

Capstone Project 1 Statistical Data Analysis

After the data is wrangled and cleaned, and the preliminary EDA complete, the hypothesis testing begins. Questions arise through exploring the data.

-Is there a connection between certain colors of automobiles being in more accidents due to their color?

Null Hypothesis:

There is no statistical significance in the likelihood of an Automobile getting into an accident related to color.

Alternative Hypothesis:

Certain colors of Automobiles show a higher likelihood for getting in an accident.

A Chi-Square Test was run to discover if there is a significant association between the color of an automobile and its likelihood for being in an accident.

The outcomes were:

Significance level: 0.05

Degree of Freedom: 1

chi-square statistic: 488.4484795288744

critical_value: 3.841458820694124

p-value: 0.0

Comment:

Therefore the Null Hypothesis is rejected. Accepting the Alternative Hypothesis that certain colors show a greater or reduced likelihood for an Automobile to get into an accident.

(Detailed testing can be found in Capstone_1_Data_Story notebook)

-Can accidents be predicted based on the month?

Null Hypothesis:

There is no statistical significance in the likelihood of an Automobile getting into an accident related to the month.

Alternative Hypothesis:

Certain months of the year show a greater or reduced likelihood for an Automobile to get into an accident.

A Paired T-Test was run to discover if there is an association in the likelihood of an Automobile getting in an accident related to the month.

6.672627141580349e-69

reject null hypothesis

Comment:

The test shows to reject the null hypothesis. Therefore certain months of the year do show a greater or reduced likelihood for an Automobile to get in an accident.

Conclusion:

Certain colors of vehicles have a slightly higher risk for being in an accident when looking at the percentages related to each individual color. The risks overall vary by <1%, and therefore it doesn't seem a strong statement to make for clients to base decisions from.

Certain months of the years showed a variance in accidents that could vary by more than 200 compared to other months in the same year. This merits further investigation by a client or company who wishes to use this information. The weather may have played a factor or perhaps the economy.