

HYPOTHESIS TESTING

INTRODUCTION

Autolib is an electric car-sharing service company. People usually take the vehicle and later return them. Lastly, they also offer charging services for the vehicles.

PROBLEM STATEMENT

Autolib provided there dataset that we will be making a research on. There has been a rise in claim on the blue cars they usually provide to people. This is the variable we will be investigating for the time being.

- Null: (claim) is the number of blue cars taken in postal code 94700 same as 92170?
- alternative: the number of blue cars taken in postal code 94700 different from 92170?
- Significance level = 0.05

The importance of this hypothesis that will be carried out is to guide the company on its logistics. If there will be a difference, then we will be forced to take more vehicles to postal code 94700 since they will have concluded that there is a rise in demand for the blue car in the area.

DATA DESCRIPTION

The dataset is a daily aggregation by date and postal code of the number of events on the Autolib network.

The source of the dataset is from autolib database that is normally updated when an event is carried out.

HYPOTHESIS TESTING PROCEDURE

On the dataset we will be working on the following variables:

- bluecars_taken_sum
- Postal code
- Type of day

The hypothesis was raised as a matter of concern from the company. They claim that the two regions they have provided have a difference in the number of cars taken. They want us to investigate whether this is true so that they can get to know whether the stated fact is true. This will guide them to come up with a conclusive decision as to whether to add more cars on postal code 94700. The company wants to determine whether there is a rise in demand, so as to increase profit.

The hypothesis will be conducted using t-statistic for we are comparing two samples that have been drawn from the same population. We shall get the mean from the two samples to check whether they are significantly different from each other. This is done by calculating the standard error in the difference between the means, which can be interpreted to see how likely the difference is.

The t-statistic is calculated by the test will be interpreted by comparing it to critical values from the t-distribution. The critical value is calculated using the degrees of freedom and a significance level with the percent point function(PPF). Then, the p-value will be achieved by using the cumulative distribution function(CDF), that will later be used to compare it to the significance level, that is, 0.05. This will enable us to reach a conclusion as to whether we accept or reject the null hypothesis.

HYPOTHESIS TESTING RESULTS

After conducting the hypothesis, the following are the results that I got:

- T-statistic result = 4.874320053778534
- We reject the null hypothesis since the p-value of 1.9999999999989644 is greater than our significant value

SUMMARY AND CONCLUSIONS

The project was able to give results about the blue cars. Its now up to the company to reach a decision whether to go for their plan or not.

The hypothesis was done to a satisfying level and the results will help the company make a decision since they have concrete information.