**Autolib Hypothesis Testing Report**

1. **Problem Statement**

This project aims to demonstrate the concept of hypothesis testing by investigating a claim about blue-cars using a dataset from an electric car-sharing service company called Autolib.

I will be comparing the blue car usage rates in Paris and Val-de-Marne, which are two French cities. Below are the specific hypotheses of interest in this analysis:

Null Hypothesis: For both Paris and Val-de-Marne the average number of blue-cars picked up during weekends is the same. Ho : μ1 = μ2

Alternate Hypothesis: The average number of blue cars picked up during the weekend is different for Paris and Val-de-Marne . Ha: μ1 ≠ μ2

1. **Data Description**

The dataset used in the analysis contains records of electric car usage in France from January 2018 to June 2018.

This dataset was initially sourced from opendataparis.com. Some of the dataset information includes postal code, day of the week, and total cars returned or picked up for blue cars, Utilib cars, and Utilib 1.4 cars.

The data available for this analysis is valid and useful towards achieving the project goal, given the availability of both numeric and categorical attributes.

1. **Hypothesis Testing Procedure**

We used a two-sample z-test and p-value to either accept or reject our null hypothesis to perform hypothesis testing. Below are the reasons why we choose to use the z-test as the right test statistic:

• Data points are independent of each other.

• The sample size is greater than 30.

• The sample data has been randomly selected from a population, so there is an equal chance of being selected for each item.

The significance level is defined as the probability of rejecting the null hypothesis when it is true. For this analysis, the significance level chosen is 0.05 or 5%.

If the calculated P-value from the test statistic is less than 0.05, we reject the null hypothesis. If the value is greater than or equal to 0.05, then we accept the null hypothesis.

Since α = 0.05, and the test statistic is a one-tailed test, the critical value of z is 1.645, as per the z critical value table.

1. **Hypothesis Testing Results**

The p-value of the two sample z test was calculated and found that the p values are 3.790179011261841e-52, lower than 0.05. Therefore, we reject the null hypothesis. This means that the average number of blue cars picked up on weekends is not the same for Paris and Val-de-Marne.

In this case, we accept the alternate hypothesis. The calculated z statistic is 15.195455826186167 against the expected critical value of 1.645. It means that the data points are 15.2 standard deviations away from the mean. This is further evidence that we need to reject the null hypothesis.

1. **Discussion of Test Sensitivity**

I imported the *starsmodel* library in the python notebook to check the power of the hypothesis test. Then calculated the two-sample size based on hypothesis testing parameters.

The power of the test results is 64%. Preferably, this should be closer to 90% for a more significant statistical test.

Some of the factors that could have affected the statistical power of the test include:

• The population did not follow a normal distribution.

• The sample sizes for the two strata were not the same.

1. **Summary and Conclusions**

The null and alternate hypothesis was successfully defined. Performed the sampling technique and carried out hypothesis testing, which resulted in the rejection of the null hypothesis. The results concluded that the average number of blue cars picked up in Val-de-Marne is not the same in Paris.