Lab 5 (Part B)

Code

Julian Adames-Ng

2022-03-09

Hide

```
library(tidyverse)
library(openintro)
library(infer)
```

Exercise 1

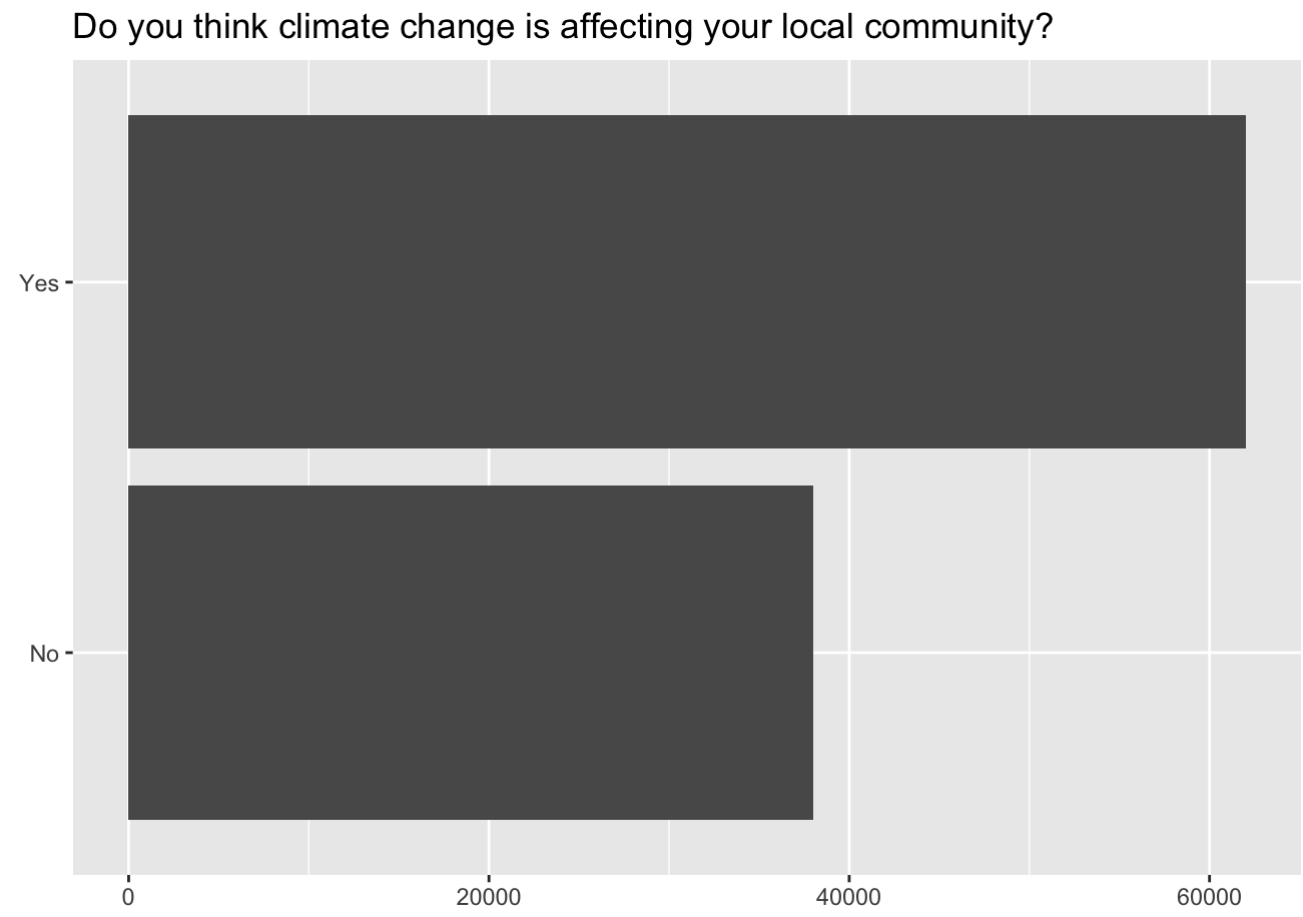
What percent of the adults in your sample think climate change affects their local community? Hint: Just like we did with the population, we can calculate the proportion of those in this sample who think climate change affects their local community.

