Moringa School

DSC 14

IP Week Four

Yves Havugimana

Project Report

Mean comparison between Bluecars taken per day on weekdays and Bluecars taken per day on Weekend

Github link: <https://github.com/yves-moringa/CoreWeekFourIp.git>

# Problem statement

As is commonly known, traffic on weekends is different from the traffic during weekdays.

As a way to find out if that behavior affects the business. A comparison will be made between the number of Bluecars taken on weekdays and the number of Bluecars taken on weekends.

# Data Description

On every given date, our dataset provides the following information:

Dataset: http://bit.ly/DSCoreAutolibDataset

Glossary: http://bit.ly/DSCoreAutolibDatasetGlossary

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| BlueCars\_taken\_sum | Number of bluecars taken that date in that area |
| BlueCars\_returned\_sum | Number of bluecars 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 |

The dataset has 16085 records and 13 fields.

Our focus is on the “Bluecars\_taken\_sum '' variable and we perform a significance test to identify a difference between the means of Bluecars taken during weekdays and Bluecars taken during weekends.

# Hypothesis Testing Procedure

The null hypothesis is that there is no difference between the mean of Bluecars taken per day during weekdays and the mean of Bluecars taken per day during weekends.

We will perform a two tailed z-test statistic as we assume that the population is normally distributed.

The other conditions for a z test are satisfied by drawing a sample above 30 units and below 10% of the population.

The significance level applied to our test statistic will be alpha = 0.05.

As the dataset contains 11544 records for weekdays and 4541 records for weekend, our sample will have 1154 record for weekdays and 454 records for weekend.

# Hypothesis Testing Results

The results of the test have confirmed that there is no significant difference between the mean number of Bluecars taken per day on weekends and the mean number of Bluecars taken per day on weekdays. The p-value calculated (0.316216) is higher than the significance level of 0.05, we set for the test.

The sample mean for Bluecars taken per day on weekdays is 109 bluecars taken per day with a population standard deviation of 169.62 and on weekends is 156 bluecars taken per day with a population standard deviation of 218.56.

# Discussion of Test Sensitivity

By randomly varying the sample, the conclusion has constantly been reached that there is no significant difference between the two means

# Conclusion

The project concludes that the population mean number of Bluecars taken during weekdays is not significantly different from the population mean number of Bluecars taken during weekends, with a significance level of alpha equal to 0.05.