

**Independent Project: Part 2- Hypothesis Testing Report**

# Problem Statement

## **The Dataset**

autolib\_daily\_events\_postal\_code.csv is a dataset on electric car usage derived from Autolib, an electric car-sharing service company. From it the area of interest chosen randomly (by simple Random sampling) was postal code 75012 and 93310.The periods, weekdays, were chosen non-randomly.

## **Null and alternate hypothesis**

The claim here is that there is a difference between the mean of blue cars taken in postal code 75012 and postal code 93310.

The null hypothesis is that there is no difference between the mean of blue cars taken in postal code 75012 and postal code 93310.

The alternate hypothesis is that the claim .i.e. there is a difference between the mean of blue cars taken in postal code 75012 and postal code 93310.

# Data Description

The data was obtained from <https://bit.ly/DSCoreAutolibDataset> and the data variables for the dataset are explained in this link <https://bit.ly/DSCoreAutolibDatasetGlossary>. A sample from the two postal codes mentioned in the problem statement was selected by was od simple random sampling. The number of samples chosen for postal code 75012 was 40 while the number of samples chosen from the postal code 93310 was 50.

# Hypothesis Testing Procedure

The test statistic chosen was the Z-Score. This is because the samples that the researcher was working with are greater than 30 and because the sample variances can be obtained from the data.

A confidence level of 0.05 is used.

# ****Hypothesis Testing Results****

The test statistic (Z- score yielded), was 1.125. The table below obtained from <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-statistics/what-is-a-z-score-what-is-a-p-value.htm> was used in the interpretation of the results.

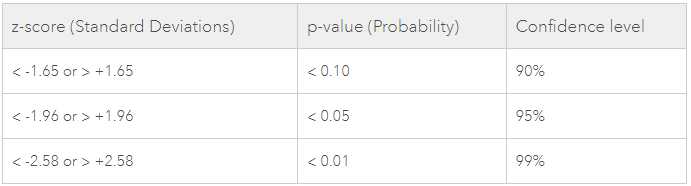


Figure : Z-Score Table

The highlighted row above in figure 1 is used. Our test statistic is les than 1.96 thus it lies in the rejection region area. This therefore leads the researcher to reject the null hypothesis meaning that *the mean of blue cars taken in the postal code 75012 area is* ***not*** *the same as the mean of blue cars taken in the postal code 93310 area.* It also follows that the alternate hypothesis is retained.

The point estimate parameter is the sample mean which estimates the population mean. Since the researcher was working with two populations, the area of postal code 75012 is chosen whose sample mean is 441.95. The Confidence interval around the parameter is further calculated yielding 421.917 < populationmean < 461.983. In reality the mean of this particular population is 440.58035714285717 which lies with the range shown above.

# Summary and Conclusion

The hypothesis testing was successfully carried out. It begun with coming up with the null and alternate hypothesis as indicated in the problem statement. A test statistic was then chosen, which in this case was the z-score. From computations, done in python the z-score obtained was 1.125 which lies in the rejection area of the normal curve and thus the alternate hypothesis was retained i.e. there is a difference between the mean of blue cars taken in postal code 75012 and postal code 93310.

With this knowledge the researcher would delve deeper in the other phase of this project and find out now that the means differ where, between the two postal codes do we have higher means and try to find out the factors that contribute to that.