")
```

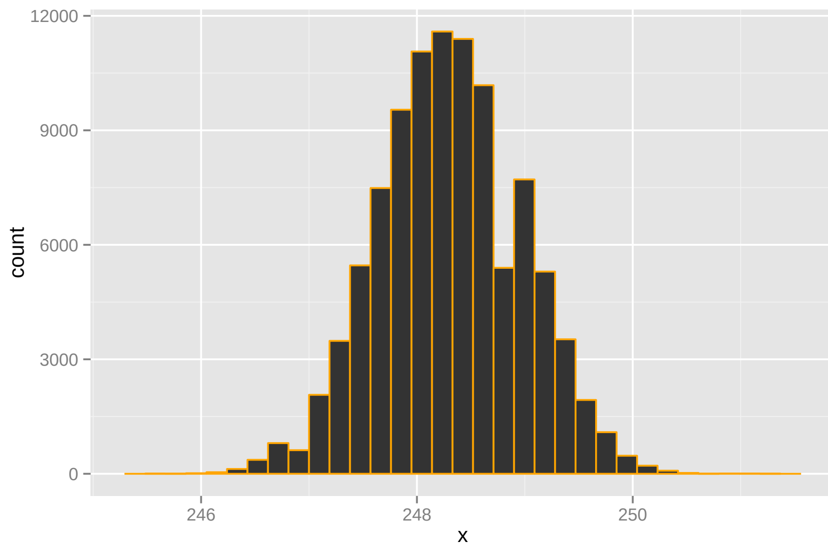
5. How does the BCa confidence interval differ most from the others?

Solution
The BCa confidence interval is much high than the other three confidence intervals. It seems that this may be so since the BCa interval adjusts for bias and skew. Our distribution has these features, and thus, it seems natural that the BCa interval would be different than the other intervals.

R problem 2 The data set CaffeineTaps contains two samples of size 10 where each value is the number of times a student tapped his finger in a minute. One group had consumed 200mg of caffeine in coffee two hours before the experiment, and one group had decaf. Details of the experiment are on page 240.

1. Use the bootstrap to estimate the mean number of taps per minute with a 95% confidence interval separately for each group. Use both the SE and percentile versions.

Solution
We begin by creating a bootstrap distribution to estimate the mean number of taps for the caffeine group. We first calculate the sample mean, which is 248.3. We then take 10,000 samples of size 10 with replacement and calculate the mean of each sample. The distribution is shown in the histogram below.



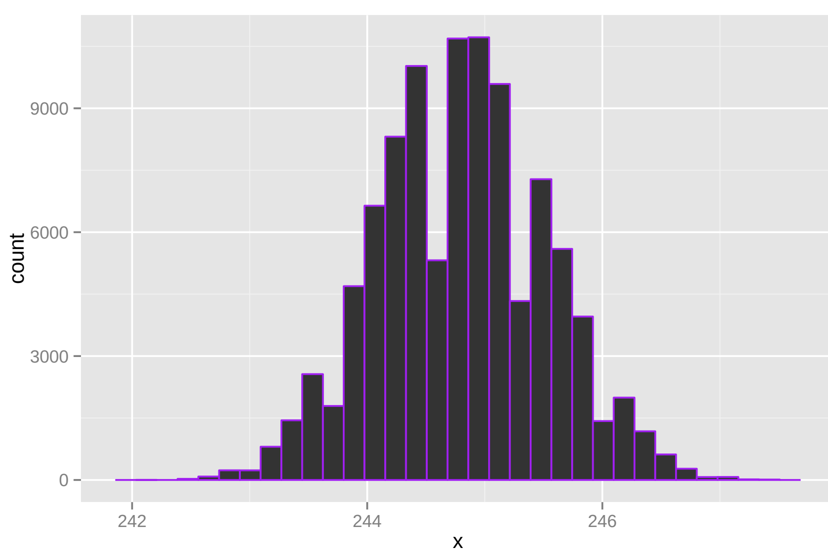
In R, we are able to determine that the SE of the distribution is 0.66. Thus, we calculate a 95% confidence interval for the population mean of taps for the caffeine drinkers as

$$248.3 \pm 2 * 0.66 \rightarrow (246.97, 249.63)$$

Using the 2.5th and 97.5th quantiles, we find a confidence interval for the population mean of taps for the caffeine drinkers to be

$$(247.0, 249.6)$$

We do the same procedure for the students who drank decaf. Below is a histogram of the bootstrap distribution.



The mean of the sample is 244.8, and the SE of the distribution is 0.72. Thus, we calculate a 95%

confidence interval for the population mean of taps for the decaf drinkers as

$$244.8 \pm 2 * 0.72 \rightarrow (243.36, 246.24)$$

Using the 2.5th and 97.5th quantiles, we find a confidence interval for the population mean of taps for the decaf drinkers to be

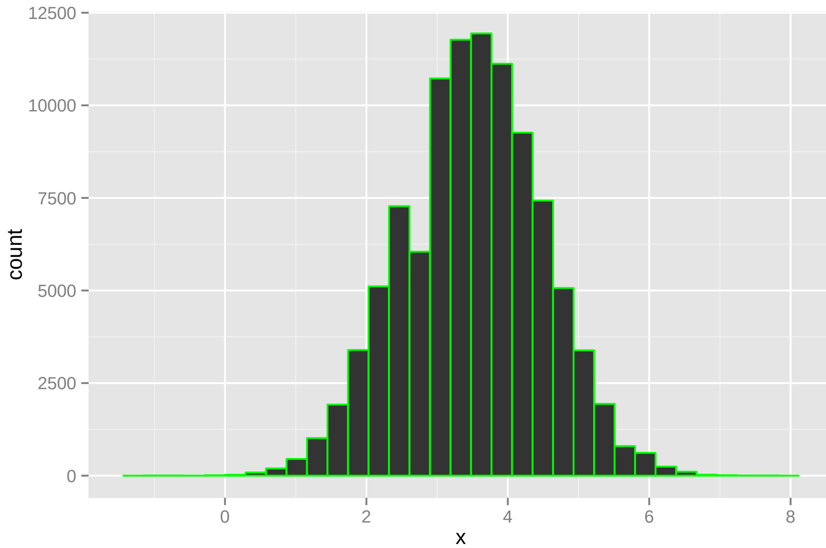
$$(243.4246.2)$$

