This challenge was to provide insights to a manufacturing team about production data for a new prototype known as MechaCar that was experiencing production problems and has delayed its progress at AutosRUs. To execute this analysis we will do the following:

1. Perform multiple linear regression analysis to determine the variables in the dataset that would predict the MechaCar prototypes mpg.
2. Collect summary statistic on pounds per square inch (PSI) of the suspension coils from the manufacturing lots.
3. Run T- tests to determine if the manufacturing lots are statistically different from the mean population.
4. Design a statistical study to compare MechaCar vehicles performance with vehicles from other manufacturers. And write a summary of each statistical analysis with the interpretation of the findings.

## Linear Regression to Predict MPG

The Linear Regression conducted on a MechaCar dataset with a sample size of 50 prototypes produced the follow results.

Graphical user interface, text, application, email

Description automatically generated

According to the findings the r-squared value from the linear regression model is 0.711, that gives a confidence that this model has a 71% accuracy. The p-value of the linear regression

analysis is 6.712e-11, this is less than the significance level of 0.05%. Hence there is sufficient evidence to reject the null hypothesis and the slope of this model is not equal to zero.

## Summary Statistics on Suspension Coils

The Summary Statistics for the Manufacturing Lot produced below displays the mean is 1498.78 and the population mean for the sample was 1500.

Table

Description automatically generated

The different Manufacturing Lots’ means are similar to that of the sample mean and population mean.

Table

Description automatically generated

Based on the required design specification for the Suspension Coils mandate to not exceed 100 pounds per square inch (PSI). The Total Manufacturing Lot variance is 62.29 < 100 and this is approximately the stated requirement. Upon reviewing by lots and performing a deeper analysis, Lot 1 had a low variance of 1 and Lot 2 a variance of 7. But Lot 3 had a significantly high variance of 170 > 100 which did not meet the design specifications.

## T-Tests on Suspension Coils

All the Manufacturing Lots: p-value = .60. This means the total manufacturing lot is not statistically significant from the normal distribution and is exemplified as .06 > .05. The mean is within the 95% confidence interval.

Graphical user interface, text, application

Description automatically generated

Lot 1 has the sample mean of 1500 and a p-value of 1. Hence, we cannot reject the null hypothesis that there is no difference between the sample mean and Population mean.

Text

Description automatically generated

Lot 2 had very similar outcome with a sample mean of 1500.02 and a p-value of 0.61. Hence the null cannot be rejected similarly to Lot 1. The sample mean and the population mean are very close to 1500.

Text

Description automatically generated

Lot 3 produced a sample mean of 1496.14 and a p-value of 0.04. This was less than the significance level of 0.05. Hence the null hypothesis should be rejected. The sample mean and the population are not statistically different.

Text, letter

Description automatically generated

## Study Design: MechaCar vs Competition

What metric or metrics are you going to test?

The metrics would be Vehicle Weight comparative to the Safety Ratings test as this would be a competitive advantage for MechaCar.

What is the null hypothesis or the alternative hypothesis

The null represents the possibility that the results are a random probability and not influenced by any outside factors or other variables. The alternate Hypothesis represents the contrast of the above stated being, any other scenario yielded from the results.

H0: PH = 0.1 Ha: PH ≠ 0.1

The PH represents the safety ratings to vehicle weight

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

The two-tailed test would help to describe both sides of distribution and utilized descriptions namely, “equal to” or “not equal to”.

If both the Hypothesis and statistical test are 2 two tailed, then the statistical test p-value may be used as is. If the statistical test is two-tailed and the hypothesis is one-tailed, then divide the statistical test p-value by 2.

What data is needed to run the statistical test?

The sample of data must be representative of the number of and types of Mecha Cars for observing the relationship between accidents and weights. To minimize type I error (reject the null hypothesis when it is actually true), decrease the P to 0.1 to limit type II error (fail to reject the null). Then include other variables for measurement namely; weather, maintenance, horsepower and terrain.