HW#7 87 F21

Patrik Schweika

12/3/2021

```
knitr::opts_chunk$set(echo = TRUE)
```

Question 1. A hypothesis test

Write your text answers here

In my case p-value was 0.018 which is less than 0.05, so I reject the null hypothesis that 50% students would say they are better drivers than average, in favor of alternative hypothesis that more than 50% students would say they are better drivers.

```
# Suppose you surveyed a random sample of 200 UVM students, and 115 of them said they are 'Be
tter than Average' or "Way Better than Average" drivers.    Do a hypothesis test to see if th
e true proportion of UVM students who would say they are better or way better drivers is a ma
jority.   That is, use the null hypothesis "50% of students would say they are better or way
 better than average".   Generate 1000 trials, compute the 'p-value' and state your decision
 (above), in terms of the problem.

Trials <- rbinom(n = 1000, size = 200, prob = 0.5)

table(Trials)
```

```
## Trials
##  76  77  80  81  82  83  85  86  87  88  89  90  91  92  93  94  95  96  97  98
##   1   1   2   4   3   5   7   4  13  16  17  19  26  28  37  50  57  46  45  48
##  99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 119
##  51  53  38  44  46  48  53  49  36  34  26  25  11  19  15   8   7   3   4   1
```

```
p <- (4 + 5 + 2 + 2 + 2 + 2 + 1) / 1000
p
```

```
## [1] 0.018
```

Question 2. Confidence Intervals for Percentages

part a – Driving

The point estimate was given to be 115/200 which is 57.5%.

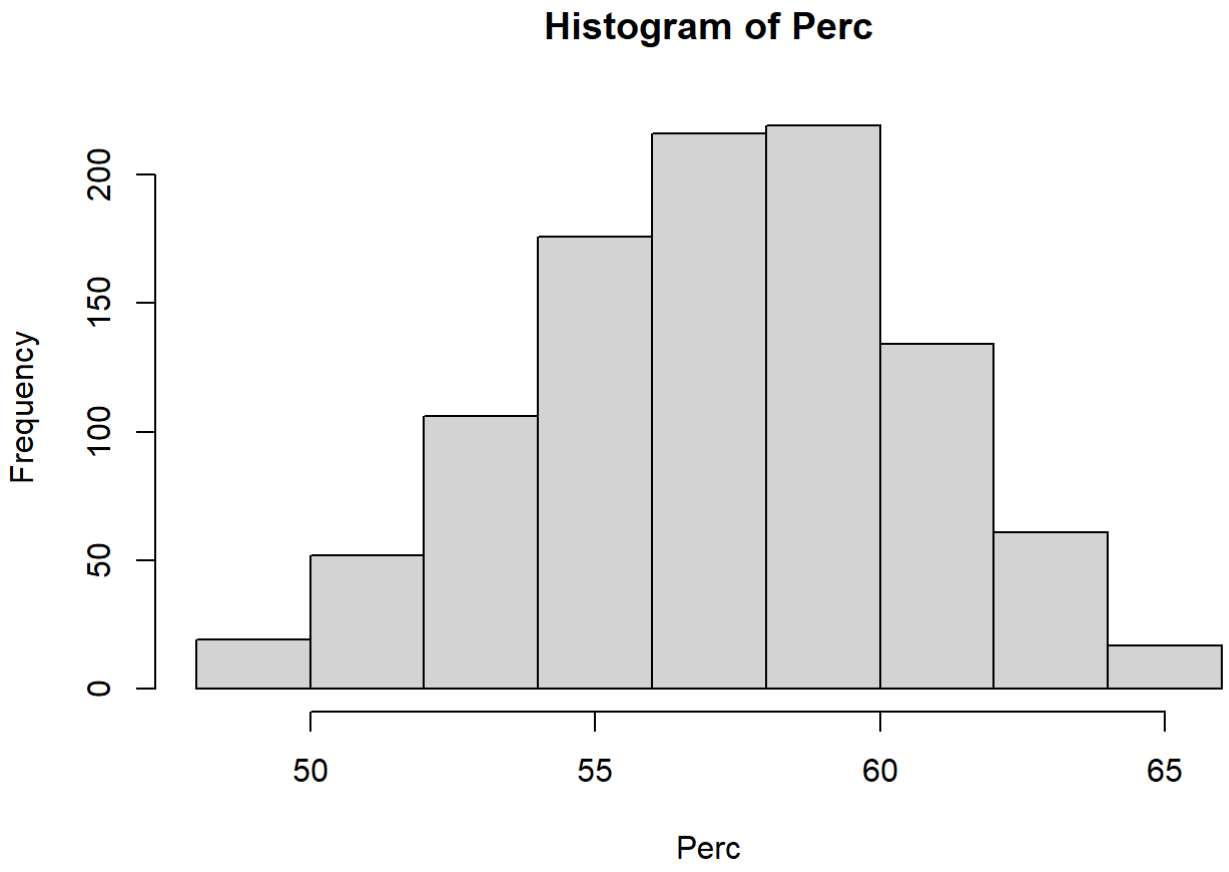
I'm 95% confident that the percentage of UVM students who say they are better driver than average is captured in the interval from 50.83% to 64.17%.

