**MechaCar\_Statistical\_Analysis**

**Overview**

AutosRUs’ newest prototype, the MechaCar, is suffering from production troubles that are blocking the manufacturing team’s progress. In this project, we are going to review the production data by completing the following tasks:

* Perform multiple linear regression analysis to identify which variables in the dataset predict the mpg of MechaCar prototypes.
* Collect summary statistics on the pounds per square inch (PSI) of the suspension coils from the manufacturing lots.
* Run t-tests to determine if the manufacturing lots are statistically different from the mean population.
* Design a statistical study to compare vehicle performance of the MechaCar vehicles against vehicles from other manufacturers.

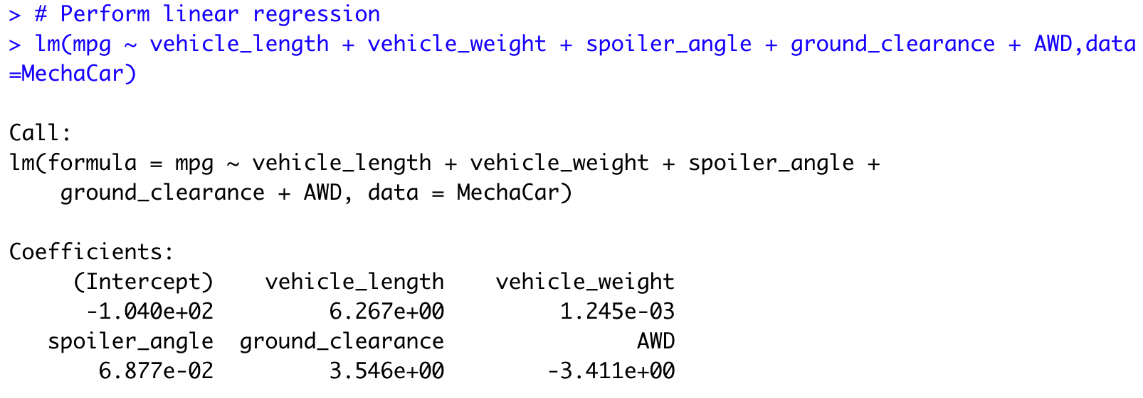
**Resources**

* Data Source: MechaCar\_mpg.csv - Suspension\_Coil.csv
* Software: R 4.2.2 - RStudio 2022.07.2

