**HYPOTHESIS REPORT FOR AUTOLIB DATASET**

1.*Problem statement*

The dataset was an autolib dataset that entailed an electric car program within Paris. It showed a compilation of dates when the blue cars were picked from and returned to the particular addresses. The claim being investigated was whether or not the average number or blue cars taken was different from the average number of blue cars returned during that period.

The null hypothesis was set to: The average number of blue cars taken is not different that of the blue cars returned

Whereas the alternative hypothesis or claim was set to: The average number of blue cars taken is different that of the blue cars returned

The hypothesis is important as it will enable me to demystify the data, its distribution and come up with insights that are meaningful to the general population.

2.*Data description*

Dataset used for this investigation was an open dataset about cars in Paris.

It contains variables like;

Postal code of the area (Paris), the dates of data collection.(The dates ranged between January and July of 2018), the number of daily data points that were available for aggregation on the particular days of aggregation within the specified time periods,The blue cars that were taken and returned, the utilib data and the slots set of data .

The problem under investigation was on the averages which would make the null and alternative hypotheses.

It was a set of data that was already collected. However, if i were to collect such comprehensive data, i would use my data response team to go out in the field, collect the data and perform the analysis from which conclusions would later on be made.

3*.Hypothesis testing procedure*

The dataset is big as it contains 16,085 rows and 13 columns. Therefore, I picked a sample from the dataset using a stratified sampling method. This is because the sample would be unbiased.

Stratified sampling required me to group the data into different sets of strata and then randomly pick a sample from the strata. I used python programming to generate my sample.

I used the count and dates of blue cars taken and returned to determine my hypothesis.

The choice behind my null and alternative hypothesis was that, since it was not going to be easy for me to manually group the data or identify the clusters and the sample in them, I decided to work with an average.

Interestingly, I wanted to find out whether the average number of cars that were picked in a day was similar to the average number of cars that were returned on that very day. This information could help determine the future trends of business operations relating to blue cars in the autolib electric car sharing company.

From my stratified sampling, I got a sample of 264 \* 4 entries. .

Therefore, I used the z-score to determine the p-value. I did perform normality tests on the data since the statistic I was using demanded so. The alpha level of significance that I used was 0.05.

4.*Hypothesis testing results*

From the hypothesis test, we found that there was not sufficient evidence to prove that the average means of the blue cars taken and the bluecars returned are not equal. The z-score was -1.1908605153511547 and as a result, the null hypothesis was not rejected. The z-critical value was 0.16478012998031033 with the confidence interval being :

Confidence interval:

(58.26612023051733, 61.81802216430468)

The p-value was 0.1102561727001386 which was greater than the significance level.

5*.Discussion of test sensitivity*

Sensitivity in a statistical test is the measure of performance of a binary classification test. It measures the proportion of the actual positive i.e. the probability of a null hypothesis being true. In this case the sensitivity was 93%.

6. Summary and conclusions

The project was challenging and demanding. I performed exploratory data analysis with hypothesis testing as its implementation.

In conclusion, I failed to reject the null hypothesis because there was not enough evidence to prove that the average means of the blue cars taken and returned are not equal.