**MechaCar\_Statistical\_Analysis**

**Summary**

AutosRUs’ newest prototype, the MechaCar, is suffering from production troubles that are blocking the manufacturing team’s progress. AutosRUs’ upper management has requested a review of the production data for insights that may help the manufacturing team.

**Deliverable 1: Linear Regression to Predict MPG**

Text

Description automatically generated

Figure - Mechacar Statistical Analysis

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

Vehicle length and vehicle ground clearance have significantly lower than normal p-values which would indicating that variances are most likely non-random and therefore will have the most significant impact on mpg values. Conversely, vehicle weight, AWD and spoiler angle all have p-values greater than .05, indicating that any variances are more likely random.

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

The p-Value for this model, p-Value: 5.35e-11, is much smaller than the assumed significance level of 0.05%. This indicates there is sufficient evidence to **reject our null hypothesis**, which further indicates that the slope of this linear model is **not zero**.

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

Our R-squared value is .7149 (adjusted at .6825), which means roughly 71% (68% adjusted) of the time the model will predict mpg values correctly. There may be other factors that were not captured in the datasaet that could contribute to the mpg variability of the MechaCar prototypes.

**Deliverable 2: Summary Statistics on Suspension Coils**

Graphical user interface, text

Description automatically generated

Figure - Suspension Total PSI All Lots



Figure - Suspension PSI Summary by Lot

1. 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?

When analyzing the total lots, it does appear that the Variance is within margin, with the Median and Mean being very close to the design specifications. However, when you dive into the lot breakdown, it becomes apparent that there is a problem. While Lot 1 and Lot 2 are both within design specifications and have nearly the same exact mean and median with their Variance and Standard Deviation being well withing margin, Lot 3 shows the most variance and exceeds the manufacturers specs. The variance within lot 3 should be investigated.

**Deliverable 3: T-Tests on Suspension Coils**

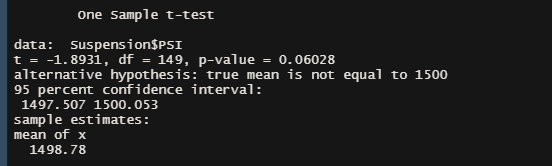
****

Figure - Suspension Coil Statistics - All Lots

1. Looking at all Lots in aggregate, with a p-Value of 0.06, which is higher than the common significance level of 0.05, there is NOT enough evidence to support rejecting the null hypothesis. That is to say, the mean of all three of these manufacturing lots combined (1498.78) is statistically similar to the presumed population mean of 1500.

Text

Description automatically generated

Figure - Suspension Coil Statistics - Lot 1

1. Lot 1 sample contains the true sample mean of 1500. With a p-Value of 1, clearly we cannot reject the null hypothesis that there is no statistical difference between the observed sample mean and the presumed population mean (1500).

Text

Description automatically generated

Figure - Suspension Coil Statistics - Lot 2

1. As with Lot 1, Lot 2 has essentially the same outcome with a sample mean of 1500.02 and a p-Value of 0.61. Therefore it still holds true that the null hypothesis cannot be rejected, and the sample mean and the population mean of 1500 are statistically similar.

Text

Description automatically generated

Figure - Suspension Coil Statistics - Lot 3

1. As we saw with the PSI statistics earlier, Lot 3 is problematic. Here the sample mean is 1496.14 and the p-Value is 0.04, which is lower than the common significance level of 0.05. In this case, we need to reject the null hypothesis that this sample mean and the presumed population mean are not statistically different.

Looking at the aggregate data for both the PSI figures and T Tests on the individual lots, it is clear that there is a problem with Lot 3 and that their production cycle needs to be investigated to determine why this particular lot is experiencing the variances that are not present at the other lots.

**Deliverable 4: Study Design: MechaCar vs Competition**

Metric: 5 year Cost vs Fuel Efficiency

Determine whether the ownership cost of MechaCar is competitive based on car fuel efficiency. The study will examine the estimated cost of the MechaCar vs its competitors over the course of the first 5 years of ownership and whether that is reflective within the vehicle’s MPG range.

Hypothesis: Null and Alternative

After determining the 5 year cost for MechaCar and its competitors:

• Null Hypothesis: MechaCar is cost effective based on its MPG performance

• Alternative Hypothesis (Ha): MechaCar is NOT cost effective based on its MPG performance

Statistical Test

After collating the list of competitors and estimating their combined retail and 5 year maintenance costs, we identify the mean value and use that to do a basic p-value analysis to determine whether MechaCar is priced appropriately. Further breakdown can be performed by filtering on Engine Fuel Type to determine whether the hypothesis holds up within a more direct comparison model.

Data Required (last 5 years)

* Competitors – Similar makes and models
* Engine Fuel Type – Diesel, Gas, Hybrid, Electric
* Average MPG
* Retail Value
* Average Annual Maintenance Cost