**DATA REPORT ON AUTOLIB NETWORK HYPOTHESIS TESTING**

1. **Problem statement**

* The data we are working on is an autolib dataset. The dataset contains data on the bluecar sharing on the day it is taken and returned in different postal codes.

The company would love to know the difference in the bluecars being taken from

area X and the bluecars being taken from area Y.

The postal codes for area X is 75015 and the postal code for area Y is 75017.

* Our null hypothesis;

Our null hypothesis is that area X has greater or equal number of bluecars being

taken than area Y on the weekend.

Our Alternative hypothesis ;

Area X has less bluecars being taken compared to area Y on the weekend

* Finding out the difference in the two postal codes is interesting because , getting to know the difference between two regions and their contribution in the business is important to the company. Getting a deep insight on the two postal codes also helps the company weigh on their investment and helps them track on the usage of their services.

Getting to know the postal code that is better than the other would help them single out on the things they do to have such and why it is doing well.

They would also know why the other postal code is not doing so well.

1. Data description

The dataset is from the autolib car sharing network.The autolib dataset contains aggregation of a number of events by date and postal code.

The dataset contains 13 columns which has an attached dataset that explains each column. Postal codes are based in paris, the date for the aggregation is given, the number of the daily data points that were available for aggregation , identifier of the day of the week given in numbers, (0 ; monday to 6;sunday) . the day type is provided if it is a weekday or a weekend, the type of car , the sum of a utilib or a bluecar or utilib 1.4 when taken or returned is recorded, the number of recharging slots released that date in that area and the sum of the recharging slots taken that date in that region.

The rows in the dataset are 16085.

In our research, we are interested in the sum of bluecars being taken and their postal codes . The type of day is an important variable since we want to know the days that the company is getting more sales. We are going to sample the data from the company's data. We create another dataframe that contains the columns we need. Splitting the dataset to weekdays and weekends makes our work easier.

1. Hypothesis testing procedure

* To test our hypothesis, we first split the dataset to weekdays and weekends dataframe.

We will find their central tendencies to ensure that the most profit is made over the weekend. This helps us make an assumption. From our data, more bluecars were taken on weekends more than weekdays. Weekends had a mean of 151 whereas weekdays had a mean of 161 .

* We then formulate our null hypothesis and our alternative hypothesis as we have stated them above.

H0 = area X has greater or equal number of bluecars taken than area Y over the weekend

Ha = area X has less number of bluecars taken than area Y over the weekend

Comparing two postal codes, makes sense getting their differences and similarities. To compare the two, we need an hypothesis that makes one better than the other and the alternative hypothesis gives the opposite if the null hypothesis

* We then calculate the mean and the standard deviations of the bluecars taken over the weekend in the two postal code areas.

Area X has a mean of 1072 whereas area Y has a mean of 831.

Their standard deviation 211 and 159 respectively

* We then declare the alpha level

Our alpha level is 0.05. Our samples data is small we can use 0.05 as our alpha level

* Get the p value

We will use t statistic to get our p value. Our sampled data is less than 30. That is why I chose to use it. The X p value is 0.4958 and the p value for Y is 0.14522

* We then compare the p value to the alpha level and choose to either reject or fail to reject the null hypothesis.

The p value is more than the alpha level so we fail to reject the null hypothesis.

1. Hypothesis testing results

* We did not reject our null hypothesis because the p value was greater than the alpha level.we accepted the null hypothesis.
* From t statistic we found the value of x and y as x = 0.3434 , y = -1.0571

Our p value for X and Y as 0.4958 and 0.14522 respectively

The point estimate for the parameter is 1000

We did set our confidence interval at 95% to get the standard error and we are sure the mean is between 517 and 1144 bluecars.

1. Discussion of test sensitivity

Our null hypothesis was not rejected. We are using a parameter x to test our hypothesis and we have our confidence level. A change in n would result in a decrease of beta.

1. Summary and conclusions

In summary, the sum of bluecars in area X is greater or equal to the sum of bluecars in area Y over the weekend. We failed to reject our null hypothesis. This could be due to increased social activities in area X, most people run personal errands over the weekend or area X could be having a higher working population and the service being offered are good there.

our assumptions and hypotheses can be relied on. The project was a success and the findings can be relied on to make a decision.

To conclude , the company should consider have bluecars readily available in area X to meet the demend and to maximize on their profit. This will be good for the company.