**Linear Regression to Predict MPG**

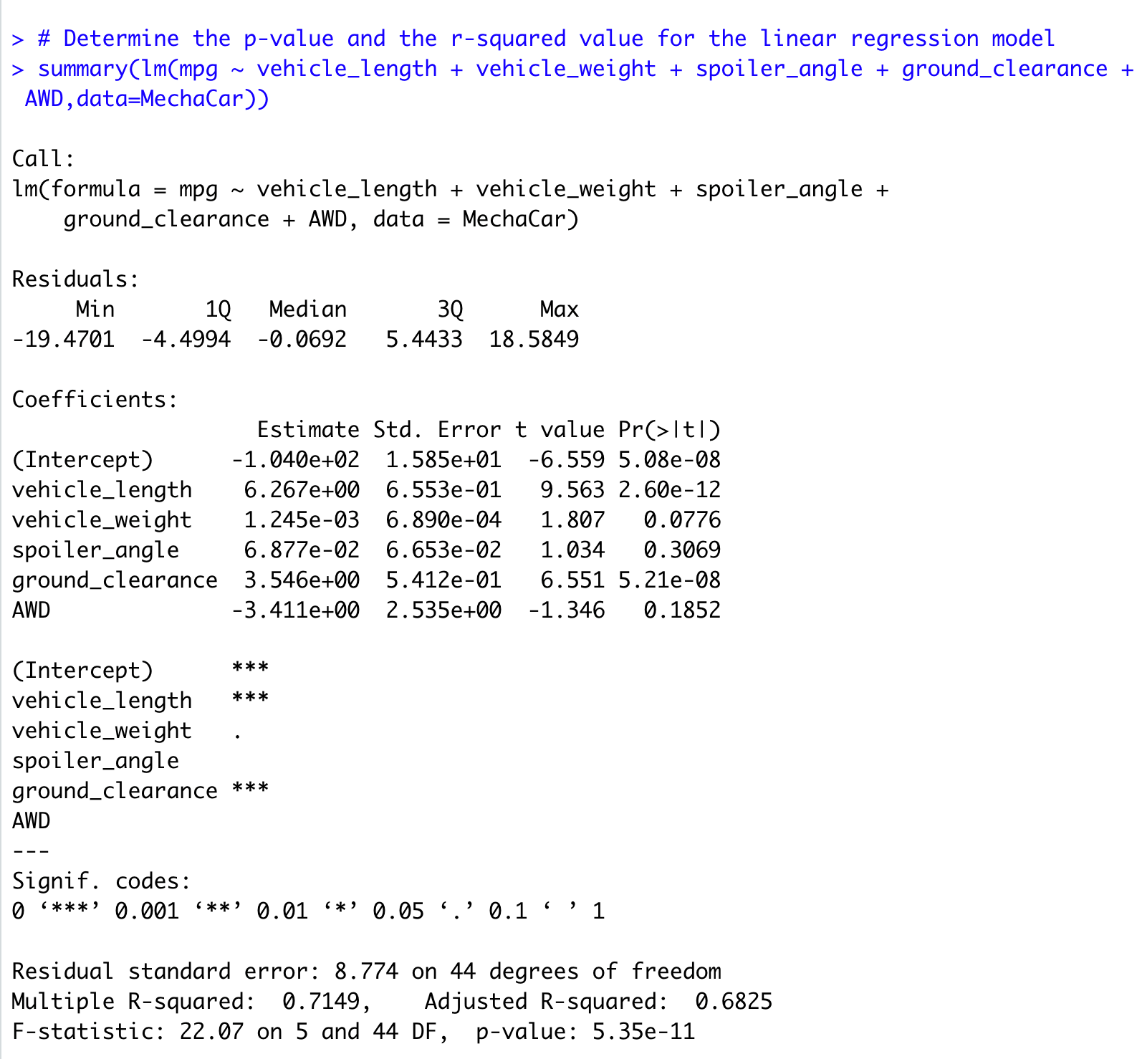
The MechaCar\_mpg.csv dataset contains mpg test results for 50 prototype MechaCars. The MechaCar prototypes were produced using multiple design specifications to identify ideal vehicle performance. Multiple metrics, such as vehicle length, vehicle weight, spoiler angle, drivetrain, and ground clearance, were collected for each vehicle.

Using lm() we have created a multiple linear regression model and in order to determine the p-value and r-squared value we have used the summary() function. (See Fig1. and Fig2)

[](https://user-images.githubusercontent.com/109363759/206569339-d48a27a8-2631-4142-81a7-1a157d5ced47.png)

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

Figure 1 shows that all independent variables have a slope different than zero. However, one can notice that the slope of “vehicle length” and “ground clearance” are significantly bigger than other slopes while the slope of AWD is negative. Which indicates that “vehicle length” and “ground clearance” accounts more than other independent variable for the the total variance in the dependent variable “mpg”.

[](https://user-images.githubusercontent.com/109363759/206567174-8cb504e0-7dbb-4010-bec7-9989d8365727.png)

This could be confirmed by looking at the values of Pr(>|t|) in Figure 2:

* The p-value for “vehicle length” is 2.60e-12 while it is 5.21e-08 for “ground clearance” which are significantly less than the assumed significance level of 5%. Thus we can say that these coefficients have a significant impact on the mpg values of the dataset.
* The p-value for “vehicle\_weight” is 0.0776, for “spoiler\_angle” it is 0.3069 and 0.1852 for “AWD”. All these values are bigger than the assumed significance level of 5%. Therefore, theses variables do not have a statistically significant relationship with the response variable in the model.

