**MechaCar Statistical Analysis**

**Overview**

AutosRUS’ newest prototype, the MechaCar is suffering from production issues that are blocking the manufacturing teams’ progress. In this analysis, we are performing data analytics on the production data that could provide insights to the manufacturing team. In this analysis, we will perform linear regression analysis to predict the MPG, perform Summary Statistics on the Suspension Coils and perform T-Tests on Suspension Coils of all three manufacturing lots for MechaCar.

**Linear Regression to Predict MPG**

Using the MechaCar data file, we created linear regression, using the lm() function with the six variables provided in the data file. Then we determined the p-value and r-squared value for the linear regression model, using the summary() function. We used the following summary() formula to generate the desired results:

summary(lm(mpg ~ vehicle\_length + vehicle\_weight + spoiler\_angle + ground\_clearance + AWD, data=MechaCarData)) Graphical user interface, text, application

Description automatically generated

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

According to the screenshot above, the Vehicle Length and Ground Clearance variables are most likely to provide a non-random amount of variance to the mpg values in the dataset.

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

The p-value of our linear regression analysis is 5.35e-11, which is smaller than the assumed significance level of 0.05%. Therefore, we can conclude that there is sufficient evidence to reject our null hypothesis, which means the slope of our linear model is not zero.

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

Based on our linear regression analysis, the r-squared value of this analysis is 0.7149, approximately 71% of the mpg predictions in this data will be determined by this model. In a simple linear regression model, the higher the correlation between two variables, the more than one variable can explain/predict the value of the other.

Therefore, we can conclude that this linear model can effectively predict the mpg of MechaCar Prototypes.

**Summary Statistics on Suspension Coils**

To determine the Summary Statistics on Suspension Coils, we ran analysis on the total summary and lot summary of the suspension coils. To determine the total summary, we used the following function:

total\_summary <- suspension\_coil\_data %>% summarize(Mean=mean(PSI),Median=median(PSI),Variance=var(PSI),SD=sd(PSI), .groups = 'keep')

After running the above the function, we got the following result:

Table

Description automatically generated

According to the our result, the mean is 1498.78, the Median is 1500, the Variance is 62.29456 and Standard Deviation or SD is 7.892627

We used the following function to determine the lot summary:

lot\_summary <- suspension\_coil\_data %>% group\_by(Manufacturing\_Lot) %>% summarize(Mean=mean(PSI),Median=median(PSI),Variance=var(PSI),SD=sd(PSI), .groups = 'keep')

We got the following result after running the lot summary function:

Table

Description automatically generated

The design specifications for the MechaCar suspension coils dictate that the variance of the suspension coils must not exceed 100 pounds per square inch. Does the current manufacturing data meet this design specification for all manufacturing lots in total and each lot individually? Why or why not?

Based on our result, we can conclude that Lot 1 and 2 are within the desired specifications with 0.98 and 7.47 variance, however, Lot 3 does not meet specifications as it is above the desired specification of 100 pounds as Lot 3 variance is 170.79 approximately.

**T-Tests on Suspension Coils**

In this deliverable, we created a T-test that compares all the manufacturing lots against mean PSI (1,500) of the population. Then we created three T-Tests for each of the 3 manufacturing lots against mean PSI of the population.

We used the following function to create T-test for all the manufacturing lots:

t.test(suspension\_coil\_data$PSI,mu=1500)

Text

Description automatically generated

The p-value of all the manufacturing lots is 0.06 and has a mean of 1498.78 compared the presumed population mean of 1500

Then we created T-Tests for all three manufacturing lots using the formula above.

Text

Description automatically generated

The mean of Lot 1 sample is 1500, which is the same as the PSI mean. The p-value for Lot 1 is 1, therefore, we cannot reject the null hypothesis and must accept because there is no statistical difference between the observed sample mean and the presumed population mean

Text, letter

Description automatically generated

The result for Lot 2 is similar to that of Lot 1 with a sample mean of 1500.2 and a p-value of 0.61. The null hypothesis cannot be rejected as the two samples are statistically similar.

Text

Description automatically generated

The p-value for Lot 3 is 0.04 which is lower than the significance level of 0.05. Therefore, we can conclude that there is sufficient evidence to reject our null hypothesis.

**Study Design: MechaCar vs Competition**

At this point, we are creating a design to see how MechaCar performs against the competition. When shopping for cars, most consumers (myself included) consider the cost of the car, fuel efficiency, maintenance cost, horsepower, and safety rating before deciding on what to purchase. In this design, we are going to consider metrics we believe is most important to the average consumer to help determine how MechaCar compares to its competitors. We will use the following metrics to perform our statistical analysis:

* Cost of MechaCar compared to its competition
* MPG/Fuel efficiency
* Safety Rating

**Null Hypothesis:**

MechaCar is more fuel efficient than its competition and there is no difference in the cost compared to its competition

**Alternative Hypothesis:**

MechaCar consumes more fuel than its competition and is costlier than its competition.

**Statistical Testing:**

To test our hypothesis, we would use the Two-Sample t-Test as it is the best test to perform our analysis. Two Sample T-test analyses if there is a statistical difference between the distribution means from two samples. With this Statistical testing type, we can compare the price and fuel efficiency of MechaCar and its competition.