

Lab 3.1 - Bootstrapping

Student

2025-04-16



Creating a bootstrap simulation

For this assignment, we will again be using data from the Seoul bicycle sharing system. In particular, we want to try and create accurate confidence intervals for various features of the dataset.

You can find the variable definitions [here](#).

Expectations

1. Before starting, write a sentence or two on your expectation of the confidence interval for two variables, `Rental.Count` (number of bicycles rented that day) and `Temperature.C.`.

Setup

2. Install the library `infer` (documentation [here](#)) and load it
3. Review the basic code process here: [here](#)

Create a sample

4. Make a subset of the data that is a random sample of size 50 using the following command:

```
seoul.bicycles.sample <- seoul.bicycles %>%
  slice_sample(n = 50, replace = TRUE)
```

Classical confidence intervals

5. Check the conditions and create the 95% confidence intervals by hand for your chosen variable
6. Compare the result to the overall dataset values - did your confidence interval cover the true value? How close was your sample to reality?

Bootstrapping

7. Using the `rep_sample_n(size = 50, replace = TRUE, reps=100000)` function, sampling with replacement. Make sure to follow the code example on the [moderndive](#) [web-page](#) to generate your sampling distribution of means.

To run create replications, you can do so by modifying the following sample code:

Note: **VERY IMPORTANT** - use your sample of size 50, and then sample from THAT sample.

```
virtual_resamples <- seoul.bicycles.sample %>%
  rep_sample_n(size = 50, replace = TRUE, reps = 10000)
```

To generate a list of sample means, in particular, you'll need to change the variable in the `mean()` function to be the one you chose:

```
virtual_resampled_means <- virtual_resamples %>%
  group_by(replicate) %>%
  summarize(mean_year = mean(<variable you selected>))
```

Remember that when bootstrapping, you should make sure that the `size` argument is set to be the size of your sample - you want to use all of the information possible from your sample!

8. Create a confidence interval using the `quantile()` function on this sampling distribution. How do these compare to your confidence intervals calculated in the classical way?

In particular, you want to check cutoff at the 0.025 and 0.975 range of your data (the range in which 95% of the sample means fell).

Hint: you can find documentation on using the `quantile()` function by typing `?quantile` in the Console window

9. Compare the result of this confidence interval you generated by bootstrapping to the ones calculated by classical methods. How close were they? Do the differences surprise you or not?
10. Think carefully about what the difference is between a confidence interval calculated by classical methods and the one generated by bootstrapping. What are the differences in key assumptions?
 - If you have extra time, redo this analysis but for the other variable
 - If you still have extra time, you can try to use the alternative workflow described [here](#).