STAT 120 HW 3

Victor Huang

4/18/2021

Section 3.1

3.1

parameter, μ

3.2

statistic, r

3.5

statistic, $\overline{x}$

3.8

statistic, $\hat{p} = 0.9$

3.9

statistic, $r = 0.037$

3.14

sample mean = 300, SE = 5

3.18

a. iii
b. i
c. ii

3.22

We are trying to measure the proportion of the population p by using proportion of the sample in context $\hat{p}$ which is equal to 0.66

3.26

a. A goes with n = 50, B goes with n = 100, C goes with n = 200
b. It wouldn't be surprising for the n = 20 sample as well as the n = 100 looking at the scale below.It would be more surprising, if not nearly impossible, for that to happen with the n = 500 sample.
c. The same applies for n = 50 and n = 100. However, it also wouldn't be surprising for the n = 500 to also be off by 0.05.
d. As the sample size increases, the accuracy greatly improves.

3.30

a. 0.13

b. For n = 100, min is approximately 0.05 while max is 0.23. For n = 1000, min is approximately 0.10 while the max is 0.16

c. For n = 100, the SE is about 0.045 whereas the n = 1000 has an SE around 0.015

d. It wouldn't be surprising for the n = 100 sample since it falls within the 95%. It would, however, be surprising if it was found in the n = 1000.

3.38

a. The proportion of inductees that also were performers is approximately 0.680. Meaning that $\hat{p} = 0.680$.

b. We should expect to see a bell curve centered around the 34 value.

Section 3.2

3.43

a. Estimating difference in mean $\mu_1 - \mu_2$

b.Best estimate difference in mean $\overline{x}_1 - \overline{x}_2$

3.44

a. Estimating difference in proportion $p_1 - p_2$

b. Best estimate difference in proportion $\hat{p}_1 - \hat{p}_2$

3.46

A range of plausible values ranges from 0.33 to 0.41

3.50

a. Not plausible

b. Plausible

c. Not plausible

3.51

A 95% confidence interval wold range from 0.24 to 0.40

3.59

a. The quantity being estimated is the proportion of men and women who would prefer an electric shock to thinking in solitude, this quantity can be represented with the parameters $p_1 - p_2$ with p_1 and p_2 respectively representing the proportion of male and female.

b. The notation that will best estimate this would be the statistical proportional difference provided by the sample shown in the problem. This quantity can be represented with the parameters $\hat{p}_1 - \hat{p}_2$ with $\hat{p}_1$ and $\hat{p}_2$ respectively representing the proportions of male and female.

c. A 95% confidence interval would be between 0.1087 and 0.7247

d. From the answer above, while 95% of cases should fall within the interval, a chance of "no difference" while slim, is not impossible. But it would be extremely rare given how far it is from the 95% confidence interval.

3.63

This means we are 95% sure that the proportional of US adults who think cars are a nessecity fall between the interval of 0.83 and 0.89.

3.70

a. The relevant parameter is the mean weight gain experienced years later

b. We could find the exact value of the parameter by going to each individual and measure their weight, find the difference, and find the average of all the weight gained from all the individuals.

c. A 95% confidence interval would be between 4.4lbs and 9.2lbs

d. The margin of error is plus or minus 1.2lbs. This means that 95% of all cases should fall below 9.2lbs and 4.4lbs. These values are found by adding and subtracting the maring of error from the sample mean.

Section 3.3

3.77

The sample statistic is $\hat{p} = 0.7$ with an SE = 0.1. A 95% confidence interval can be found between 0.5 and 0.9.

3.85

a. The best proportion estimate would be 0.15

b. The standard error is approximately 0.03

c. The 95% confidence interval ranges between 0.06 and 0.21. What this means is that we are 95% sure that the proportion of snails who survived the birds digestive track fall between 0.06 and 0.21.

3.87

a. The original data is the graph on the left while the bootstrap is on the right

b. An estimation for the sample mean would be around 7600. This would be represented as $\overline{x} = 7600$

c. The standard error appears to be around 200

d. The standard error of the original data appears to be larger than the standard error.

e. The 95% confidence interval would be between 7200 and 8000. What this means is that we are 95% sure that the average of football players hippocampal size falls between 7200 and 8000.

