**Hypothesis Testing Report**

1. **Problem statement**

The research problem is to create a null hypothesis from the Autolib dataset and do testing to either accept or reject the hypothesis tested.

The problem statement is to determine if the difference in the means of blue cars taken in random postal codes equivalent to 0 or not.

Hypothesis testing will be conducted. The significance level which is 0.5 will be compared to the P-Value.

If :

P-Value < Significance level, the null hypothesis will be **accepted.**

or:

P-Value > Significance level, the null hypothesis will be **rejected.**

**Null hypothesis**

The difference in the means of blue cars taken in postal code 75003 and 92420 is equal to 0.

**Alternative hypothesis**

The difference in the means of blue cars taken in postal code 75003 and 92420 is not equal to 0.

1. **Data Description**

The dataset and glossary to use for this project can be found in this link [http://bit.ly/DSCoreAutolibDataset] and here [Link] respectively.

The dataset has 16085 rows and 13 columns.

To work on this project, we will perform the following analysis with Python;

* Load datasets
* Find and deal with outliers, anomalies, and missing data within the dataset.
* Plot appropriate univariate and bivariate summaries recording our observations.
* Implement the solution by performing hypothesis testing.

1. **Hypothesis testing procedure.**

First, we will create a sample of the dataset. In this case, we chose n= 2.This is done using a simple random sample method.The significance level = 0.5, confidence interval = 95%

After calculating the sample n, we will calculate the test statistic using z-test which gives us a p-value.

Finally, we compare the p-value with the significance level. If :

P-Value < 0.05, the null hypothesis will be **accepted.**

or:

P-Value > 0.05, the null hypothesis will be **rejected.**

1. **Hypothesis results**

The p-value = 9.209.

P-value > 0.05

Hence, it is not statistically significant and indicates weak evidence against the null hypothesis. We, therefore, fail to accept the null hypothesis.

1. **The sensitivity of testing/Power of the test**

In normal cases, the lower the significance level, the lower the power of the test. If one reduces the significance level (e.g., from 0.05 to 0.01), the [region of acceptance](https://stattrek.com/Help/Glossary.aspx?Target=Region%20of%20acceptance) gets bigger. As a result, you are less likely to reject the null hypothesis. This means you are less likely to reject the null hypothesis when it is false, so you are more likely to make a Type II error.

1. **Summary and Conclusions**

According to our hypothesis testing the p-value is 9.209.

A p-value higher than 0.05 (> 0.05) is not statistically significant and indicates weak evidence against the null hypothesis.

Therefore, we fail to accept the null hypothesis.