-62% of the adults in the sample think climate change affects their local community.

Hide

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

ggplot(us_adults, aes(x = climate_change_affects)) +
  geom_bar() +
  labs(
    x = "", y = "",
    title = "Do you think climate change is affecting your local community?"
  ) +
  coord_flip()
```



Hide

```
us_adults %>%
  count(climate_change_affects) %>%
  mutate(p = n /sum(n))
```

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

Hide

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

Exercise 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 would not expect another student's sample proportion to be identical, but I would expect it to have a similar outcome.

...

Exercise 3

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

-95% confidence means that we are 95% sure that a given interval actually contains the population parameter.

...

Hide

```
set.seed(95)
samp %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.95)
```

```
## # A tibble: 1 × 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1     0.45      0.7
```

Exercise 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, the true value of the population proportion is 62% and falls between the 51.667% and 75% bounds of my confidence interval. ...

Exercise 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'd say that 95% of the intervals would capture the true population mean because the confidence level is 95%.

...

Exercise 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.

-45/50 of the confidence intervals include the population proportion. This does not equal to the 95% confidence interval as this is 90%.

...

Exercise 7

Choose a different confidence level than 95%. Would you expect a confidence interval at this level to be wider or narrower than the confidence interval you calculated at the 95% confidence level? Explain your reasoning.

-If I were to choose a 90% confidence interval, I'd expect this level to be narrower than the 95% confidence interval because we are less confident in our value falling within the bounds. The true value has a better chance of being in a larger interval of values vs with a narrower interval, the slimmer the chance.

...

Exercise 8

Using code from the infer package and data from the one sample you have (samp), find a confidence interval for the proportion of US Adults who think climate change is affecting their local community with a confidence level of your choosing (other than 95%) and interpret it.

-Here, the confidence level chosen was 77%. The confidence interval corresponding to this is smaller than the 95% interval that we previously looked at.

...

Hide

```
set.seed(77)
samp %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.77)
```

```
## # A tibble: 1 × 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1    0.483     0.65
```

Exercise 9

Using the app, calculate 50 confidence intervals at the confidence level you chose in the previous question, and plot all intervals on one plot, and calculate the proportion of intervals that include the true population proportion. How does this percentage compare to the confidence level selected for the intervals?

-The confidence level chosen was 77%. 39 out of 50 of the intervals included the true population. This equates to 78% which is 1% shy of the 77% confidence level that was used.

...

Exercise 10

Lastly, try one more (different) confidence level. First, state how you expect the width of this interval to compare to previous ones you calculated. Then, calculate the bounds of the interval using the infer package and data from samp and interpret it. Finally, use the app to generate many intervals and calculate the proportion of intervals that capture the true population proportion.

-Here, I chose a 25% confidence level. I expect only about 25% of the intervals to include the true proportion. The bounds of my 25% confidence interval are 61.667% and 65%, which is pretty narrow. Using the app, 13 out of 50 of the intervals or 26% of the intervals included the true proportion which is also 1% shy of the 25% confidence level.

...

Hide

```
set.seed(25)
samp %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.25)
```

```
## # A tibble: 1 × 2
##   lower_ci upper_ci
##      <dbl>    <dbl>
## 1     0.55    0.590
```

Exercise 11

Using the app, experiment with different sample sizes and comment on how the widths of intervals change as sample size changes (increases and decreases).

-As the sample sizes increase, the widths of the intervals decrease.

...

Exercise 12

Finally, given a sample size (say, 60), how does the width of the interval change as you increase the number of bootstrap samples. Hint: Does changing the number of bootstrap samples affect the standard error?

-Increasing the number of bootstrap samples doesn’t change the width of the interval at all.

...