3.90

a. A reasonable estimate for the standard error would be 0.015. Considering the sample mean to be approximately 0.15, we should have a 95% range between 0.12 and 0.18. Since the margin of error is twice that of the standard error, 0.015 makes sense to be the standard error value.

b. The 95% confidence interval would be between 0.12 and 0.18. What this means is that we are 95% sure that the mean difference between kids and adults would fall between the interval of 0.12 and 0.18.

Section 3.4

3.116

a. 100

b. A 99$ confidence interval would be between 90.5 and 110. Give the graph below, we see that the SD is 4.789 and the mean is 100.104. Adding and subtracting double the SD from the mean would give us a 99% confidence interval.

3.120

a. The difference in population mean represented with $\mu_1 - \mu_2$

b. Looking at the bootstrap distribution, we see that the sample mean is around 24.5, represented as $\overline{x}_1 - \overline{x}_2 = 24.5$

c. A 99% confidence interval would be between 16 and 34.

d. What the interval means is that we are 99% sure that the difference in mean of 3-PBA concentration would fall between 16 and 34.

e. Yes it does. The interval exists between 16 t0 34. Even on the low end we still see a difference of 16. This is a significant difference and is evidence that there is indeed a difference between the mean concentration of 3-PBA before and after switching to organic.

f. Yes we can. First off, this is not an observation but an experiment. Which means that variables are isolated and only the ones in question are changed to measure difference. The experiment also reflected a significant difference (as shown in part e) for a causation relationship between organic eating and 3-PBA concentration.

3.122

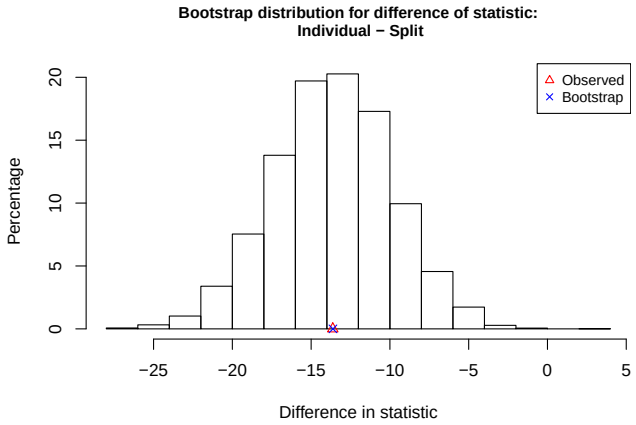
a. decrease

b. decrease

c. decrease

3.131

```
##
##  ** Bootstrap interval for difference of statistic
##
##  Observed difference of statistic:  Individual - Split =  -13.625
##  Mean of bootstrap distribution: -13.61533
##  Standard error of bootstrap distribution: 3.76538
##
##  Bootstrap percentile interval
##       2.5%      97.5%
## -21.001042  -6.291667
##
##      *--------------*
```



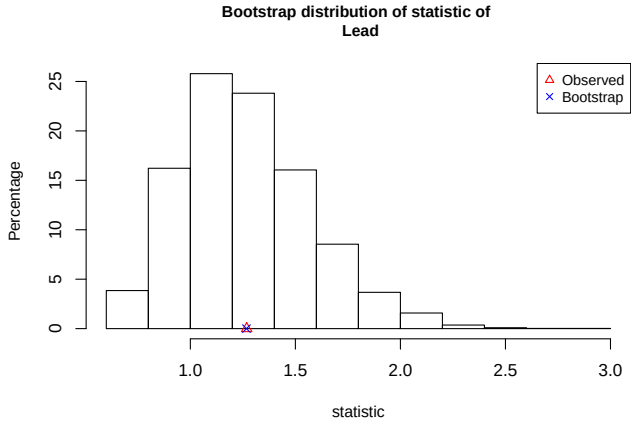
From the graph above, we can see the 95% confidence intervals would be between -24 to -5.