I believe that majority of UVM students will say they are better drivers than average, since the whole 95% confidence interval is above 50%. I'm 95% sure about this conclusion.

```
# Use the same data as in question 1 (a random sample of 200 UVM students, and 115 of them sa
id they are 'Better than Average' or "Way Better than Average" drivers.)    Using methods fro
m class, find a 95% bootstrap confidence interval for the true percentage of students.  Have
 R calculate and print the lower and upper limit, rounding each value to 2 decimal places. (x
x.xx  to xx.xx).  Also have R do a histogram of the bootstrap percentages.

p = 115/200

Trials <- rbinom(n = 1000, size = 200, prob = p)
Perc <- 100 * Trials / 200

PE <- 100 * p

SE <- sd(Perc)
ME <- 2 * SE

low <- PE - ME
high <- PE + ME

low <- round(low, 2)
high <- round(high, 2)

print(paste("(", low, " to ", high, ")"))
```

```
## [1] "( 50.62  to  64.38 )"
```

```
hist(Perc)
```



```
# In the text before this chunk, state the point estimate, and write a sentence of interpreta
tion above (I'm 95% sure…..)  Based on your interval, do you believe that a majority of all U
VM students would say they are better or way better than average drivers?   How do you know?
How sure are you of the conclusion about a majority? (write answers above)
```

part b – Favorite app

Point estimate was given to be 50/200 which is 25%.

