Math 300 Lesson 25 Notes

Constructing Confidence Intervals

YOUR NAME HERE

June, 2022

Contents

Objectives

1. construct bootstrap percentile and standard error confidence intervals for a single mean or median using the `infer` package.

Reading

Chapter 8.4

Lesson

Work through the learning check LC 8.5.

- We will be using the `infer` package to create the bootstrap confidence intervals. Last lesson we used `rep_sample_n()`. The `infer` package will be used for inference in the rest of the course.
- The `infer` package gives a framework to think about and conduct inference. It makes hypothesis testing and confidence interval construction more structured and puts computational resources at the center versus mathematical tools.
- the `infer` package uses the verbs `specify()`, `generate()`, `calculate()`, and `visualize()` to complete the construction process.

Libraries

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

Review

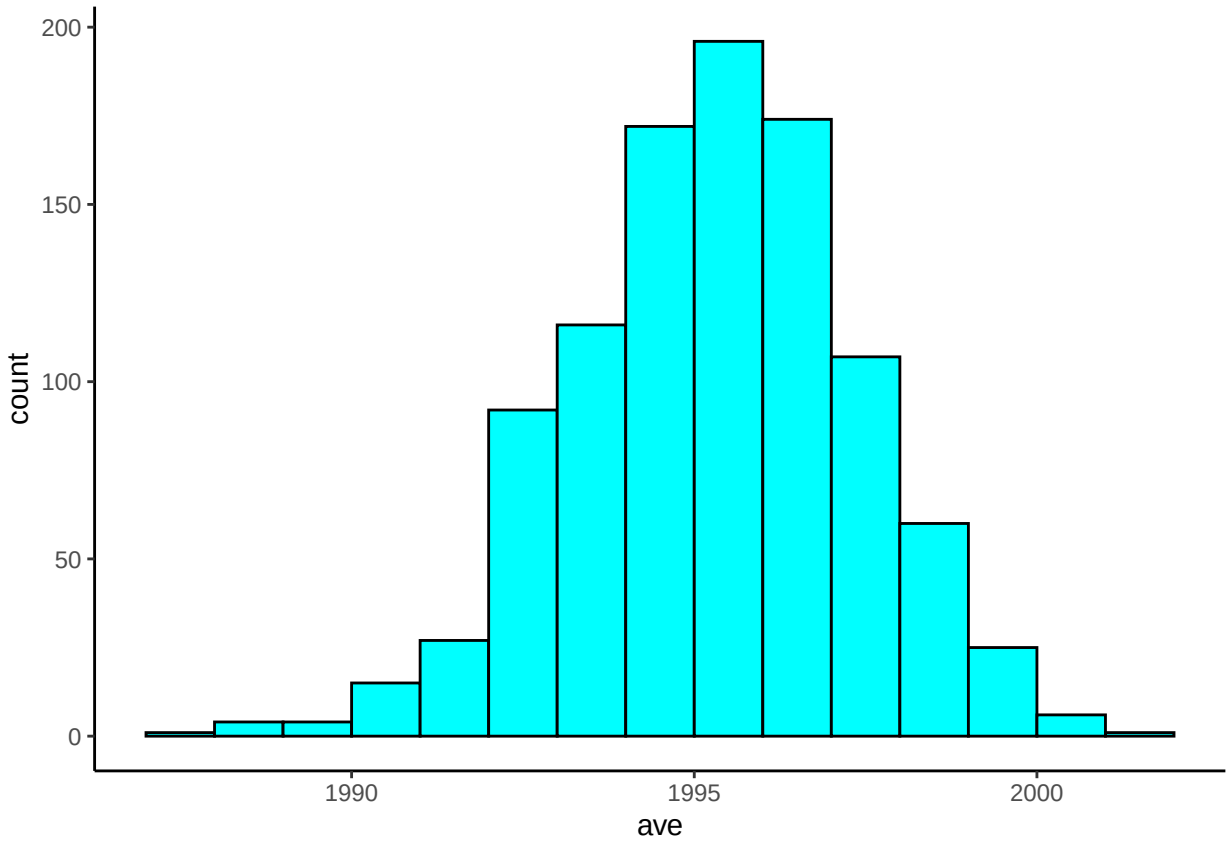
As a review, let's construct the bootstrap distribution of the sample mean for our pennies data. This requires use to use our skills on wrangling data frames.

```
head(pennies_sample)
```

```
## # A tibble: 6 x 2
##      ID  year
##   <int> <dbl>
## 1     1  2002
## 2     2  1986
## 3     3  2017
## 4     4  1988
## 5     5  2008
## 6     6  1983
```

```
set.seed(52249)
bootstrap_dist<-pennies_sample %>%
  rep_sample_n(size = 50, replace = TRUE, reps = 1000) %>%
  group_by(replicate) %>%
  summarize(ave=mean(year))
```

```
bootstrap_dist %>%
  ggplot(aes(x=ave)) +
  geom_histogram(binwidth = 1, color = "black",
                 boundary = 1990, fill = "cyan") +
  theme_classic()
```




```
bootstrap_dist %>%
  summarize(center=mean(ave),
            lower=quantile(ave,probs=0.25),
            upper=quantile(ave,probs=.975))
```

```
## # A tibble: 1 x 3
##   center lower upper
##    <dbl> <dbl> <dbl>
## 1  1995. 1994. 1999.
```

