**Multiple Linear Regression – MechaCar Prototypes**

**Background**

The MechaCar\_mpg.csv data file provides a collection of performance related data for 50 potential prototype MechaCars. The performance metrics supplied are Vehicle Length, Vehicle Weight, Spoiler Angle, Ground Clearance, AWD and MPG as revealed by the table column headers.

In this analysis, Vehicle.length, Vehicle Weight, ground.clearance, AWD and MPG are the independent variables of interest for our Multiple Linear Model. In the R system, we applied the lm() function with the four variables as input. Also, calculated the summary statistics for MechaCar as indicated.

**Hypotheses**

H0: The slope of the linear model is zero, or m = 0 Ha: The slope of the linear model is not zero, or m ≠ 0

**Observation**

1. The summary output of this Linear Model shows that the **Pr(>|t|)** values of some of the independent variables such as Ground.clearence, Spoiler.angle and the Intercept are statistically unlikely to contribute random amounts of variance to the mpg values.
2. The slope of this Linear model is not considered to be zero based on the result of R-Squared and P-value. There is sufficient evidence to reject the null hypothesis.

Multiple R-squared: 0.0942 and the P-value: 0.4801.

1. The Multiple Linear model predicts the mpg of the MechaCar prototypes effectively because the model applied multiple independent variables in the prediction.

Vehicle.weight and AWD are more likely to contribute to the random variance of mpg.

**Population sample size and variables**

The result shows that there are other variables and factors that may contribute to the variation which are not included in the supplied data, and therefore not in the Linear Model.

Larger sample size and more test will yield a reliable and more realistic output to support predictions. Considering the fact that the product is at a prototype stage, the data works well enough for the findings.

summary(MechaCar)

vehicle.length vehicle.weight spoiler.angle ground.clearance AWD mpg

Min. :12.00 Min. : 2000 Min. : 0.00 Min. : 6.00 Min. :0.0 Min. :10.00

1st Qu.:13.55 1st Qu.: 5038 1st Qu.:46.82 1st Qu.:11.09 1st Qu.:0.0 1st Qu.:33.51

Median :14.60 Median : 5929 Median :58.55 Median :12.98 Median :0.5 Median :43.40

Mean :15.02 Mean : 6154 Mean :57.12 Mean :12.71 Mean :0.5 Mean :45.13

3rd Qu.:16.00 3rd Qu.: 7505 3rd Qu.:68.82 3rd Qu.:14.50 3rd Qu.:1.0 3rd Qu.:54.53

Max. :20.00 Max. :10000 Max. :90.00 Max. :18.00 Max. :1.0 Max. :80.00

**Suspension Coil-MechaCar Prototypes**

**Background**

The Suspension Coil data provides measures of independent variable for the design of coil system. The only quantifiable variable available for this analysis is the PSI.

Design specification for the variance of the Suspension Coil is set at 100 Pounds-per-inch maximum. Our analysis revealed that the measure of the average of the spread or the squared deviation of the PSI from its mean of 1500 is at about 76.23436.

It is reasonable to conclude that the current manufacturing data of less than or equal to 100 is within design specification and therefore meet specified design requirement.

**Statistical Summary Table**

|  |  |  |  |
| --- | --- | --- | --- |
|  | VehicleID | Manufacturing\_Lot | PSI |
| nbr.val | NA | NA | 1.500000e+02 |
| nbr.null | NA | NA | 0.000000e+00 |
| nbr.na | NA | NA | 0.000000e+00 |
| min | NA | NA | 1.463064e+03 |
| max | NA | NA | 1.535568e+03 |
| range | NA | NA | 7.250402e+01 |
| sum | NA | NA | 2.249297e+05 |
| median | NA | NA | 1.499747e+03 |
| mean | NA | NA | 1.499531e+03 |
| SE.mean | NA | NA | 7.129030e-01 |
| CI.mean | NA | NA | 1.408706e+00 |
| var | NA | NA | 7.623459e+01 |
| Std.dev | NA | NA | 8.731242e+00 |
| Coef.var | NA | NA | 5.822649e-03 |

**Suspension Coil T-Test**

A One Sample t-test is applied in this analysis since we are limited to one independent variable and one sample. Population mean provided is 1500.

**Hypotheses** H0: There is no statistical difference between the observed sample mean and its presumed population mean. Ha: There is a statistical difference between the observed sample mean and its presumed population mean.

**One Sample t-test**

data: log10(Suspension\_Coil$PSI) t = -7247086, df = 149, p-value < 2.2e-16 alternative hypothesis: true mean is not equal to 1500 95 percent confidence interval: 3.175540 3.176356 sample estimates: mean of x 3.175948

**Observation**

In comparing the calculated P-value of < 2.2e-16 to the significance level interval at 3.175540 3.176356, which is significantly lower, there is enough evidence to reject the null hypothesis and support the fact that there is a statistical difference between the observed sample mean and its population mean.

**Study of competitive comparisons**

Developing a broad metrics is a vital component in creating a successful product. Crucial metrics offer a solid path to decision making that leads to a widely acceptable and competitive product.

Metrics offer three major benefits or advantages to product design process: 1) Shed light on the past performance of the product 2) Reveal the existing situation of the product 3) Provide motivation for future decisions and actions about the product

Among automotive product design metrics of interest to consumers are: Cost, Interior Color option, Exterior Color option, Fuel Efficiency, Vehicle Size, Vehicle Height, Interior Spacing, upholstery Material Quality, Interior Spacing, Dashboard Design,, Engine Horsepower/Displacement, Cargo Space, Wheel Design, Vehicle Durability, Safety/Impact Test, general Curb Appeal and Number of cars produced.

With respect to the design of MechaCar, more metrics are needed in order to improve the design and competitive level of the product. By offering more metrics, we can apply more robust Linear Model and test methodologies that incorporate or allow multiple variables thereby yielding more reliable predictions.

**Questions to ask**

Expanding the population size will increase the choice and availability of samples. The most important question will always surround the variability of correlated variables. In addition, the chances of predicting values of one variable looking at the outcome of another.

**Type of test to apply**

Multiple Linear Regression is ideal for predictive modeling along with quantifiable measurements.