

AUTOLIB HYPOTHESIS TESTING REPORT

1) Problem statement:

To determine whether there is a difference in the number of bluecars taken from stations in Paris and the number of bluecars returned to the stations during the weekends.

Null Hypothesis: There is no difference in the sum of bluecars taken from stations and the sum of bluecars returned during the weekend.

Alternate Hypothesis: There is a difference in the sum of bluecars taken from stations and the sum of bluecars returned during the weekend.(CLAIM)

This hypothesis is interesting as it will help us find out which of the two days of the weekends is the busiest and will enable the company determine the car provision and distribution. It will also monitor the cars returned /taken for accountability.

our claim test will be: Is the number of blue cars taken from stations in Paris and the number of bluecars returned to the stations equal during the weekends.

2) Data description

This study seeks to determine if the total number of bluecars taken/shared differs during the 2 days of the weekends. Data used was sourced from Autolib electric car-sharing service company to investigate a claim about the blue cars from the provided Autolib dataset. <https://bit.ly/DSCoreAutolibDataset>.

The data set contains variables like the postal code of the area which was paris, the dates of data collection i.e between Jan - July 2018. It has two categoricals in the day_type variable (Weekdays (Mon- Fri) & Weekends (Fri & Sat) encoded numerically.

Our key variables are; dayofweek, bluecars_taken_sum, and bluecars_returned_sum.

3) Hypothesis testing procedure

I used stratified sampling to pick my sample from the main data set, which had **(16,085 * 13)** using python programming.

From my stratified sampling, I got a sample of **454 * 4 entries**. This meant that $n > 30$. As such, I used the z-score to determine the p-value. I did perform normality tests on the data since the statistic I used is designed using assumption that the underlying distribution is normal. This data set had a normal distribution.

I used an alpha level of significance of **0.05**.

4) Hypothesis testing results

From the hypothesis test, the p was **0. 530385573** which is larger than the alpha of 0.05, we found that there was sufficient evidence to prove that the sum of blue cars taken and the bluecars returned are not equal on the weekend.

The z-score was **0.076**. The z-critical value was **1.959963984540054** with the confidence interval being : **(148.6968556616749, 188.90666856739998)**

The point of estimation was -17.71169384763587

5) Summary and conclusions

I rejected the null hypothesis as there was some evidence for me to reject the null hypothesis and accepted the Alternate Hypothesis/claim.

Null Hypothesis: There is no difference in the sum of bluecars taken from stations and the sum of bluecars returned during the weekend.

Alternate Hypothesis: There is a difference in the sum of bluecars taken from stations and the sum of bluecars returned during the weekend.