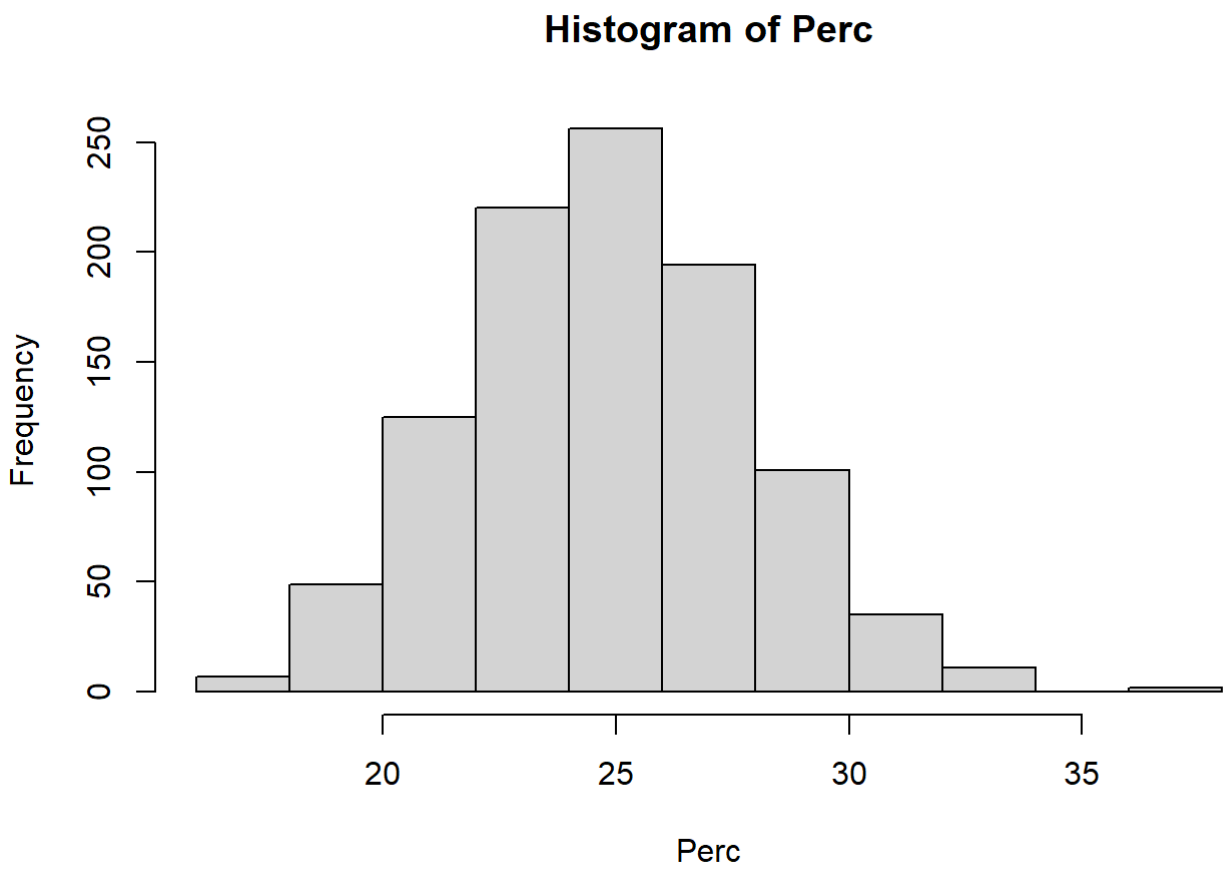
I'm 95% confident that the percentage of UVM students who say that their favorite app is Instagram is in interval from 18.84% to 31.16%.

I believe that minority of UVM students will say that their favorite app is Instagram, since the whole 95% confidence interval is under 50%. I'm 95% sure about this conclusion.

```
# Suppose in the same random sample of 200 UVM students, 50 of them said their favorite app i
s Instagram (the most popular app in this sample).    Using methods from class, find a 95% bo
otstrap confidence interval for the true percentage of students who prefer Instagram.  Have R
calculate and print the lower and upper limit, rounding each value to 2 decimal places. (xx.x
x  to xx.xx). Also have R do a histogram of the bootstrap percentages.

p = 50/200

Trials <- rbinom(n = 1000, size = 200, prob = p)
Perc <- 100 * Trials / 200

PE <- 100 * p

SE <- sd(Perc)
ME <- 2 * SE

low <- PE - ME
high <- PE + ME

low <- round(low, 2)
high <- round(high, 2)

print(paste("(", low, " to ", high, ")"))
```

```
## [1] "( 18.91  to  31.09 )"
```

```
hist(Perc)
```



