

Business Understanding

Problem Statement

We have been tasked to understand electric car usage. We will work as a Data Scientist for the Autolib electric car-sharing service company to investigate a claim about the blue cars from the provided Autolib dataset.

In order to achieve this, we will respond to the following claim:

- On a given weekday, the number of BlueCars taken is more than the total number of Utilib and Utilib_14 cars. {Claim- Alternative Hypothesis}
- On a given weekday, the number of BlueCars taken is not more than the total combined number of Utilib and Utilib_14 cars {Null Hypothesis}

This claim is important in guiding the team on determining where to focus their efforts to ensure even distribution of resources in Paris.

Data Description

autolib_daily_events_postal_code.csv

This dataset is an extract of the number of BlueCars, Utilib and Utilib_14 vehicles taken and returned on every day of 2018 in the different postal codes in Paris. It also highlights the number of charging slots taken and freed for each of the days.

The fields are as highlighted below:

<u>Field Name</u>	<u>Description</u>
Postal code	Postal code of the area (in Paris)
Date	Date of the row aggregation
N_daily_data_points	Number of daily data points that were available for aggregation, that day
dayOfWeek	Identifier of weekday (0: Monday -> 6: Sunday)
Day_type	Weekday or weekend
BlueCars_taken_sum	Number of BlueCars taken on that date in that area

BlueCars_returned_sum	Number of BlueCars returned on that date in that area
Utilib_taken_sum	Number of Utilib taken on that date in that area
Utilib_returned_sum	Number of Utilib returned on that date in that area
Utilib_14_taken_sum	Number of Utilib 1.4 taken on that date in that area
Utilib_14_returned_sum	Number of Utilib 1.4 returned on that date in that area
Slots_freed_sum	Number of recharging slots released on that date in that area
Slots_taken_sum	Number of recharging slots taken on that date in that area

The distribution of the three cars is shown during our analysis in the Google colab notebook in Github.

https://github.com/esther-nyokabi/Moringa-Core-Statistics/blob/main/Independent_Project_Hypothesis_Testing.ipynb

Sampling

a) Target population

The target population in this study constituted 16,085 records linked to postal codes in Paris for the year 2018.

b) Sampling Method - Simple random sampling

This study will focus on weekdays. In the dataset, weekdays are represented by 0 to 4 where 0 is Monday and 4 is Friday.

Simple random sampling was used to select one value between 0 and 4 and number 3 was selected.

c) Sample Size

Using the random number '3' selected, all entries that were made on the 3rd day of the week were selected from the total population. This resulted in a sample size of 2,374 records.

Hypothesis Testing Procedure

Null Hypothesis (H_0) - The number of BlueCars taken is less than or equal to the total combined number of Utilib and Utilib_14 cars.

$H_0: \mu \leq$ total combined number of Utilib and Utilib_14 cars.

Alternate Hypothesis(H_a) - The number of BlueCars taken is more than the total combined number of Utilib and Utilib_14 cars.

$H_a: \mu >$ total combined number of Utilib and Utilib_14 cars

Test Statistics & Level of Significance

- The z statistic will be used in this test because the number of samples selected is more than 30
- The confidence level for this test will be 0.95 hence an alpha value of 0.05.

Hypothesis Testing Results

On a given weekday, the number of BlueCars taken is more than the total combined number of Utilib and Utilib_14 cars.

The z-statistic was used to test this hypothesis and the calculated score was -0.95.

From the Z-Score above, the calculated p-value was 0.83. This value was greater than the alpha value of 0.05. For this reason, the null hypothesis that claims that on a given weekday, the average number of vehicles (BlueCars, Utilib and Utilib_14) taken is more than the average number of vehicles being returned was accepted.

Discussion of Test Sensitivity

High sensitivity tests have low specificity. The specificity of a test (also called the True Negative Rate) . The specificity of our test for example refers to how well a test identifies whether On a given weekday, the number of BlueCars taken is more than the total number of Utilib and Utilib_14 cars taken.

A test that is 90% specific will identify 90% of Blue cars taken compared to utilib and utilib_14 cars taken on a given weekday. For our model/ hypothesis we conducted,the specificity is 95%.

Summary and Conclusion

Our results provide support for our research hypothesis as there is still a slight probability that the results occurred by chance and the null hypothesis was correct.