Assignment_5b

AZM

2022-10-03

Contents

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.1
## v readr   2.1.2     v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(openintro)
```

```
## Loading required package: airports
## Loading required package: cherryblossom
## Loading required package: usdata
```

```
library(dplyr)
library(infer)
library(metan)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
## |=========================================================|
## | Multi-Environment Trial Analysis (metan) v1.17.0        |
## | Author: Tiago Olivoto                                   |
## | Type 'citation('metan')' to know how to cite metan      |
## | Type 'vignette('metan_start')' for a short tutorial     |
## | Visit 'https://bit.ly/pkgmetan' for a complete tutorial |
## |=========================================================|
##
## Attaching package: 'metan'
##
## The following object is masked from 'package:forcats':
##
##     as_factor
##
## The following object is masked from 'package:dplyr':
##
##     recode_factor
```

```
##
## The following object is masked from 'package:tidyr':
##
##     replace_na
##
## The following objects are masked from 'package:tibble':
##
##     column_to_rownames, remove_rownames, rownames_to_column
```

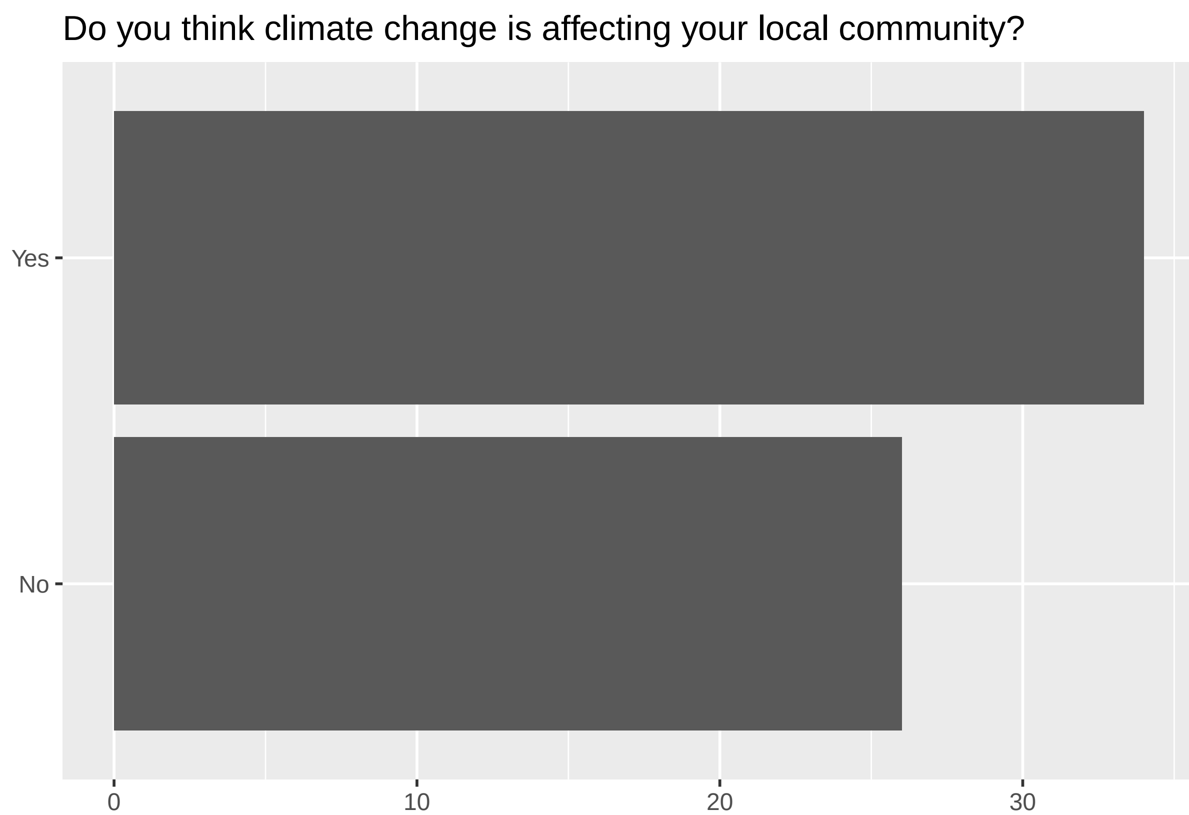
```
#Plot CI function source from https://rdrr.io/github/OpenIntroOrg/oilabs-r-package/src/R/plot_ci.R
# Instead of loading and implementing OpenIntro Labs companion library
plot_ci <- function(lo, hi, m) {
  par(mar=c(2, 1, 1, 1), mgp=c(2.7, 0.7, 0))
  k <- length(lo)
  ci.max <- max(rowSums(matrix(c(-1*lo,hi),ncol=2)))