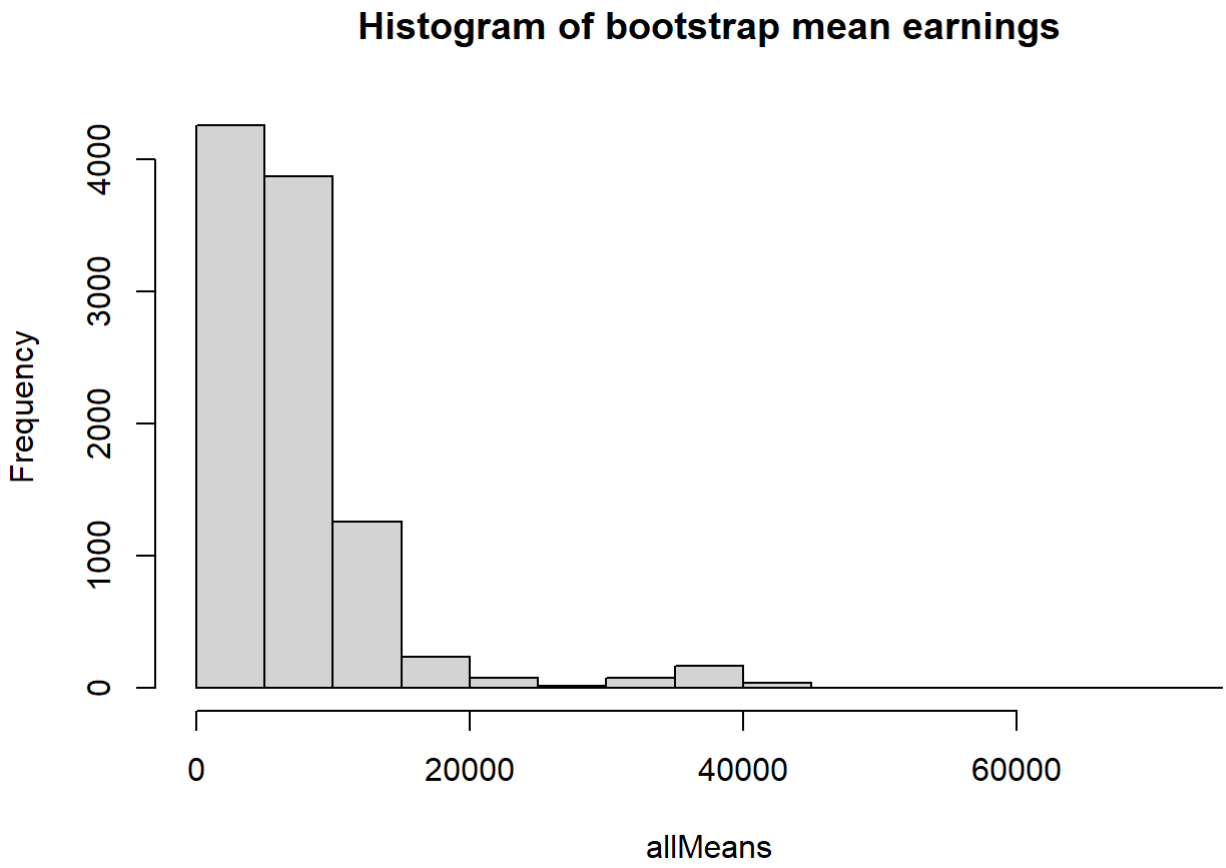

```
# In the text before this chunk, state the point estimate, and write a sentence of interpreta
tion above (I'm 95% sure…..)  Based on your interval, do you believe that a minority of all
 UVM students (less than 50%) like Instagram the best?  How do you know?  How sure are you
 of the conclusion about a minority? (write answers above)
```

Question 3 – Confidence Intervals for Means

part a - CI for mean earnings

I'm 95% confident that the true mean earnings in dollars for UVM students is in interval from -5768.12 to 20685.28.

```
# Use the attached data file surveyC_S21.csv, download it, and read in as a data frame calle
d, s.  Assume the data is a random sample of 200 UVM students.  Using the bootstrap method, a
nd code similar to the code from class, find a 95% confidence interval for the true mean earn
ings for UVM students.  Have R do a histogram of the bootstrap means, calculate the standard
 deviation of the bootstrap means (the SE), calculate the margin of error, and Upper and Lowe
r limits of the confidence interval.  Have R calculate and print the lower and upper limit, r
ounding each value to 2 decimal places. (xx.xx  to xx.xx).

s <- read.csv("surveyC_F21.csv")

earnings <- na.omit(s$Earnings)

allMeans <- c()

for (i in 1:10000) {
  samp <- sample(earnings, size = length(s), replace = TRUE)
  allMeans[i] = mean(samp)
}

hist(allMeans, main = "Histogram of bootstrap mean earnings")
```



```
PE <- mean(earnings)
SE <- sd(allMeans)
ME <- 2 * SE

low <- PE - ME
high <- PE + ME

low <- round(low, 2)
high <- round(high, 2)

PE
```

```
## [1] 7458.578
```

```
print(paste("(", low, " to ", high, ")"))
```

```
## [1] "( -5693.26  to  20610.42 )"
```

