Statistical Analysis of New York City TLC Data

Executive summary report

Overview

The New York City Taxi and Limousine Commission (TLC) has approached the data consulting firm Automatidata to develop an app that enables TLC riders to estimate the taxi fares in advance of their ride. The New York City TLC would like the data team at Automatidata to analyze the relationship between fare amounts and payment type. The team agrees that the next step is to perform a hypothesis test using the data.

Objective

- Explore the project data
- Compute descriptive statistics
- Conduct a hypothesis test using the New York City TLC dataset

Results

- There are 15,265 recorded taxi rides paid with credit card. The mean fare amount is 13.43
- There are 7,267 recorded taxi rides paid with cash. The mean fare amount is 12.21
- The hypotheses for the hypothesis test is as follows:
  - H_0: There is no difference in the average fare amount between customers who use credit cards and customers who use cash.
  - H_A: There is a difference in the average fare amount between customers who use credit cards and customers who use cash.
- The p-value (6.79 x 10^-12) is less than the significance level 0.05
- Reject the null hypothesis.
- There is a statistically significant difference in the average fare amount between customers who use credit cards and customers who use cash. Customers who pay in credit card tend to pay a larger fare amount than customers who pay in cash.

Next Steps

- Consider implementing strategies to encourage taxi customers to pay with credit card.