  xR <- m + ci.max*c(-1, 1)
  yR <- c(0, 41*k/40)

  plot(xR, yR, type='n', xlab='', ylab='', axes=FALSE)
  abline(v=m, lty=2, col='#00000088')
  axis(1, at=m, paste("mu = ",round(m,4)), cex.axis=1.15)
  #axis(2)
  for(i in 1:k){
    x <- mean(c(hi[i],lo[i]))
    ci <- c(lo[i],hi[i])
    if((m < hi[i] & m > lo[i])==FALSE){
      col <- "#F05133"
      points(x, i, cex=1.4, col=col)
      #       points(x, i, pch=20, cex=1.2, col=col)
      lines(ci, rep(i, 2), col=col, lwd=5)
    } else{
      col <- 1
      points(x, i, pch=20, cex=1.2, col=col)
      lines(ci, rep(i, 2), col=col)
    }
  }
}
```

```
set.seed(0144)
us_adults <- tibble(
  climate_change_affects = c(rep("Yes", 62000), rep("No", 38000))
)
samp1 <- us_adults %>%
  sample_n(60)

ggplot(samp1, aes(x = climate_change_affects)) +
  geom_bar() +
  labs(
    x = "", y = "",
    title = "Do you think climate change is affecting your local community?"
  ) +
  coord_flip()
```




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

```
##                        [,1]    [,2]
## climate_change_affects "No"    "Yes"
## n                      "38000" "62000"
## p                      "0.38"  "0.62"
```

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.

In this sample, 61.67% of adults believe that climate change affects their community.

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 expect another student's sample size to be similar to mine, but not exactly the same. These are random samples, so there should be some degree of variation in the data set, but it should be remarkably close.

Exercise 3

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

We used the phrase 95% confident to indicate that we are 95% sure that the value falls within a range.

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?

My confidence interval has a 95% chance to contain the true population portion of UUS adults who think climate change affects their local community. And if I was working in a lab, my neighbors interval would have the same 95% chance to capture it. (assuming they followed the same code as above)

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 would expect 95% of the ci intervals to capture the true population mean. Because all the intervals have a 95% ci, I would expect 95% of all of them to capture the true population mean.

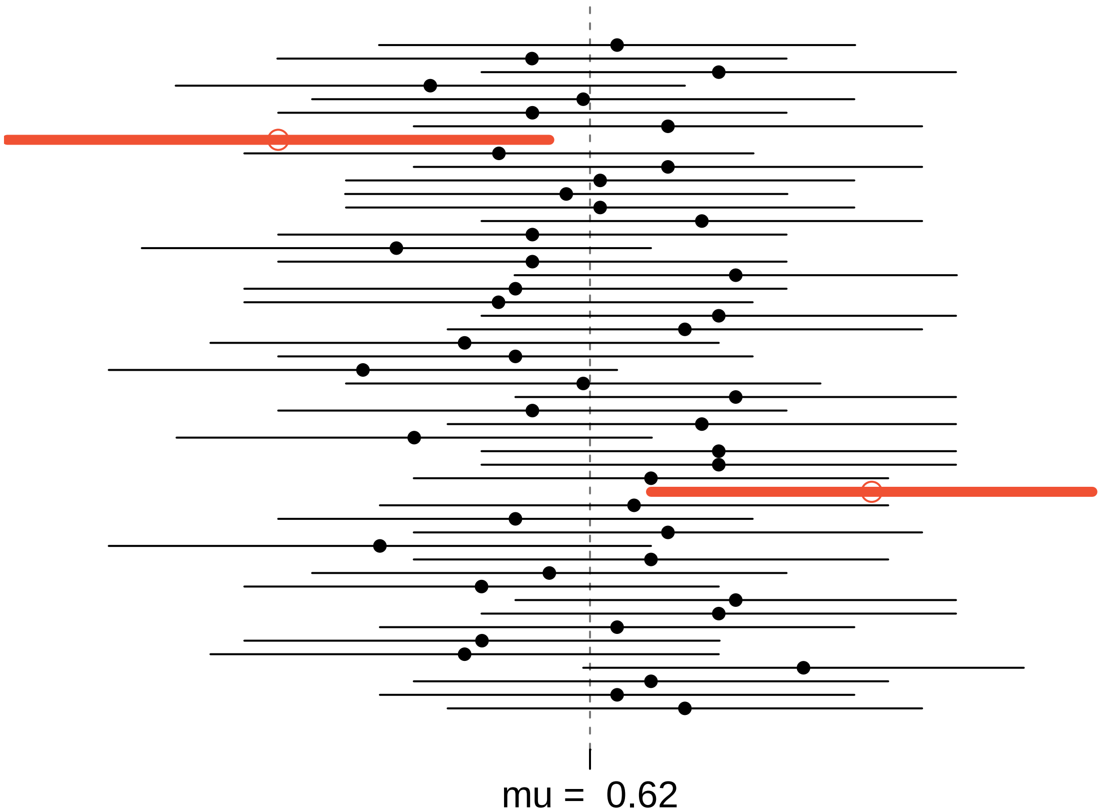
Exercise 6

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.

