Autolib

Introduction

Today we are going to analyse the autolib data. The data has 16,085 records in it and 13 columns.We will be analysing the probability that the number of blue cars being used has decreased.

Problem Statement

Since we can calculate the mean, I calculated the mean, once that was calculated we divided the mean with the sample size to get the probability since E(X) = n * p. With that I calculated the p which is the probability. I used this to state the null and alternate hypothesis.

H_0 : p = 0.14

H_1 : p < 0.14 (Since we are checking if the usage has decreased.

Data Description

The data is a csv file that has the following columns:

- Postal Code
  This is the postal code of the area in Paris
- Date
  Date of the raw aggregation
- N_daily_data_points
  Number of daily data points that were available for aggregation that day
- Day of the week
  Identifier of weekday (0: Monday -> 6: Sunday)
- Day type
  Weekend or weekday
- Blue cars taken sum
  Number of blue cars taken that date in that area
- Blue cars returned sum
  Number of blue cars returned that date in that area
- Utilib taken sum
  Number of utilib taken that date in that area
- Utilib returned sum
  Number of utilib returned that date in that area
- Utilib 14 taken sum

Number of utilib 1.4 taken that date in that area

- Utilib 14 returned sum
  Number of utilib 1.4 returned that date in that area
- Slots freed sum
  Number of recharging slots released that date in that area
- Slots taken sum
  Number of recharging slots taken that date in that area

However for this project I intend to work with the number of blue cars taken. The blue cars are one of the products Autolib sells, they use renewable energy to fuel them, hence good for the environment .

Hypothesis testing procedure

The hypothesis that we came up with , is whether the number of blue cars being taken has reduced. So to do this I took a random sample using the cluster method. I then began testing on this sample, So i calculated the mean which was 125.669. From there since we n which is the number of the sample I divided the mean by n to get a p_value. After calculating the p-value. I modelled my sample using the binomial distribution. In addition to this I took a random number to test whether the hypothesis is true. Considering the numbers were too large I had to convert it to normal distribution. After converting it to normal distribution I calculated the Z score of the normal distribution.This gave me a figure of -0.445. I went to the tables to check for positive 0.445 , then I subtracted the answer from one.

Hypothesis Testing Results

When I tested the hypothesis , i got a significance level of 32% percent which is too high, this led me to accept the H_0, as it did not lie within the significance level of 5%.The point interval is 0.259 and hence this shows that our estimator over estimates the true mean by 0.259.We also identified the confidence level of 95% interval which lies between 124.82 and 126.51.In addition to that there was a standard error of 0.4308 which translates to 43% standard error.

Summary and conclusions

My conclusion is that the hypothesis is false, there is insufficient evidence to show that there has been a decrease in the taking of cars. 32.9% is what I got as the significance level. This led me to accept the null hypothesis. In addition to this