Overview

A few weeks after starting his new role, Jeremy is approached by upper management about a particular project. AutosRUs’ latest prototype, the MechaCar, is suffering from production troubles blocking the manufacturing team’s progress. AutosRUs’ upper management has called on Jeremy and the data analytics team to review the production data for insights that may help the manufacturing team.

In this challenge, you’ll help Jeremy, and the data analytics team do the following:

* Perform multiple linear regression analysis to identify which variables in the dataset predict the mpg of MechaCar prototypes
* Collect summary statistics on the pounds per square inch (PSI) of the suspension coils from the manufacturing lots
* Run t-tests to determine if the manufacturing lots are statistically different from the mean population
* Design a statistical study to compare the vehicle performance of the mechanical vehicles against vehicles from other manufacturers. For each statistical analysis, you’ll write a summary interpretation of the findings.

The project aims to analyze metrics that can affect the manufacturing of a new car prototype and compare vehicle performance across different manufacturer lots. The metrics for this project include vehicle length, weight, spoiler angle, ground clearance, AWD capabilities, MPG, and PSI.

Results

1. Linear Regression to predict MPG

Text

Description automatically generated

Text

Description automatically generated

* The vehicle length and ground clearance coefficients can statistically provide a non-random amount of variance to the mpg values. Based on the p-values, there is a random amount of conflict within the dataset.
* At p-value = 5.35e-11 < 0.05, the null hypothesis is rejected because of the p-value.
* An adjusted R-square of 0.6825 concludes that this linear model predicts the mpg of MechaCar prototypes relatively well.

1. Summary Statistics on Suspension Coils

Graphical user interface, application, table

Description automatically generated

Table

Description automatically generated

Based on the summary statistics, lot 3 has a more considerable variance in performance and consistency, with a variance = 170.29 PSI. The conflict in lot 3 increases the variance of the overall lot. The design specifications for the MechaCar suspension coils dictate that the variance of the suspension coils must not exceed 100 pounds per square inch. With the overall variance of 62.29 <100, thus the manufacturing data of the lot meet the design specifications.

1. T-test suspension coils

Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated

1. Lot 1 sample has a sample mean of 1500; with a p-Value of 1 > 0.05, we cannot reject the null hypothesis that there is no statistical difference between the observed sample mean and the presumed population mean (1500).
2. Lot 2 has a sample mean of 1500.02 and a p-Value of 0.61 > 0.05; the null hypothesis cannot be rejected, and the sample mean, and the population mean of 1500 are statistically similar.
3. Lot 3 has a sample mean of 1496.14, and the p-Value is 0.04 < 0.05. All indicate to reject the null hypothesis that this sample mean, and the presumed population mean are not statistically different.
4. Study Design: MechaCar vs. Competition

Metrics

Collecting data for comparable models across all major manufacturers for the past three years for the following metrics:

* Current Price (Selling): Dependent Variable
* Safety Feature Rating: Independent Variable
* Drive Package: Independent Variable
* Engine (Electric, Hybrid, Gasoline / Conventional): Independent Variable
* Resale Value: Independent Variable
* Average Annual Cost of ownership (Maintenance): Independent Variable
* MPG (Gasoline Efficiency): Independent Variable

Hypothesis: Null and Alternative

* Null Hypothesis (Ho): MechaCar is priced correctly based on its performance of critical factors for its genre.
* Alternative Hypothesis (Ha): MechaCar is NOT priced correctly based on the performance of critical factors for its genre.

Statistical Tests

A multiple linear regression would be used to determine the factors that have the highest correlation with the list selling price (dependent variable); which combination has the most significant impact on cost (it may be all of them!)