**Is the slope of the linear model considered to be zero? Why or why not?**  
From figure 1 we can see that the slopes of each coefficients are different than zero. This is not enough to say that the slope of the linear model is not zero since we have one negative slope. However, the p-value (Fig.2) can confirm that since it’s equal 5.35e-11, which is much smaller than our assumed significance level of 0.05%. Therefore, we can state that there is sufficient evidence to reject our null hypothesis, which means that the slope of our linear model is not zero.

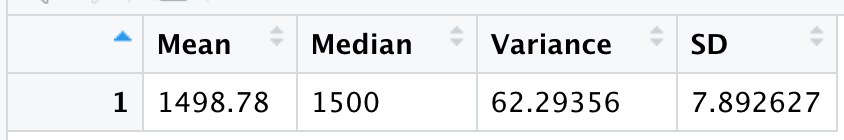
**Does this linear model predict mpg of MechaCar prototypes effectively? Why or why not?**  
The r-squared value for this linear regression model is equal 0.7149 (Fig.2), which means that this linear model will predict mpg MechaCar prototypes to 70% accuracy. In other words, 70% of all predictions using this model will be correct.

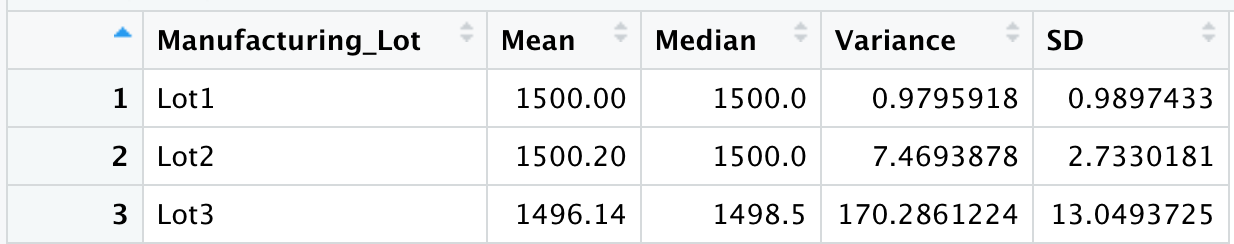
**Summary Statistics on Suspension Coils**

The MechaCar Suspension\_Coil.csv dataset contains the results from multiple production lots. In this dataset, the weight capacities of multiple suspension coils were tested to determine if the manufacturing process is consistent across production lots.

We have created a summary statistics table to show:

* The suspension coil’s PSI continuous variable across all manufacturing lots (Fig.3)
* The following PSI metrics for each lot: mean, median, variance, and standard deviation. (Fig4)

[](https://user-images.githubusercontent.com/109363759/206567269-921e3d92-ec72-4692-b1dd-c141038ee249.png)

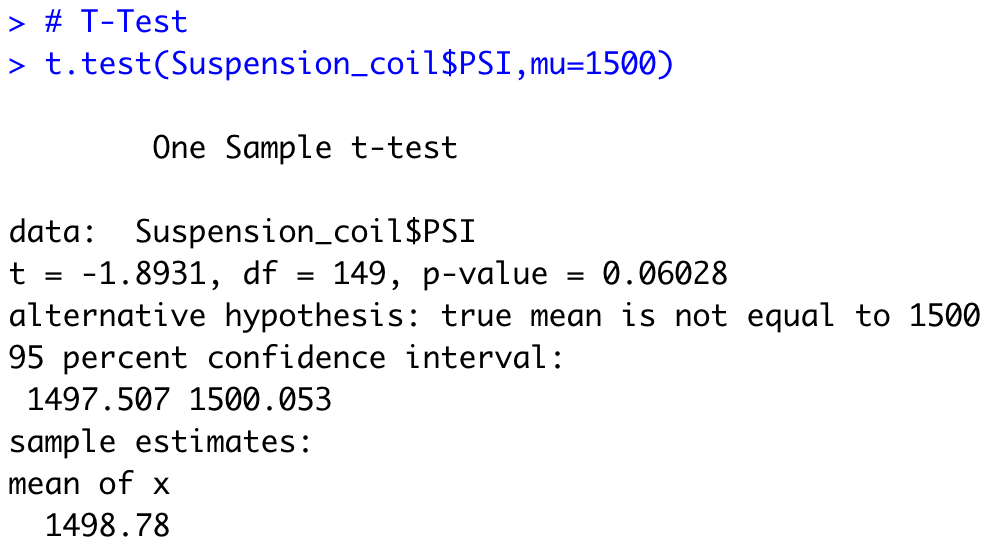
[](https://user-images.githubusercontent.com/109363759/206567321-050272aa-3fc8-4f0b-9947-d2511b6581ba.png)