```
R Code

library(Lock5Data)
data(CaffeineTaps)
str(CaffeineTaps)

caffeine = with(CaffeineTaps, Taps[Group=="Caffeine"])
decaf = with(CaffeineTaps, Taps[Group=="NoCaffeine"])
caffeine.mean = mean(caffeine)
decaf.mean = mean(decaf)
caffeine.n = length(caffeine)
decaf.n = length(decaf)
caffeine.mean
decaf.mean

B = 100000
caffeine.boot1 = numeric(B)
for ( i in 1:B ) {
  caffeine.boot1[i] = mean(sample(caffeine,size=caffeine.n,replace=TRUE))
}
ggplot(data.frame(x = caffeine.boot1), aes(x = x))+geom_histogram(color="orange")
CI1 <- caffeine.mean+c(1,-1)*2*sd(caffeine.boot1)
CI2 <- quantile(caffeine.boot1,c(0.025,0.975))

B = 100000
decaf.boot1 = numeric(B)
for ( i in 1:B ) {
  decaf.boot1[i] = mean(sample(decaf,size=decaf.n,replace=TRUE))
}
ggplot(data.frame(x = decaf.boot1), aes(x = x))+geom_histogram(color="purple")
CI3 <- decaf.mean+c(1,-1)*2*sd(decaf.boot1)
CI4 <- quantile(decaf.boot1,c(0.025,0.975))
```

2. Use the bootstrap to estimate the difference, caffeine minus decaf, in the population means, with a 95% confidence interval. Use both the SE and percentile versions.

Solution
We now want to create confidence intervals for the difference in population means of the caffeine minus decaf group. The observed difference in means is 3.5. We create a bootstrap distribution by taking 10,000 random samples of size 10 with replacement from both the caffeine drinkers' data and from the decaf drinkers' data. We then take the mean of each of the samples, subtract the

decaf means from the caffeine means, and the resulting values compose our bootstrap distribution. We perform this process, and below is a histogram of the bootstrap distribution.



We first calculate a confidence interval by using the standard error, which we compute to be 0.98 from this bootstrap distribution. Thus, our confidence interval is

$$3.5 \pm 2 * 0.98 \rightarrow (1.53, 5.47)$$

Using the 2.5th and 97.5th quantiles, we find a confidence interval for the population mean of taps for the decaf drinkers to be

$$(1.6, 5.4)$$

```
R Code
caffeine.mean - decaf.mean
B = 100000
caffeine.boot = numeric(B)
decaf.boot = numeric(B)
for ( i in 1:B ) {
  caffeine.boot[i] = mean(sample(caffeine,size=caffeine.n,replace=TRUE))
  decaf.boot[i] = mean(sample(decaf,size=decaf.n,replace=TRUE))
}
boot.stat = caffeine.boot - decaf.boot
ggplot(data.frame(x = boot.stat), aes(x = x))+geom_histogram(color="green")
CI5 <- (caffeine.mean - decaf.mean)+c(-1,1)*2*sd(boot.stat)
CI6 <- quantile(boot.stat, c(0.025,0.975))
```

3. Compare the widths of the confidence intervals for the individual population means and for the difference in population means. What is the ratio of these widths? Which is larger?

Solution

The widths and ratios for the six confidence intervals computed in parts 1 and 2 are as follows:

Confidence Interval	Width	Ratio
Caffeine SE	2.7	1.02
Caffeine Quantiles	2.6	
Decaf SE	2.9	1.03
Decaf Quantiles	2.8	
Caffeine-Decaf SE	3.9	1.03
Caffeine-Decaf Quantiles	3.8	

By comparing the the widths of the confidence intervals for the individual population means and for the difference in population means, we see that the confidence intervals constructed from the difference in population means are wider.

```
R Code
CI1[2]-CI1[1]
CI2[2]-CI2[1]
CI3[2]-CI3[1]
CI4[2]-CI4[1]
CI5[2]-CI5[1]
CI6[2]-CI6[1]

(CI1[2]-CI1[1])/(CI2[2]-CI2[1])
(CI3[2]-CI3[1])/(CI4[2]-CI4[1])
(CI5[2]-CI5[1])/(CI6[2]-CI6[1])
```

4. Write an interpretation of one of the confidence intervals for the difference in means.

<u>Solution</u>
We are 95% percent confident that the difference in populations means of number of finger taps per minute of students who drink decaf subtracted from the students who drink caffeine is between 1.53 and 5.47 taps.