**AUTOLIB BLUE CARS HYPOTHESIS TEST.**

1.Problem statement:

Autolib is an electric car sharing service. Blue cars are one of the electric vehicles that are included in this service. The following research will test whether the mean number of blue cars returned is the same as the mean of the blue cars taken, on the weekends in postal code 75001.

Given that blue cars returned = A and blue cars taken = B, then our hypothesis will be:

***Null hypothesis:***

***Alternative hypothesis:***

The hypothesis being tested is important because it is assumed that most times the number of cars returned during the weekends is often less than the number of those taken. This test seeks evidence to support this claim in the postal code stated above.

2. Data Description:

The data necessary for this report will be:

* Blue Cars taken and returned.
* Postal code 75001 only.
* All weekends of the given time period.

The count of blue cars taken and returned is a discrete random variable since it can only take a specific value (whole numbers).

The data is a collection of all electric cars that were provided by Autolib from January 2018 - June 2018. It comprises other cars aside from blue cars. Furthermore, it shows the number of cars taken and returned for every day of the week. This data was collected by autolib’s software that would automatically log information whenever a car was taken, returned.

3. Hypothesis Testing Procedure:

The number of blue cars returned depends on the number of those taken. Therefore, a paired t-test (with =0.05) will be used to test the hypothesis given. This will be done by:

* Clearly stating the null and alternative hypothesis.
* Cleaning the data we have by removing anomalies and outliers.
* Univariate and Bivariate analysis.
* Getting a sample.
* Derive the p-value
* Get confidence interval
* If the Null hypothesis is not rejected then conclude. If it is, find the power of the test before making a conclusion.

The null and alternative hypothesis may be important to study because often, it is assumed that a lot of people do not return blue cars with the same frequency that they are taken.

4. Hypothesis Testing Results:

The result of our test was as follows:

* T-statistic with degrees of freedom=11 is 1.0477
* P-value is 0.317
* Point estimator of the tested parameters was (to the nearest whole number ) =192 for blue cars returned and =191 for blue cars taken.

Since p is greater than the significance level; fail to reject the null hypothesis.

Furthermore, the confidence interval -0.7889, 1.1257, shows that our tested values lie within it further proving the results we have.

5. Test Sensitivity:

Even though the evidence present does not support the alternative hypothesis, more analysis needs to be done. Since the null hypothesis was not rejected the power of the test is not significant. However, was tested, and the result(0.0035) showed that the probability of a type 2 error occuring is very low.

6. Summary and Conclusion:

This project primarily involved sifting our dataset and picking out the needed values only. It also involved removing anomalies, analysis and finally testing the given hypotheses.

In conclusion, there is evidence that shows that the number of blue cars returned is not less than those taken. This does not mean that the quantity is proportional and may need further analysis.