Module 15 Challenge write-up

**MechaCar Statistical Analysis**

R & R scripts

**Linear Regression to Predict MPG**

Multiple linear regression models allow us to determine if there is a significant statistically relationship between any number of variables to a single output. This module’s challenge had us create a linear regression model in “R script” to determine the significance for several design variables on a vehicle’s miles per gallon (mpg), we were using data provided for 50 of MechaCar prototypes. We selected the following five features for our analysis: vehicle weight, spoiler angle, ground clearance, All Wheel Drive (AWD) and vehicle length.

Text

Description automatically generated

* Which variables/coefficients provided a non-random amount of variance to the mpg values in the dataset?

Only two of the five variables proved to have statistically significant impact on the mpg based on the above result. Those with “\*\*\*” are the variables that are significant. Both vehicle length (p-value of 2.60e-12) and ground clearance (p-value of 5.21e-08), meet the standard 0.05 p-value that indicates a statistical significance, while all the other variables did not. The estimate numbers indicate the increase in MPG’s per one unit increased in the variable measurement.

* Is the slope of the linear model considered to be zero? Why or why not?

The slope of the linear model would not be considered zero. Since the overall p-value for the multiple linear regression, all five variables, is 5.35e-11, which is below the significance level of 0.05; therefore, we can reject the null hypothesis of a zero slope.

* Does this linear model predict mpg of MechaCar prototypes effectively? Why or why not?

With r-squared value of 0.7149, which means that the five selected variables account for over 70% of the changes in gas milage; and with a significant p-value of 5.35e-11 we can determine that our model effectively predicts mpg of MechaCar prototypes.

## Summary Statistics on Suspension Coils

Suspension coil spring specifications require that their variance must not exceed 100 pounds per square inch (PSI). As seen in the total\_summary table below, the variance across the three manufacturing lots we analyzed is 62.29 PSI, which easily meets the outlined design specifications for the MechaCar suspension coils.

Graphical user interface, application, table

Description automatically generated

While the variance for the manufactured suspension coils overall meets the required PSI specification, when we analyze the coils at a manufacturing lot level, we find that one lot exceeded the specification. The lot\_summary  table below shows the variance of suspension coils for each manufacturing lot. Both Lot 1 and Lot 2 meet the design specifications of 100 PSI or less with variances of 0.98 and 7.47 PSI, respectively. But Lot 3 does not, and with a variance of 170.29 PSI significantly exceeds the 100 PSI limit.

Graphical user interface, application, table

Description automatically generated

## T-Tests on Suspension Coils

Four t-tests were performed to evaluate whether the PSI of suspension coils for all manufacturing lots and each individual lot is statistically different from the population mean of 1,500 PSI.

**All manufacturing lots:**

The t-test of the PSI across all manufacturing lots verses the mean of the population (1500 PSI) resulted in a p-value of 0.06028, which is above the significance value of 0.05. Based on this, there is not sufficient evidence to reject the null hypothesis and therefore we can conclude there is no statistical difference between samples and the mean.

Graphical user interface, text, application

Description automatically generated

**Lot 1**

Text

Description automatically generated with medium confidence

The t-test of the PSI of Lot 1 and the population mean of 1,500 PSI resulted in a p-value of 1. Since the p-value is above the significance value of 0.05, there is not sufficient evidence to reject the null hypothesis. We can conclude there is no statistical difference between the lot 1 samples and the population mean, so we cannot reject the null hypothesis.

**Lot 2**Graphical user interface, text

Description automatically generated

The t-test of the PSI of Lot 2 and the population mean of 1,500 PSI resulted in a p-value of 0.6072, which is above the significance value of 0.05 There is not sufficient evidence to reject the null hypothesis and conclude that there is no statistical difference between Lot 2 samples and the population mean.

**Lot 3**

Graphical user interface, text

Description automatically generated

The t-test of the PSI of Lot 3 and the population mean of 1,500 PSI resulted in a p-value of 0.04168 which is below the significance level of 0.05. This indicates that there is sufficient evidence to reject the null hypothesis and we can state that there is a statistical difference between the Lot 3 samples and the population mean.

## Study Design: MechaCar vs Competition

This challenge delt with comparing MechaCar and parts which is useful in making manufacturing decision. In addition, it would be valuable to do a statistical analysis comparing MechaCar against competitors in the auto business for marketing decisions.

* What metric or metrics are you going to test?

The additional metrics to test comparing MechaCar with its competitors could be safety rating, highway fuel efficiency, and resale value.

* What is the null hypothesis or alternative hypothesis?

The null hypothesis for each of these would be that the variance for each metric is the same when compared to MechaCar.

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

Using a multiple linear regression statistical summary would show how the variables impact for each of the variable. Then a simple T test to compare the competitor’s statistical output to see if there see if we can reject the null hypothesis.

* What data is needed to run the statistical test?

For a reliable statistical analysis, we would need a random sample of 50 or more events for MechaCar and each of their competitor, reporting a numerical value for each of the variables we want to compare.