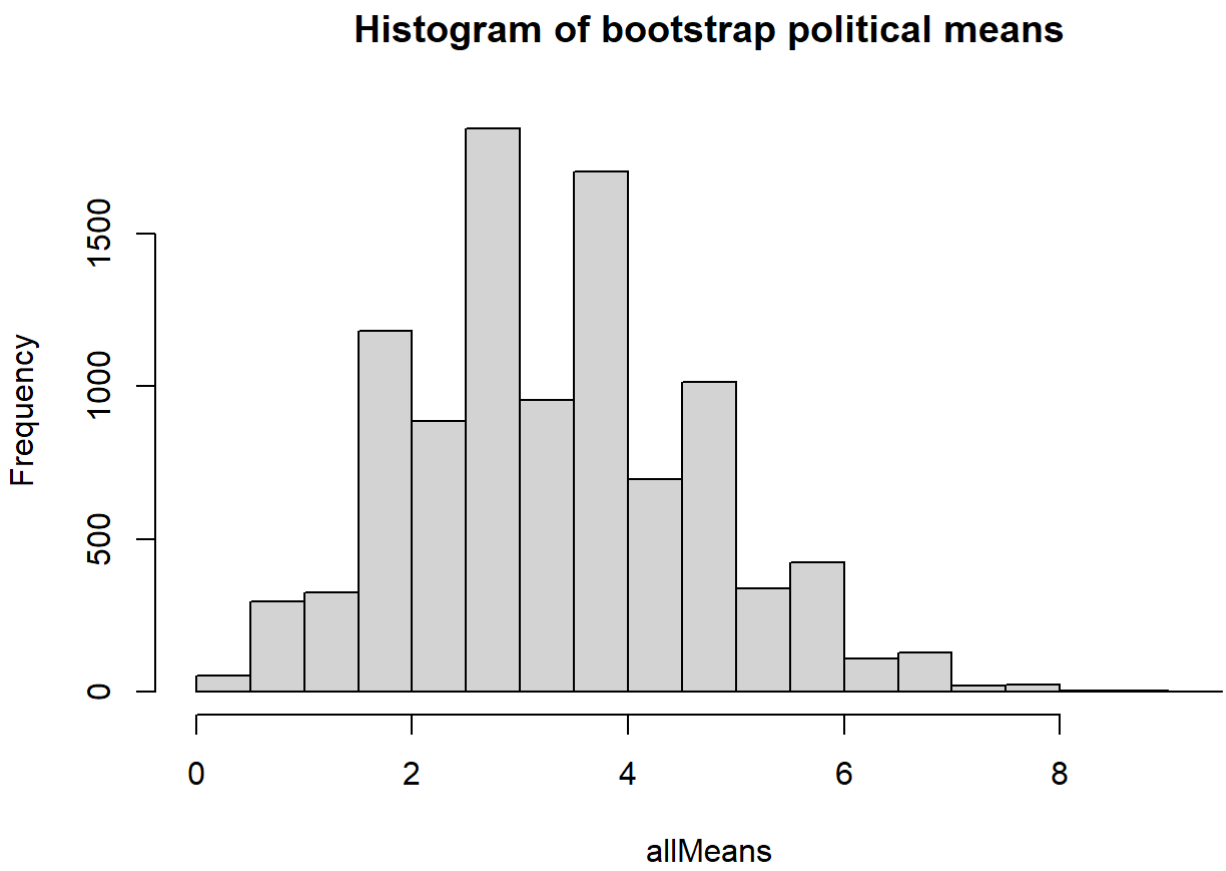
```
#  Above this code chunk, state the point estimate, then state the interval in a complete sen
tence in terms of the problem, as we've done in class.  (I'm 95% sure…..)  (Hint:  Remember
that you need to specify that Earnings is in the data frame, s:  os <- s$Earnings and you ma
y need to remove missing values)
```

part b - CI for mean political leaning

I'm 95% confident that the true mean of political standing is in interval from 0.71 to 6.1.