```
n <- 50
high <- rep(NA, n)
low <- rep(NA, n)
for(i in 1:n){
  test <- sample_n(us_adults, 60)
  ci <- test %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.95)
  high[i] <- ci$upper_ci
  low[i] <- ci$lower_ci


}
```

```
plot_ci(low, high, 0.62)
```



96% of my sample (48/50) contains the true population proportion. This proportion is approximately equal to the Ci, within a margin of error.

Exercise 7

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

I choose a 1% ci! This would have a significantly smaller ci than the 95% as there only needs to be a 1% chance that the value is within the range.

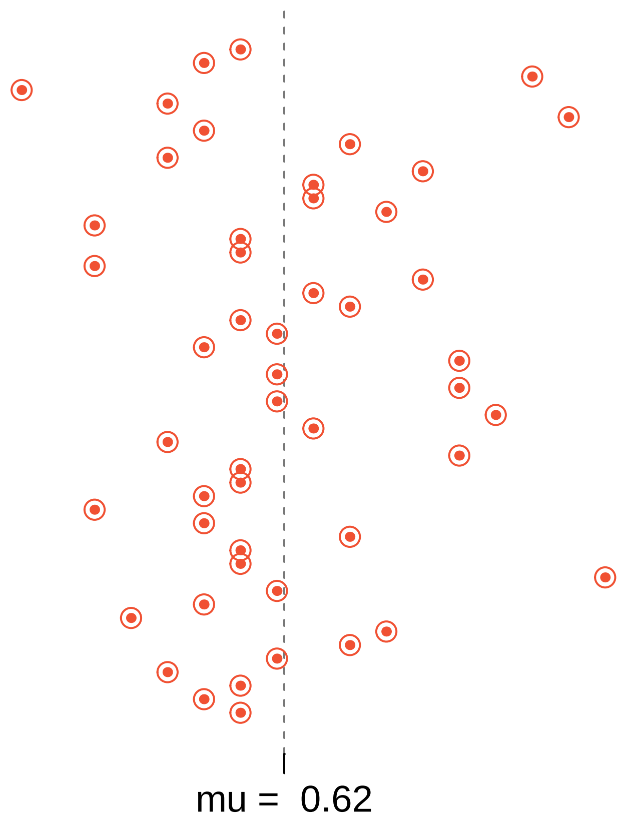
Exercise 8 & 9

Using code from the infer package and data fromt 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.

