Project 1 Inferential Statistics
Anna Apa

After studying trends between the Airbnb listings, we apply inferential statistics techniques to see if any of these relationships are statistically significant. We mainly focus on how these features are related to the number of cancellations of the listings.

For most of the variables, we make use of permutation tests to see if their respective correlations with the number of cancellations are statistically significant.

Cancellations vs. Price

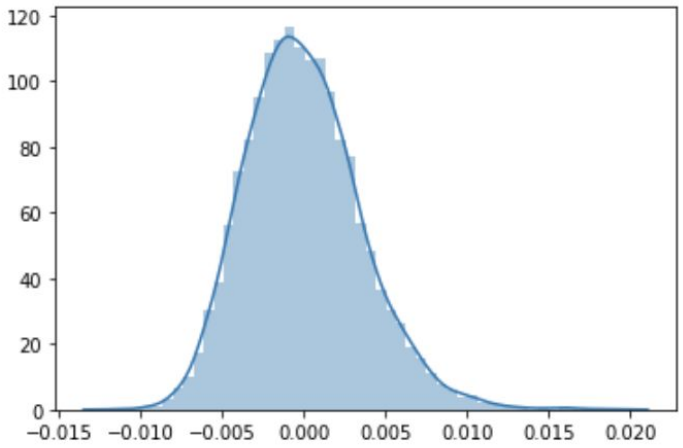
Null Hypothesis: There is no significant correlation between the number of cancellations and the price of the listing.

Alternative Hypothesis: There is a significant correlation between the two.

First, we calculate the actual correlation between the number of cancellations and listing prices. The observed correlation is -0.02526.

We test to see if this observed correlation is statistically significant by performing a permutation test. We use np.random.permutation to reorder the listing prices and get their correlation with the number of cancellations each time.

In this case, our p-value is the ratio of the amount of times the absolute value of our correlation replicates was greater or equal to our observed correlation.




```
observed correlation between price and number of cancellations: -0.02526
p-value: 0.0000000000
We reject the null. There is a significant correlation between price and
number of cancellations
```

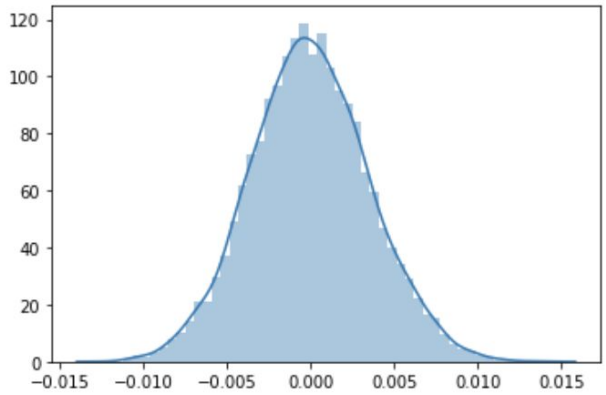
The p-value we get is really small. We then conclude that the correlation between price and the number of cancellations is statistically significant at the 1% level.

Cancellations vs Popularity (Number of Reviews)

Null Hypothesis: There is no significant correlation between the number of cancellations and the number of reviews a listing gets.

Alternative Hypothesis: There is a significant correlation between the two.

We perform the same test and conclude that their observed correlation of 0.21254 is statistically significant at the 1% level.



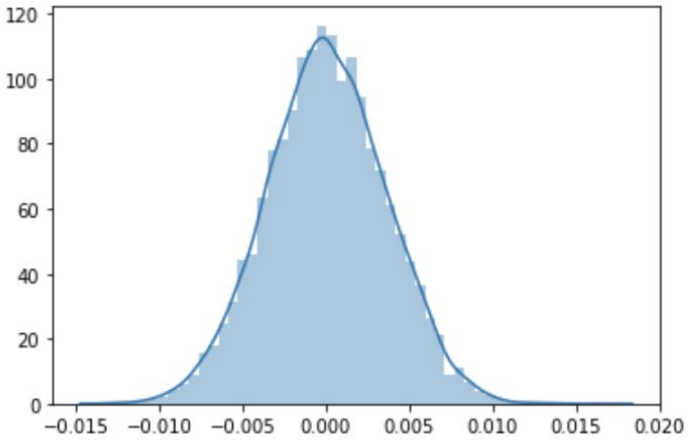

```
observed correlation between number of reviews and number of cancellation
s: 0.21254
p-value: 0.0000000000
We reject the null. There is a significant correlation between between th
e number of cancellations and number of reviews
```

Cancellations vs Demand (Days Booked)

Null Hypothesis: There is no significant correlation between the number of cancellations and the number of days the listing is booked.

Alternative Hypothesis: There is a significant correlation between the two.

We perform the same test and conclude that their observed correlation of -0.01100 is statistically significant at the 1% level.