We can say that the UVM students are more liberal on average since the 95% confidence interval mostly contains the lower half of political standing (0 to 5) which is the more liberal side.

```
# Once you have your code working for problem 3a, change it so that you can find the bootstra
p confidence interval for the true mean response to the question 'Circle where you political
 leaning falls, where 0 is most LIBERAL and 10 is most CONSERVATIVE.' Produce the same outpu
t, described above. (Include the same output, including the histogram of the bootstrap means.

political <- na.omit(s$Political)

allMeans <- c()

for (i in 1:10000) {
  samp <- sample(political, size = length(s), replace = TRUE)
  allMeans[i] = mean(samp)
}

hist(allMeans, main = "Histogram of bootstrap political means")
```




```
PE <- mean(political)
SE <- sd(allMeans)
ME <- 2 * SE

low <- PE - ME
high <- PE + ME

low <- round(low, 2)
high <- round(high, 2)

PE
```

```
## [1] 3.40201
```

```
print(paste("(", low, " to ", high, ")"))
```

```
## [1] "( 0.69  to  6.11 )"
```

```
#  Above this code chunk, state the point estimate, and state the interval in a complete sent
ence in terms of the problem, as we've done in class.  (I'm 95% sure…..)   If a value of 5 is
'average' or 'centrist," does your confidence interval allow you to say that UVM students con
sider themselves to be more liberal than average?  Explain..
```

Question 4 – Confidence Intervals for Medians

part a - CI for Earnings median

I'm 95% confident that the true median earnings in dollars for UVM students is in interval from -3954.33 to 13954.33.

```
# a.  Once you have your code working for problem 3, copy and change it so that you can find
 the bootstrap confidence interval for the true median earnings here.  Above, state the point
estimate and write the Earnings confidence interval in words, in terms of the problem here, a
s we've done in class.  (I'm 95% sure...)  You don't need to print bootstrap histograms here.

allMedians <- c()

for (i in 1:10000) {
  samp <- sample(earnings, size = length(s), replace = TRUE)
  allMedians[i] = median(samp)
}

PE <- median(earnings)
SE <- sd(allMedians)
ME <- 2 * SE

low <- PE - ME
high <- PE + ME

low <- round(low, 2)
high <- round(high, 2)

PE
```

```
## [1] 5000
```

```
print(paste("(", low, " to ", high, ")"))
```

```
## [1] "( -4320.15  to  14320.15 )"
```

part b - CI for Political median

I'm 95% confident that the true median of political standing is in interval from -0.19 to 6.19.

```
# b. Once you have your code working for problem 3, copy and change it so that you can find t
he bootstrap confidence interval for the true median political leaning here.  Above, state ea
ch point estimate, and state the interval in a complete sentence in terms of the problem, as
 we've done in class.  (I'm 95% sure…..)  you don't need to print bootstrap histograms here.

allMedians <- c()

for (i in 1:10000) {
  samp <- sample(political, size = length(s), replace = TRUE)
  allMedians[i] = median(samp)
}

PE <- median(political)
SE <- sd(allMedians)
ME <- 2 * SE

low <- PE - ME
high <- PE + ME

low <- round(low, 2)
high <- round(high, 2)

PE
```

```
## [1] 3
```

```
print(paste("(", low, " to ", high, ")"))
```

```
## [1] "( -0.23  to  6.23 )"
```

part c - CIs for medians versus means

Answer the following questions in a one short paragraph:

Were your median point estimate and mean point estimate for earnings almost the same?

Mean PE (7 458.578) of earnings is about 1.4 times higher than the median PE (5 000). Mean is much more influenced by outliers than median, so it was expected to be moved towards the higher earnings.

Were your median CI and mean CI for earnings almost the same?

Mean CI of earnings is much more broader than the median CI. Mean CI is 1.5 times broader than median CI.

Were your median point estimate and mean point estimate for political almost the same?

Median PE for political standing is 3 and mean is 3.4. Mean PE is 1.13 times higher than mean PE.

Were your median CI and mean CI for political almost the same?

In this case the median CI of political standing is a bit broader than the mean CI, but not by much. Median CI is 1.09 times broader than mean CI.

Explain above why they are very different in one case, and not a lot different in the other. (Hint: make a histogram of the original data for earnings and for political leaning.)

I think that in case of earnings the mean CI was much broader, because it contained much more outliers and the earnings also differ much more numerically, where the political standing is just a small number in range from 0 to 10. The bigger range of earnings and outliers with very high earnings can throw of the mean more easily.