

Imt573 Lab 6: Compare means

November 4, 2020

Instructions

This lab asks you to compute confidence intervals (CI) for difference of two means. In particular, we check if the AirBnB price differs for 0 and 1BR listings (data is a small subsample from 2019 AirBnB Beijing listings). First, compute the difference in mean price. Second, simulate a large number of prices from a common distribution, and compare the difference of the simulated prices. Do you find that your simulation will frequently give you comparable price difference?

We suggest you to consult lecture notes <https://otoomet.bitbucket.io/machineLearning.pdf/> section 1.4.3 (p 30 for now).

1 Simulations

First, let's simulate the polls.

1. Load the data *airbnb-beijing-br01.csv*.

It contains two variables: *price* and *bedrooms*. The data is curated so you can use it right away, just do a sanity check first.

2. Compare the mean price for 0-BR and 1-BR listings. Which one is larger?
3. How many 0-BR and how many 1-BR listings are there? Call these numbers N_0 and N_1 .
4. Compute the overall mean and standard deviation of price

5. Now create N_0 random normal “fake 0-BR listings” and N_1 random normal “fake 1-BR listings”.

Use the mean and standard deviation you computed above! Do not use standard normals! You can do this with something like `rnorm(N0, mean=mu, sd=sd)` where `mu` and `sd` are the mean and standard deviation you computed in 4.

6. Compute the difference b/w fake 1-BR and fake 0-BR listings. How does it compare to the actual difference?
7. Now repeat the points 5 and 6 1000 (or more) times. Each time compute the difference and store it.
8. Plot histogram of the differences

I recommend to use `hist()`

9. Compute and print 0.025 and 0.975 quantiles. This defines your confidence interval.

10. Does the actual difference fall into this CI?