```
n <- 50
high <- rep(NA, n)
low <- rep(NA, n)
for(i in 1:n){
  test <- sample_n(us_adults, 60)
  ci <- test %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.01)
  high[i] <- ci$upper_ci
  low[i] <- ci$lower_ci
```

```
}
```

```
plot_ci(low, high, 0.62)
```



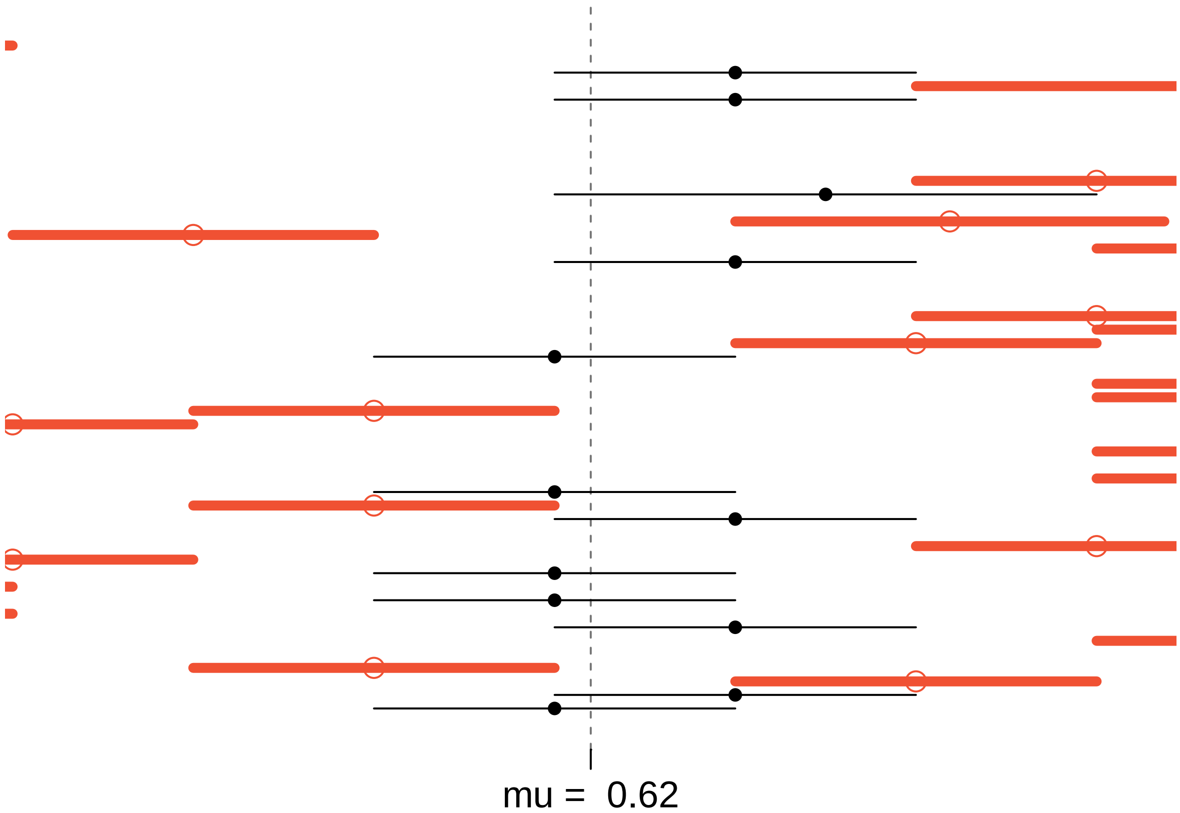
None of the cis include the true population proportion.

Exercise 10 & 11

Using code from the infer package and data fromt 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.

```
n <- 50
high <- rep(NA, n)
low <- rep(NA, n)
for(i in 1:n){
  test <- sample_n(us_adults, 60)
  ci <- test %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.25)
  high[i] <- ci$upper_ci
  low[i] <- ci$lower_ci
}
```

