# 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 are 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 ratings of customer satisfaction from a large survey as a candidate to evaluate for measurements of performance. We presume that customer satisfaction represents an indication of how the MechaCar is performing in the market.

We will develop a model that fits dependent metrics (including mpg, price, interior space, horsepower, safety rating, and recalls within first three years.) as explaining customer\_satisfaction. Other dependent variables may be considered, where data is available, that may improve the model. The survey will ask the customer to identify the MechaCar model purchased and provide a satisfaction rating on a scale of 1 to 10. This provides the data for customer\_satisfaction.

Using MechaCar’s data on the dependent variables (based on the actual car purchased) as this provides non-subjective data on the car the customer is rating. The objective would be to develop a multiple linear regression model that explains customer\_satisfaction based the values of the dependent variables from a test group of the large number of samples. The developed model estimating customer\_satisfaction will be assessed against a test group which had been excluded from the data used to develop the model to measure its effectiveness as a predictor.

For each dependent variable, the null hypothesis would be that the variable provides random amounts of variance to customer\_satisfaction. The alternative hypothesis would be that the variable does not provide random amounts of variance to customer\_satisfaction, meaning that it is meaningful in the explaining the variability of customer\_satisfaction.

For the model, the null hypothesis would be that the regression provides random amounts of variance to customer\_satisfaction. The alternative hypothesis would be that the regression does not provide random amounts of variance to the customer\_satisfaction rating, meaning that it effectively predicts customer\_satisfaction.

Once a satisfactory model is developed, dependent variable data of competitors, which is available publicly, can be run through the model to obtain their customer\_satisfaction score. This can easily be compared to MechaCar’s customer\_satisfaction score to determine how it is doing compared to its competition.