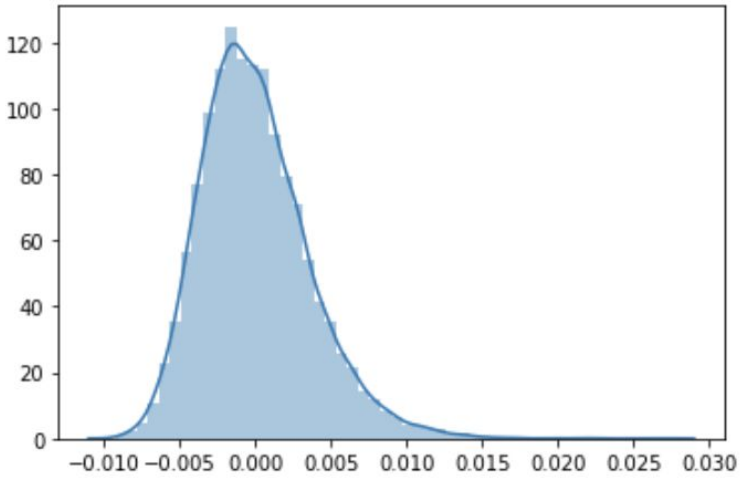

```
observed correlation between number of days booked and number of cancella
tions: -0.01100
p-value: 0.0018000000
We reject the null. There is a significant correlation between the number
of cancellations and the number of days the listing is booked.
```

Cancellations vs Minimum Nights

Null Hypothesis: There is no significant correlation between the number of cancellations and the minimum nights required by a listing.

Alternative Hypothesis: There is a significant correlation between the two.

We perform the same test and conclude that their observed correlation of -0.00473 is **not** statistically significant at the 1% level.



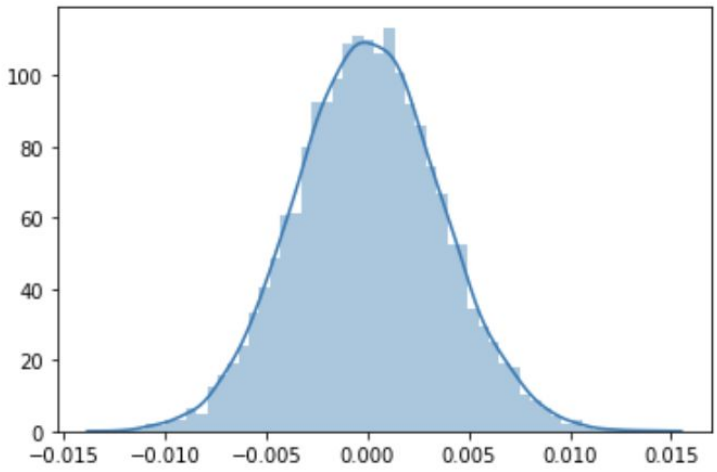

```
observed correlation between minimum nights and number of cancellations:
-0.00473
p-value: 0.1630000000
We fail to reject the null. There is no significant correlation between t
he number of cancellations and the minimum nights required by a listing.
```

Cancellations vs Crime

Null Hypothesis: There is no significant correlation between the number of cancellations and the crime rate of the neighbourhood that the listing belongs in.

Alternative Hypothesis: There is a significant correlation between the two.

We perform the same test and conclude that their observed correlation of -0.01161 is statistically significant at the 1% level.




```
observed correlation between crime rate and number of cancellations: 0.01
161
p-value: 0.0009000000
We reject the null. There is a significant correlation between the number
of cancellations and the crime rate of the neighbourhood that the listing
belongs in.
```

Besides these features, we also want to check if the number of cancellations vary by categories such as neighborhood and room type. Since there are more than two in each category, we cannot perform ttests or permutation tests. Instead, we can use an F-test. In an F-test, we are able to compare the means of various groups and determine if they are equal by looking at their variances. With this technique, we are able to check variations between and within the groups.

Cancellation by Neighborhood?

Null Hypothesis: The average number of cancellations of each neighborhood are not statistically different from each other.

Alternative Hypothesis: The average number of cancellations of each neighborhood are statistically different from each other.

```
f-stat: 4.65329
p-value:0.0000000000
The average number of cancellations of each neighbourhood are statistical
ly different from each other.
```

The f-stat of 4.65 suggests that between-neighbourhood variance is 4.65 time the within-neighborhood variance. The small p-value suggests that the difference is statistically significant.

Cancellation by Room Type?

Null Hypothesis: The average number of cancellations of each room type are not statistically different from each other.

Alternative Hypothesis: The average number of cancellations of each room type are statistically different from each other.

```
f-stat: 11.68846
p-value:0.0000084047
The average number of cancellations of each room type are statistically d
ifferent from each other.
```

The p-value suggests that the differences in the number of cancellations among the room type are also statistically significant.