**Is the price of hotel more expensive than other listings in New York City?**

*Hypothesis Testing using NYC Airbnb data*

By Muluemebet Ayalew

March,2020

Statistical test is performed to see if hotel listing price is different from non-hotel listing price in New York city from Airbnb data. Both two-samples t-test and bootstrapped tests were performed to test the hypothesis defined as:

The null hypothesis: The mean price of hotel and non-hotel listings are the same in New York city.

Alternative hypothesis: The mean price of hotel is greater than the mean price of non-hotel listing in New York city(one-tailed)

Before performing the test, it is good to see how the distribution of the two samples look (see Figure 1).

|  |  |  |
| --- | --- | --- |
|  | **Mean** | **Standard Deviation** |
| Hotel listing | 554.42 | 1714.05 |
| Non-hotel listing | 153.94 | 298.37 |

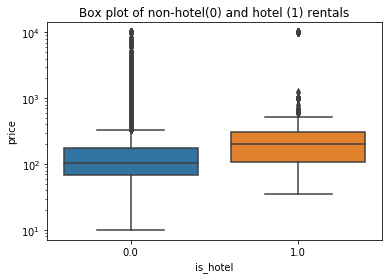


Figure 1 hotel and non-hotel listing price distribution

Once the hypothesis was defined, the test statistics was chosen (t-test statistics with the same variance) and significance level was set (0.05). The assumption of same variance had to be tested before using the test statistics. Therefore, another test statistic for difference in standard deviation was tested for the two groups (hotel and non-hotel listings). Bootstrapped test was performed to test the hypothesis:

Null hypothesis: The standard deviations of listing price for hotel and non-hotel are the same.

Alternative Hypothesis: The standard deviation of listing price for hotel and non-hotel is the same (two tailed).

The p-value was 0.4736 which is greater than the significance level (0.05). The p-value is the probability to get an outcome at least as extreme as what was observed. Therefore, the difference of standard deviation of price for hotel and non-hotel listing is statistically not significant. We can accept the null hypothesis and assume that the two groups have the same variance.

After testing equal variance between the two groups, the test statistics and p-value was computed to test the mean difference (the first hypothesis). The p-value was very low for both t-test (p-value close to zero) and bootstrapped tests(p=0.0001). Concluding that the mean price of hotel listing is statistically different from the mean price of other listing. As an evidence, the bootstrapped replicates of mean difference between hotel and other listings, and the observed difference was plot (see Figure 2). The observed difference is far from other data. And thus, the p- value is very small.

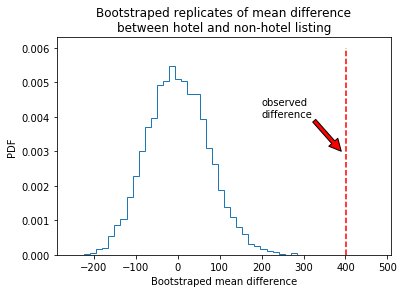


Figure 2 Bootstrapped replicates of mean difference