Problem Statement

The electric car sharing service is a viable alternative to the use of individuals using their own cars. It saves on fuel, mileage and harmful emissions. They are especially of great advantage to use in the urban areas where traffic is a major problem to the residents. Another perk of using these services is that you are able to return the car to another station within the service area making it very convenient to the users. The dataset contains the usage of the blue cars, utilib and utilib 14. In this particular study we'll be investigating the number of blue cars taken from stations within two separate postal codes, our focus being on the weekdays. The provided dataset is a daily aggregation, by date and postal code, of the number of events on the Autolib network (car-sharing and recharging).

We'll be testing two hypotheses, the null and alternative hypothesis.

Null Hypothesis, H_0: There is no difference in the average number of blue cars taken from stations with postal code 92500 and those taken from stations with postal code 92600 on weekdays.

Alternative Hypothesis, H_1: There is a difference in the average number of blue cars taken from stations with postal code 92500 and those taken from stations with postal code 92600 on weekdays.

As seen from the hypothesis we really want to find out whether there is a difference in how busy the stations within the given postal codes get during the weekdays. This is of course dictated by the number of blue cars taken from the stations.

Data Description

The dataset and glossary used for this study were gotten from [http://bit.ly/DSCoreAutolibDataset] and here [Link] respectively. The mean number of blue cars taken was found to be 49 per day while the median was 34 cars. The blue cars are used more during the weekdays as compared to the weekends.

Hypothesis Testing Procedure

The main reason we chose this particular hypothesis is because we wanted to figure out if indeed there was any remarkable difference in the average number of blue cars used in two random postal codes during weekdays or not. This would help tell us which postal code stations people frequent more during the weekdays. We'll be using the z-test to test our hypothesis since our sample size is above 30. The assumptions we'll be making while conducting the z-test include:

- ❖ The Sample is randomly selected and is representative of the whole population.
- ❖ The data, when plotted, results in a normal distribution, bell-shaped distribution curve.
- ❖ there is homogeneity of variance. The variance of the samples are approximately equal.

The significance level or alpha used for this test is 0.05

Hypothesis Testing Results

After conducting our z-test, we found the Z score to be -0.13127069818137072 and the p value to be 0.4477805853217691. The p value of 0.4478 is greater than our alpha of 0.05 meaning there is no statistical significance for us to reject our null hypothesis.we will therefore fail to reject the null hypothesis of there is no difference in the average number of blue cars taken from stations with postal code 92500 and those taken from stations with postal code 92600 on weekdays.

Discussion of Test Sensitivity

The z test is a highly sensitive test and has given us accurate results for the hypothesis we were testing.

Summary and Conclusions

The hypothesis testing process commenced with us going through the dataset to better understand it. We then loaded it into our programming environment, imported the libraries needed to manipulate the data and read through it by viewing the top and last five rows. Some data cleaning was done to ensure we were working with clean data and only what we needed for the hypothesis testing. For the random sampling, we decided to go with stratified sampling with the postal codes being our strata. A z-test was conducted to test our null hypothesis which we decided not to reject after comparing the test statistic and p value.

There's a 44.77% that there is a difference in the average number of blue cars taken from stations with postal code 92500 and those taken from stations with postal code 92600 on weekdays.

The colaboratory notebook for the hypothesis testing can be found here [Hypothesis Testing](#).