```r
plot_ci(low, high, 0.62)
```



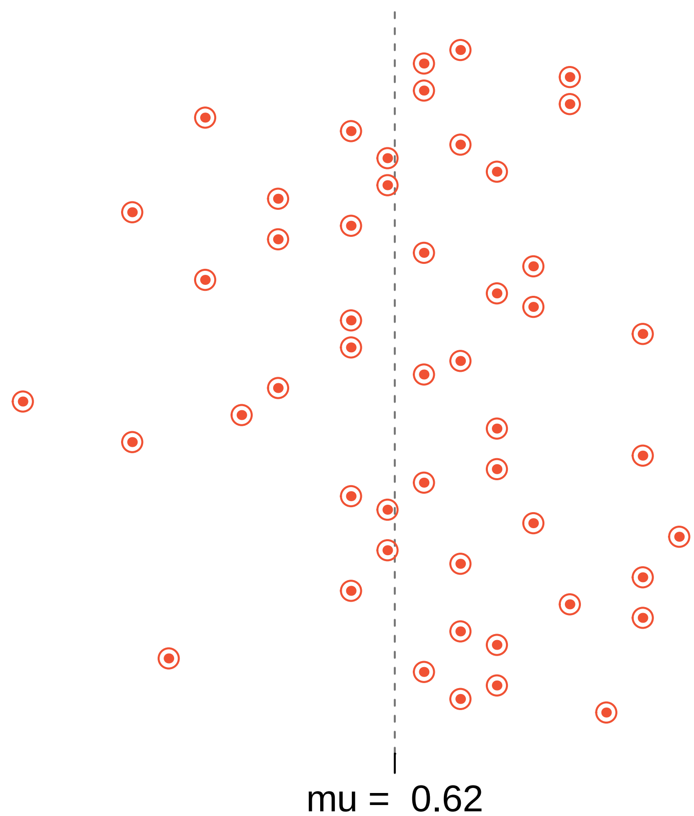
12/50 contain the true mean! The spread of the intervals is a lot higher, but less of them contain the true mean.

Exercise 11

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.

```r
n <- 50
high <- rep(NA, n)
low <- rep(NA, n)
for(i in 1:n){
  test <- sample_n(us_adults, 60)
  ci <- test %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 1000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.01)
  high[i] <- ci$upper_ci
  low[i] <- ci$lower_ci
}
```

```r
plot_ci(low, high, 0.62)
```



As the number of samples decrease, the variation of the intervals decreases

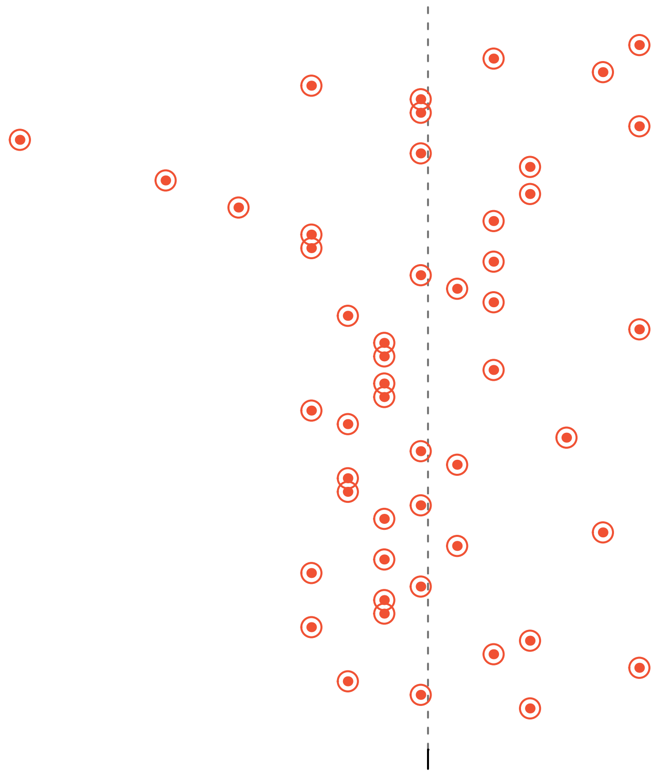
Exercise 12

Using code from the infer package and data fromt 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.

```
n <- 50
high <- rep(NA, n)
low <- rep(NA, n)
for(i in 1:n){
  test <- sample_n(us_adults, 60)
  ci <- test %>%
  specify(response = climate_change_affects, success = "Yes") %>%
  generate(reps = 2000, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  get_ci(level = 0.01)
  high[i] <- ci$upper_ci
  low[i] <- ci$lower_ci


}
```

```
plot_ci(low, high, 0.62)
```



Since the seed has been set the number of reps do not matter.