Section 4.8

1

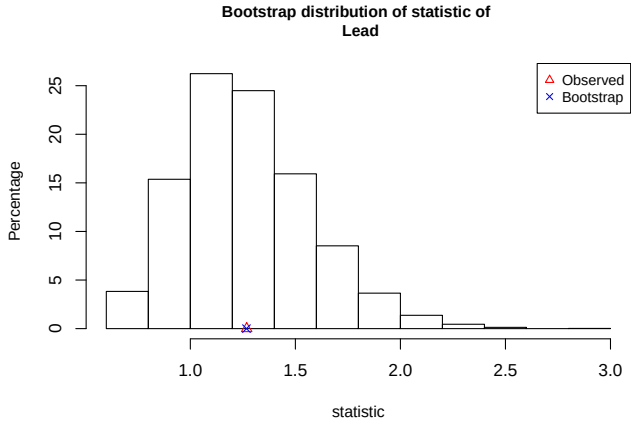
```
##
##  ** Bootstrap interval for statistic
##
##  Observed  Lead : 1.26825
##  Mean of bootstrap distribution: 1.26712
```

```
##  Standard error of bootstrap distribution: 0.30701
##
##  Bootstrap percentile interval
##      2.5%     97.5%
## 0.7737042 1.9683106
##
##      *--------------*
```



The sample mean is 1.267 and the SD is 0.306

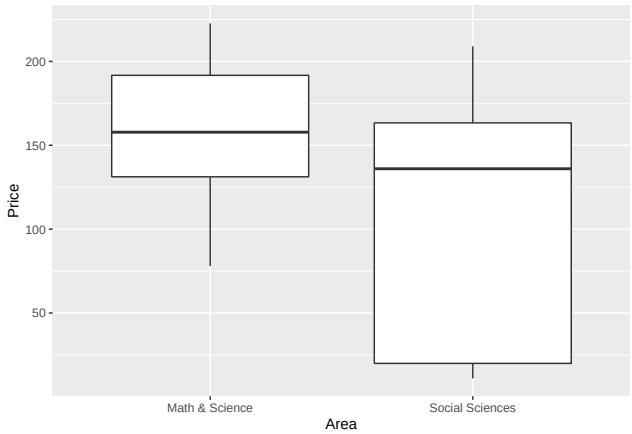
```
##
##  ** Bootstrap interval for statistic
##
##  Observed  Lead : 1.26825
##  Mean of bootstrap distribution: 1.26722
##  Standard error of bootstrap distribution: 0.30522
##
##  Bootstrap percentile interval
##      2.5%     97.5%
## 0.7726779 1.9531067
##
##      *--------------*
```



The shape is a right-skewed graph, the center is around 1.35, the spread large.

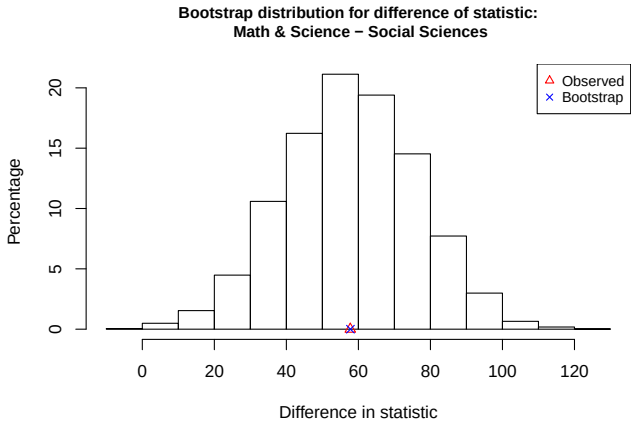
The 95% confidence interval would be between 0.643 and 1.891

2



There does appear to be a difference in the price between the two areas.

