Lab 5b

Jack Wright

10/4/2020

Load Packages

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(openintro)
```

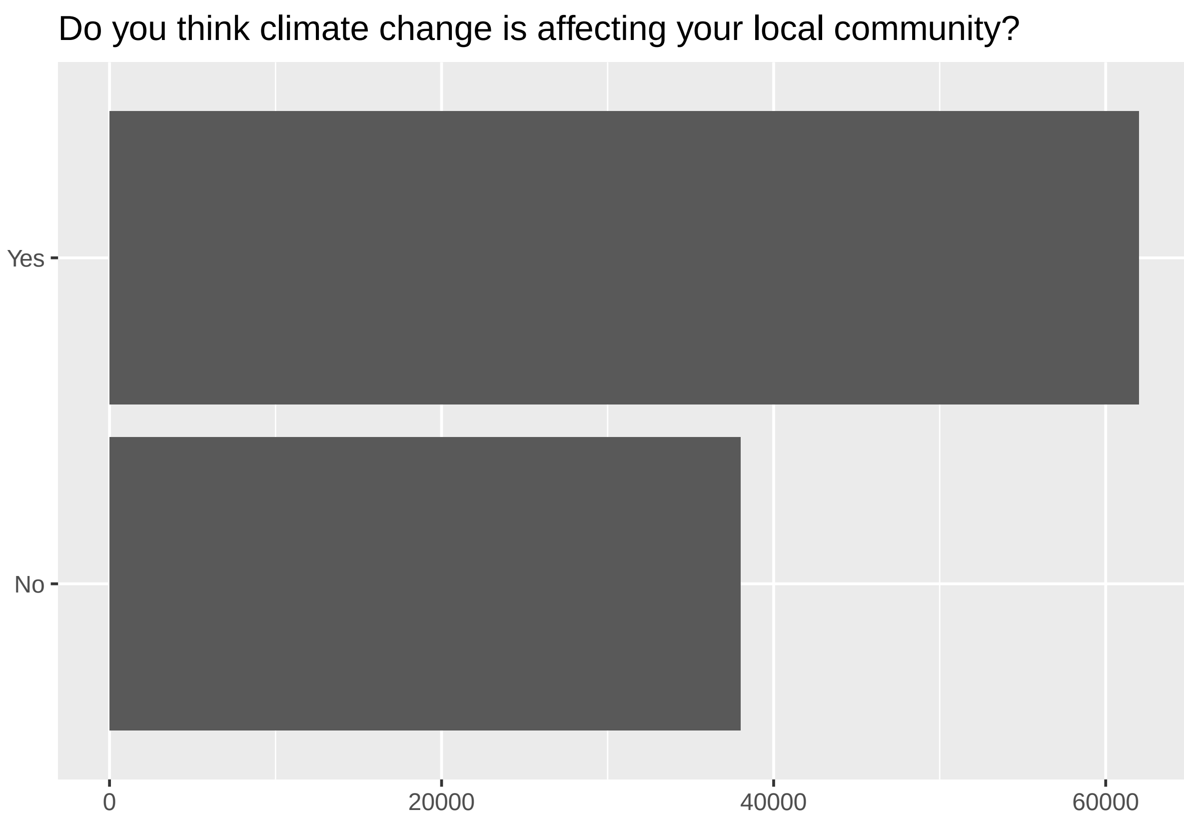
```
## Loading required package: airports

## Loading required package: cherryblossom

## Loading required package: usdata
```

```r
library(infer)
library(ggplot2)
```

The Data

```r
us_adults<-tibble(
  climate_change_affects=c(rep("Yes",62000),rep("No",38000))
)
```

```r
n <- 60
samp <- us_adults %>%
  sample_n(size = n)
```




```
## # A tibble: 2 x 3
##   climate_change_affects     n     p
##   <chr>                  <int> <dbl>
## 1 No                     38000  0.38
## 2 Yes                    62000  0.62
```

1. What percent of the adults in your sample think climate change affects their local community? Hint: Just like we did with the population, we can calculate the proportion of those in this sample who think climate change affects their local community.

I think around .62 percent think climate change is a problem.

2. Would you expect another student's sample proportion to be identical to yours? Would you expect it to be similar? Why or why not?

I think it is highly likely they would be "close" because of the Central Limit Theorem.

```
samp%>%
  specify(response = climate_change_affects, success="Yes")%>%
  generate(reps=1000, type="bootstrap")%>%
  calculate(stat="prop")%>%
  get_ci(level=.95)
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1    0.533    0.783
```

Confidence levels

3. In the interpretation above, we used the phrase "95% confident". What does "95% confidence" mean?

.95 percent of the time, the range will capture the point estimate of the population.

4. Does your confidence interval capture the true population proportion of US adults who think climate change affects their local community? If you are working on this lab in a classroom, does your neighbor's interval capture this value?

Yes it does capture the true population.

5. Each student should have gotten a slightly different confidence interval. What proportion of those intervals would you expect to capture the true population mean? Why?

I would expect .95 of the samples would capture the population mean by definition of the confidence interval.

.6 Given a sample size of 60, 1000 bootstrap samples for each interval, and 50 confidence intervals constructed (the default values for the above app), what proportion of your confidence intervals include the true population proportion? Is this proportion exactly equal to the confidence level? If not, explain why. Make sure to include your plot in your answer.

```
ci_frame<-NULL
n <- 60

for(i in 1:50){

samp <- us_adults %>%
  sample_n(size = n)

test<-samp %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.95)

ci_frame<-bind_rows(ci_frame,test)
}

capture<-ci_frame%>%
  filter(lower_ci<.62,upper_ci>.62)


p<-nrow(capture)/50

cat("The proportion of samples that include the true population proportion is ",p*100,"%.")
```

```
## The proportion of samples that include the true population proportion is  94 %.
```

This seems about right because a .96 is about as close as you can get to .95 with only 50 observations.

More Practice

7. the HIGHER the confidence level the WIDER the range it will capture. The LOWER the confidence level, the THINNER the range it will capture (because you have to be more sure you got the population proportion.)

8.

```
test<-samp %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.1)

test
```

```
## # A tibble: 1 x 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1      0.6    0.617
```

9. This is a good example because the range is 2%, and it does not capture the population proportion.

```
ci_frame<-NULL
n <- 60

for(i in 1:50){

samp <- us_adults %>%
  sample_n(size = n)

test<-samp %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = .1)

ci_frame<-bind_rows(ci_frame,test)
}

capture<-ci_frame%>%
  filter(lower_ci<.62,upper_ci>.62)


p<-nrow(capture)/50

cat("The proportion of samples that include the true population proportion is ",p*100,"%.")
```

```
## The proportion of samples that include the true population proportion is  8 %.
```

With this very narrow band, I captured the mean 10% of the time. It is a limited range which might give me a better idea about exactly what the population proportion is, but it only works .1 of the time.

10. I expect the width to be larger than the .95 interval

```r
ci_frame<-NULL
n <- 60

for(i in 1:50){

samp <- us_adults %>%
  sample_n(size = n)

test<-samp %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = .99)

ci_frame<-bind_rows(ci_frame,test)
}

capture<-ci_frame%>%
  filter(lower_ci<.62,upper_ci>.62)


p<-nrow(capture)/50

cat("The proportion of samples that include the true population proportion is ",p*100,"%.")
```

```
## The proportion of samples that include the true population proportion is  96 %.
```

This captures a about what I would expect. Around 99 % of the time.

11. As the sample size increases the width of the interval decreases.
12. As the bootstrap samples increases, the width of the interval decreases.