KELVIN NJUNGE

HYPOTHESIS TESTING REPORT

Problem Statement

This dataset was an autolib dataset that contained details about the operation of cars within Paris. It showed a compilation of dates when the blue cars were picked from and returned to the particular addresses. The claim being investigated was whether there was a difference in the average number of blue cars returned on saturday to that of blue cars returned on sunday during that period.

- **Null hypothesis**

There is no difference in the average number of blue cars returned on saturday and that of the blue cars taken on sunday?

- **Alternative hypothesis**

There is a difference in the average number of blue cars returned on saturday to that of the blue cars taken on sunday(CLAIM:)

Hypothesis testing is important here because it evaluates two mutually exclusive statements i.e the null and alternative hypothesis about a population to determine which statement is best supported by the sample data

Data Description

The dataset I used for this investigation was an open dataset about cars in Paris. It contains variables like the postal code of the area which was paris, the dates of data collection. The dates ranged between 1st of January 2018 and 19th June 2018. With also had the number of daily data points that were available for aggregation on the particular days of aggregation within the specified time periods. The days of the week were the usual monday to friday with the specifications and special assignments of days. Weekday or weekend the dataset had the specific days within the time period. The blue cars that were taken and returned, the utilib data and the slots set of data were also contained in the dataset. The problem under investigation was on the averages which would make the null and alternative hypotheses.

It was a set of data that was already collected. However, if i were to collect such comprehensive data, i would use my data response team to go out in the field, collect the data and perform the analysis from which conclusions would later on be made.

Hypothesis testing procedure.

The dataset was large as it contained 16085 rows and 9 columns . Although, from the dataset i was interested with those days that were weekend(sunday,saturday) only which blue car is returned. i located weekend and the total columns to rows were 4541 * 10. I then picked a sample here using random sampling method
I also performed a normality test from a sample dataset for a blue car taken to see if it follows a normal distribution or not. I used the Quantile_Quantile(Q-Q) Plot.

Hypothesis testing results

From the hypothesis test, we found that there was not sufficient evidence to prove that there is a difference in the average number of blue cars returned on saturday to that of the blue cars returned on sunday..

In the interpretation of the p -value, it was greater than the alpha level of significance which was 0.05. Hence i failed to reject the null(accepted the null).

Summary and conclusion.

In the course of my project , I performed exploratory data analysis using univariate and bivariate analysis and implemented my solution with hypothesis testing.Since my p-value was greater than my alpha level of significance, I failed to reject the null hypothesis because there was not enough proof that there is a difference in the average number of blue cars taken on saturday to that of the blue cars taken on sunday (CLAIM).