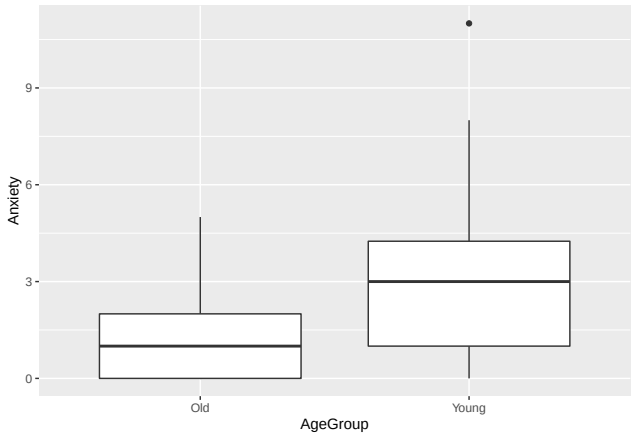
```
##
##  ** Bootstrap interval for difference of statistic
##
##  Observed difference of statistic:  Math & Science - Social Sciences =  57.74407
##  Mean of bootstrap distribution: 57.78938
##  Standard error of bootstrap distribution: 18.42124
##
##  Bootstrap percentile interval
##     2.5%    97.5%
## 21.47265 93.37643
##
##      *--------------*
```



The 95% confidence intervals would be between 21.42 and 94.52. What this means is that we are 95 sure that the average cost difference between books from the two fields fall between the values of 21.42 and 94.52.

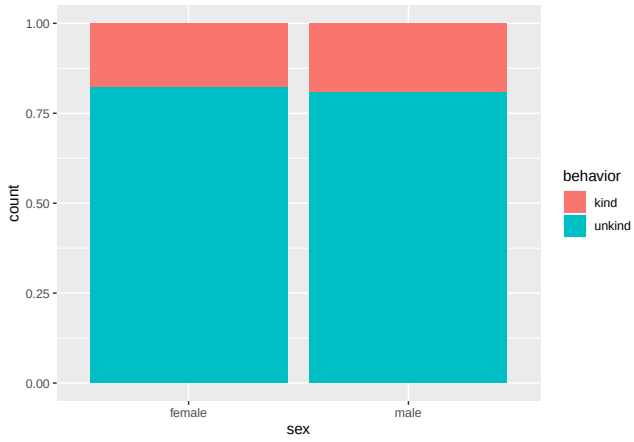
7

```
## # A tibble: 2 x 3
##   AgeGroup `mean(Anxiety)` `sd(Anxiety)`
##   <chr>              <dbl>         <dbl>
## 1 Old                 1.49          1.43
## 2 Young               2.92          2.47
```



A 95% confidence interval for the old would fall between 0.054 and 2.922 while the 95% confidence interval for the young would between 0.452 and 5.382. What this means is that we are 95% sure that the average anxiety level for the old fall between 0.054 and 2.922, and that the average anxiety level for the young fall between 0.452 and 5.382.

10

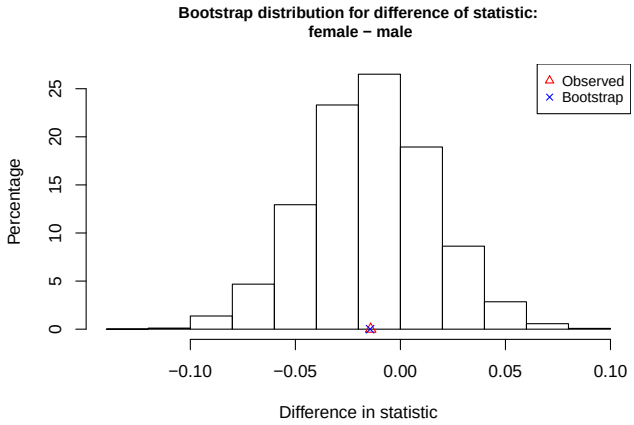


They do appear to be roughly the same

```
##
##                kind     unkind
##   female 0.08620690 0.39798851
##   male   0.09913793 0.41666667
```

The proportions are 0.086 for females while 0.099 for males.

```
##
##  ** Bootstrap interval for difference of statistic
##
##  Observed difference of statistic:  female - male =  -0.01416
##  Mean of bootstrap distribution: -0.01441
##  Standard error of bootstrap distribution: 0.02953
##
##  Bootstrap percentile interval
##        2.5%       97.5%
## -0.07168776  0.04409710
##
##      *--------------*
```



The 95% interval exists between -0.07241513 and 0.04391526. What this means is that we are 95% sure that the average difference in proportions of females and males who have seen good bahvior would fall between -0.07241513 and 0.04391526.