Math 300 Lesson 27 Notes

Case Study

YOUR NAME HERE

July, 2022

Contents

Objectives

1. Calculate, using both percentile and standard error methods, confidence intervals for a difference of two proportions.
2. Calculate a normal-based, theory-based, confidence interval for a single proportion or mean.
3. Correctly use all terms and notation to include standard error, margin of error, sampling distribution, and bootstrap distribution.
4. Compare and contrast sampling distribution and bootstrap distribution.

Reading

Chapter 8.6-8.7

Lesson

There are no learning checks for this lesson.

- The case study brings all the ideas together while introducing a two-sample test.
- The bootstrap sample needs to be the same as the original sample. What is not clear is if they sampled such that there will be 16 in the control group and 34 in the seed group. The `infer` package does not allow us to sample within groups so we sample with replacement from the entire sample.

Libraries

```
library(tidyverse)
library(moderndive)
library(infer)
```

Yawning Problem

The reading discusses the research question "Is yawning contagious?". Let's summarize the data in two different ways.

```
mythbusters_yawn %>%
  group_by(group, yawn) %>%
  summarize(count = n()) %>%
  mutate(prop = count / sum(count))
```

```
## `summarise()` has grouped output by 'group'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 4 x 4
## # Groups:   group [2]
##   group   yawn  count  prop
##   <chr>   <chr> <int> <dbl>
## 1 control no       12 0.75
## 2 control yes       4 0.25
## 3 seed    no       24 0.706
## 4 seed    yes      10 0.294
```

- What does this summary tell the reader?

Answer:

Now another summary is as follows:

```
# Complete the code
# mythbusters_yawn %>%
#   group_by(group, _____) %>%
#   summarize(count = ____) %>%
#   ungroup() %>%
#   mutate(prop = count / sum(count))
```

- What does this summary tell us about the data?

Answer:

- Which summary is better for the research question?

Answer:

Let's determine how the `infer` package samples the data.

```
# Complete the code
# set.seed(77)
# mythbusters_yawn %>%
#   specify(formula = ______ ~ _______,success="______") %>%
#   generate(reps = 1, type = "__________") %>%
#   group_by(group, yawn) %>%
#   summarize(count = n()) %>%
#   mutate(prop = count / sum(count))
```

The numbers in the control and seed group changed. The sample is with replacement from the entire data frame and not within groups.

Bootstrap Sampling Distribution

```
# Complete the code
# (bootstrap_distribution_yawning <- mythbusters_yawn %>%
#   specify(formula = yawn ~ group, success = "_______") %>%
#   generate(reps = 1000, type = "bootstrap") %>%
#   calculate(stat = "__________", order = c("_______", "________")))
```

Generate the bootstrap distribution

```
# Complete the code
#visualize(_________________)
```

Visualize the results.

- Where is the center of this distribution?

Answer:

```
# Complete the code
# bootstrap_distribution_yawning %>%
#   summarize(ave=mean(__________))
```

```
# Complete the code
# get_ci(bootstrap_distribution_yawning,type="_________")
```

Confidence Interval

- Interpret this confidence interval.

Answer:

Theory-based confidence interval

For the data object `bowl_sample_1` find a 95% confidence interval using the bootstrap percentile, bootstrap standard error method, and the theory-based method.

```
head(bowl_sample_1)
```

```
## # A tibble: 6 x 1
##   color
##   <chr>
## 1 white
## 2 white
## 3 red
## 4 red
## 5 white
## 6 white
```

- Find the bootstrap percentile confidence interval for this problem.

Answer:

```
# Complete the code
# bowl_sample_1 %>%
#   summarize(prop=__________(color=="_______"))
```

```
# Complete the code
# set.seed(45678)
# bootstrap_dist<-bowl_sample_1 %>%
#   specify(color~________,success="________") %>%
#   generate(reps=1000,type="bootstrap") %>%
#   calculate(stat="_________")
```

```
# get_ci(bootstrap_dist,type="_________")
```

- Find the bootstrap standard error confidence interval for this problem.

Answer:

```
# get_ci(bootstrap_dist,type="________",point_estimate = ______)
```

- Find the theory-based confidence interval for this problem.

Answer:

```
# Complete the code
# bowl_sample_1 %>%
#   summarize(prop=mean(color=="_______"),total=n()) %>%
#   mutate(lower_ci=prop-qnorm(________)*sqrt(prop*(1-prop)/total),
#          upper_ci=prop+qnorm(_______)*sqrt(_____________))
```

Documenting software

- File creation date: 2022-07-06
- R version 4.1.3 (2022-03-10)
- `tidyverse` package version: 1.3.1
- `moderndive` package version: 0.5.4
- `infer` package version: 1.0.2