The standard error method.

```
bootstrap_dist %>%
  summarize(center=mean(ave),
            lower=center-sd(ave)*qnorm(.975),
            upper=center+sd(ave)*qnorm(.975))
```

```
## # A tibble: 1 x 3
##   center lower upper
##    <dbl> <dbl> <dbl>
## 1  1995. 1991. 1999.
```

Steps from the `infer` package

- First specify the response variable, and explanatory variables if present. We like to use the `formula` option. Why do we pipe to the `heaD()` function?

```
# Complete the code and remove the comment symbols
# pennies_sample %>%
#   specify(formula=______~NULL) %>%
#   head()
```

It is similar to using `select()` but note the meta data has also changed.

- Generate replicates

```
# Complete the code and remove the comment symbols
# pennies_sample %>%
#   specify(formula=______~NULL) %>%
#   generate(reps = _________, type = "__________") %>%
#   head()
```

Compare this code with that using `rep_sample_n()`.

- Find the sample statistic for each replicate.

```
# Complete the code and remove the comment symbols
# pennies_sample %>%
#   specify(formula=_______~NULL) %>%
#   generate(reps = ________, type = "__________") %>%
#   calculate(stat = "________") %>%
#   head()
```

- Using results of bootstrap distribution

```
# Complete the code and remove the comment symbols
# boot_dist_mean <- pennies_sample %>%
#   specify(formula=_______~NULL) %>%
#   generate(reps = ________, type = "___________") %>%
#   calculate(stat = "_________")
```

```
# Complete the code and remove the comment symbols
# (percentile_ci <- boot_dist_mean %>%
#   get_confidence_interval(level = 0.95, type = "___________"))
```

```
# Visualize the results
# visualize(boot_dist_mean) +
#   shade_confidence_interval(endpoints = ___________________)
```

Or if we want the standard error method.

```
# Remove the comment symbols
# (mean_pennies <- pennies_sample %>%
#   summarize(ave=mean(year)) %>%
#   pull())
```

```
# Complete the code and remove the comment symbols
# (standard_error_ci <- boot_dist_mean %>%
#   get_confidence_interval(type = "________", point_estimate = mean_pennies))
```

LC 8.5 (Objective 1)

(LC 8.5) Construct a 95% confidence interval for the *median* year of minting of *all* US pennies? Use the percentile method and, if appropriate, then use the standard-error method.

Solution:

Documenting software

- File creation date: 2022-06-12
- R version 4.1.3 (2022-03-10)
- `tidyverse` package version: 1.3.1
- `moderndive` package version: 0.5.4
- `infer` package version: 1.0.0