# MechaCar\_Statistical\_Analysis

## Linear Regression to Predict MPG

Linear Regression was applied to vehicle\_length, vehicle\_weight, spoiler\_angle, ground\_clearance, and AWD variables to determine their relationship to the variable mpg. According to the results, vehicle\_length and ground\_clearance statistically unlikely to provide random amounts of variance to the linear model.

A screenshot of a computer

Description automatically generated with low confidence

With a p-value of 0.0000000000535, we reject the null hypothesis, therefore the slope of the regression is considered not to be zero.

71% of the variability of mpg is explained by the model while the p-value remained significant. The model could be said to effectively predict mpg of MechaCar prototypes.

## Summary Statistics on Suspension Coils

The summarize() function was were applied to suspension coil data that was categorized into three lots in order to determine if they meet design specifications. The design specifications for the MechaCar suspension coils dictate that the variance of the suspension coils must not exceed 100 pounds per square inch. Taking data from all lots combined, PSI variance appears to be within specifications.

Table

Description automatically generated

However, breaking data into lots, Lot 3's variance is 170PSI.

Table

Description automatically generated

According to MechaCar data, all lots do not meet specifications.

## T-Tests on Suspension Coils

T-tests were applied to were applied to suspension coil data that was categorized into three lots in order to determine if the PSI for each manufacturing lot is statistically different from the population mean of 1,500 pounds per square inch.

When assessing all lots combined, t-test results indicate that the mean is statistically similar to 1,500 as one would expect.

Graphical user interface, text, application

Description automatically generated

Whether t-testing the lot\_summary table or MechaCar\_tbl, both p-values are greater than 0.05%, thus one cannot reject the null hypothesis and mean is believed to be statistically similar to 1,500 PSI.

When t-testing each lot separately, we see that Lot 3’s p-value is less than 0.05% leading us to reject the null hypothesis and concluding the alternate hypothesis that the mean is not equal to 1,500.

Text

Description automatically generated

## Study Design: MechaCar vs Competition

We have been requested to design a statistical study to compare performance of the MechaCar vehicles against performance of vehicles from other manufacturers. In order to assess how the MechaCar performs against the competition we would identify metrics that were relevant to comparisons made by potential purchasers of cars. We have selected surveyed ratings of customer satisfaction as a candidate to evaluate for measurements of performance:

We presume that customer satisfaction represent an indication of how the MechaCar is performing in the market.

We will develop a model that fits dependent metrics (including price, city or highway fuel efficiency, interior space, horsepower, and safety rating) as explaining increases in customer satisfaction. Other dependent variables may be considered, where data is available, that may improve the model. The developed model estimating Customer Satisfaction assessed against a test group to measure its effectiveness as a predictor.

For each independent variable, a large group of consumers would be surveyed. Sample questions would be similar to those below:

In terms of overall satisfaction with a vehicle, PRICE has a significant influence on my feelings about the vehicle.

Agree

Disagree

Doesn’t matter

In terms of overall satisfaction with a vehicle, FUEL EFFICIENCY has a significant influence on my feelings about the vehicle.

For each independent variable, the null hypothesis would be that the consumer DID NOT select AGREE.

The alternative hypothesis is that the consumer either Disagreed or Didn’t Care.

What metric or metrics are you going to test?

What is the null hypothesis or alternative hypothesis?

What statistical test would you use to test the hypothesis? And why?

What data is needed to run the statistical test?

What is performance?