**The design specifications for the MechaCar suspension coils dictate that the variance of the suspension coils must not exceed 100 pounds per square inch.**

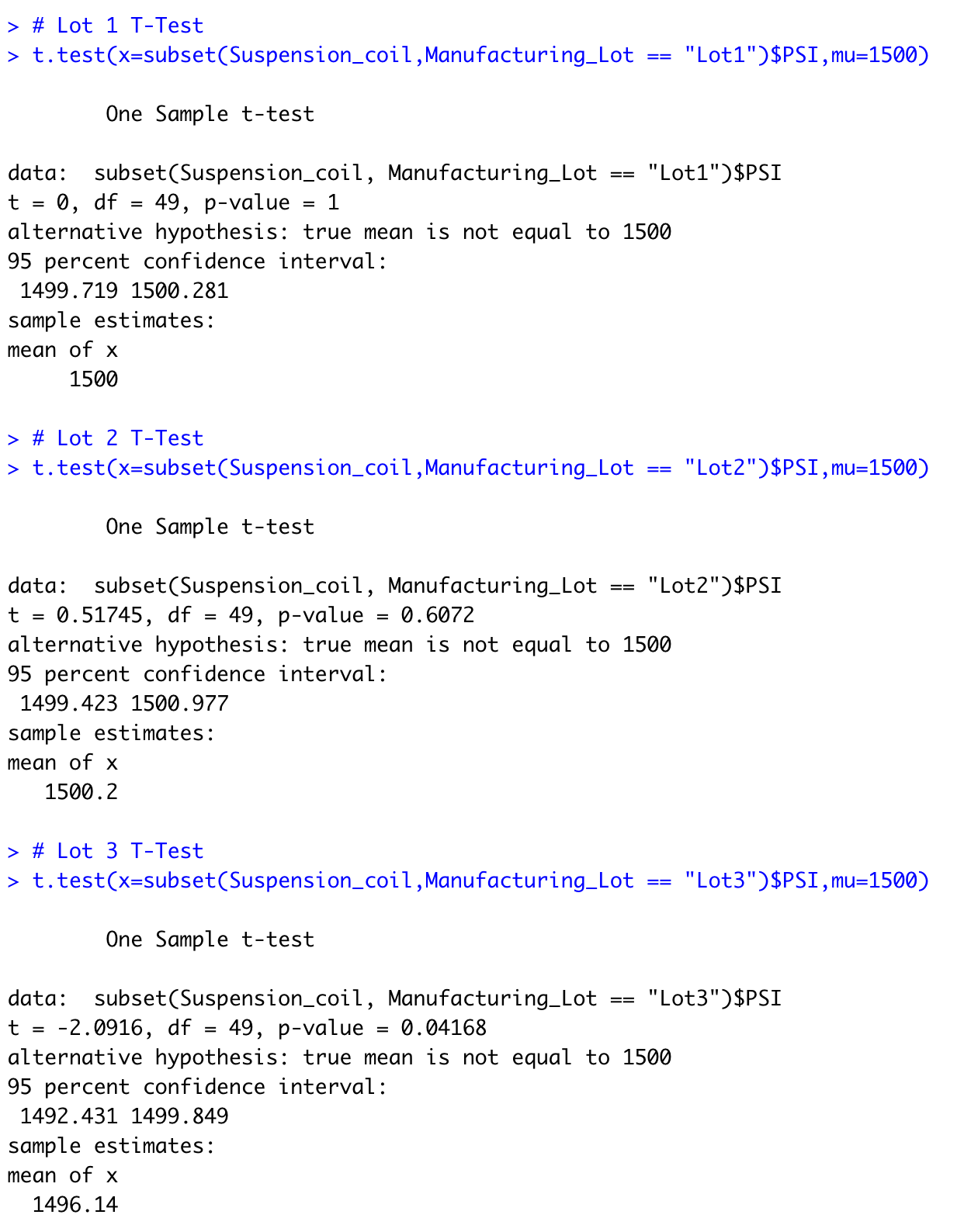
By looking at the summary statistics across all manufacturing lots (Fig.3), the variance is 62.29356 < 100 which meets the design specification. Therefore, the current manufacturing data meet this design specification for all manufacturing lots in total.  
However, the summary statistics for each lot (Fig.4) shows that variance of the suspension coils in lot 3 is 170.286 which is significantly higher than 100 and definitely not meeting the design specification.

**T-Tests on Suspension Coils**

In this part we have performed a t-test to determine if the PSI across all manufacturing lots is statistically different from the population mean of 1,500 pounds per square inch (Fig. 5) and 3 other t-tests to determine if the PSI for each manufacturing lot is statistically different from the population mean of 1,500 pounds per square inch (Fig.6)

[](https://user-images.githubusercontent.com/109363759/206567422-bb6f32e0-35b3-4c4f-bfd8-68c39c33974f.png)

The p-value of this t-test is equal 0.06 which is above the assumed significance level (5%). Therefore, we would state that there is not enough evidence to reject the null hypothesis and there is no overall statistical difference between the population mean of 1500 pounds per square inch and the PSI across all manufacturing.

[](https://user-images.githubusercontent.com/109363759/206571570-fb04d648-e9b2-4a50-97dc-f510cdcda7c2.png)

* For lot 1 and 2, the p-value are equal 1 and 0.6072 consequently. Both values are above the assumed significance level of 0.05. Therefore, there is not enough evidence to reject the null hypothesis and there is no overall statistical difference between the PSI across lot 1 and 2 and that of population mean value.
* However, the p-value for lot 3 is equal 0.04168 < 0.05. Which is sufficient to reject the null hypothesis and say that there is a statistical difference between the population mean of 1500 pounds per square inch and the PSI across manufacturing lot 3.

**Study Design: MechaCar vs Competition**

In this section we are going to design a statistical study to compare performance of the MechaCar vehicles against performance of vehicles from other manufacturers.

We are going to test the correlation between the existence of active driver assistance technologies in cars and the safety of passenger.

**Metrics to test**

* 5 Star Safety Ratings: The National Highway Traffic Safety Administration’s New Car Assessment Program (NCAP) created the 5-Star Safety Ratings Program to provide consumers with information about the crash protection and rollover safety of new vehicles beyond what is required by Federal law. One star is the lowest rating; five stars is the highest. More stars equal safer cars.
* Driver assistance technology options: Forward collision warning - Lane Departure Warning - Automatic Emergency Braking - Rearview video system - Blind Spot Warning - Pedestrian Automatic Emergency Braking…

**Null and alternative hypothesis**

* Null Hypothesis : Cars with active safety technologies are not safer than other cars.
* Adaptive Hypothesis: Cars with more active driver assistance technology options will have higher safety rating.

**Statistical test chosen to test the hypothesis**

* Chi-Squared Test: This test can be used to determine if there is a difference in categorical frequencies between groups. For example: How many vehicles with Automatic Emergency Braking have a safety level 5? Vs How many vehicles with no Automatic Emergency Braking have a safety level 5?
* It the p-value of our Chi-Squared Test is less than the assumed significance level of 0.05, then we can reject the null hypothesis and say that vehicles with active Automatic Emergency Braking are safer.

**Data needed to run the statistical test**

* Vehicles data for all models (including the MechaCar) containing the driver assistance technology options available in each model.
* The 5 star safety rating data for all vehicles.