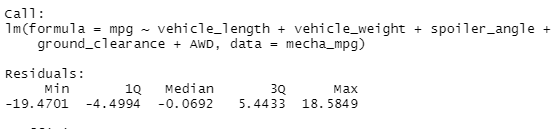
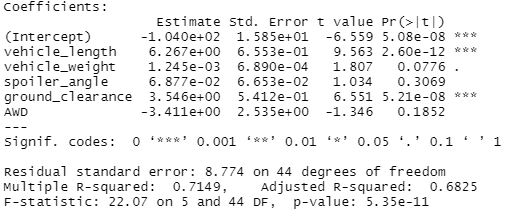
**MechaCar\_Statistical\_Analysis**

**Deliverable 1**

**Linear Regression to Predict MPG**

[](https://user-images.githubusercontent.com/104606589/189550747-525d078f-e87a-47a2-9fdf-e1a3a70eed50.png)

**Summary Statistics for Linear Regression model**

[](https://user-images.githubusercontent.com/104606589/189550775-ce4ef373-db45-4e05-b8d0-fa5f3f149bba.png)

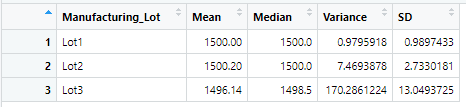
* When looking at the summary you can see that ground clearance and vehicle length have the highest likely hood of impacting the total MPG. While All Wheel Drive, Spoiler Angle, and Vehicle Weight does not seem to be as good of a predictor for the MPG of the vehicle.
* We have an assumed significance of 0.05% which is much higher than the P-Value of 5.35e-11 that is being returned. This would lead us to reject the null hypothesis, which infers that the slope of this linear model in not zero.
* With an r-squared value of 0.7149 we can be assured that this model does predict MPG reliably.

**Summary Statistics on Suspension Coils**

* Below is a summary of the total population of coils tested. This summary is not stratified in any way. It shows the values observed when looking at the largest population of samples. The mean and median are very close together, giving us the impression that the data is normally distributed.

[totals_summary](https://user-images.githubusercontent.com/104606589/189551215-8607ec70-bc2f-49fe-b357-584bfa9511ee.png)

* Below is a summary of the same data, broken up by lot.

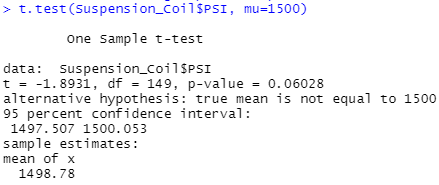
[](https://user-images.githubusercontent.com/104606589/189551220-1a30639d-2a1c-4721-861f-cc80c8a01267.png)

* When looking at the total population the minimum variance of 100 PSI has been met. When looking deeper into the sample it looks like Lot 3 does not meet the minimum requirement, and there would be justification to research what happened with this specific lot. The extreme variance in lot 3 is hidden when only looking at the total population.

**T-Test on Suspension Coils**

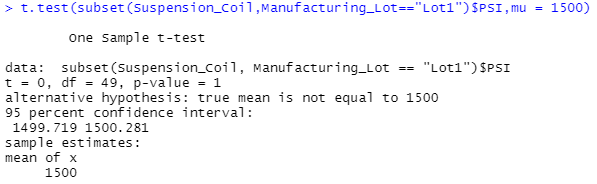
A t-test was conducted on the population as a whole and on to the individual lots. This was done to determine whether the population of samples was significantly different than the 1500 pounds per square inch requirement by the company.

When looking at all units sampled, we saw a sample mean of 1,498 and a p-value of .06. This sample mean is well within the threshold required by the company. The p-value of 0.06 is higher than the assumed statistical significance of 0.05. This will lead us to fail to reject the null hypothesis. Meaning that the average (mean) PSI across all manufacturing lots when looked at as a whole are statistically similar to the population mean of 1,500.

[](https://user-images.githubusercontent.com/104606589/189551623-1fcd921f-031a-47df-a1cd-2f64788b8dd0.png)

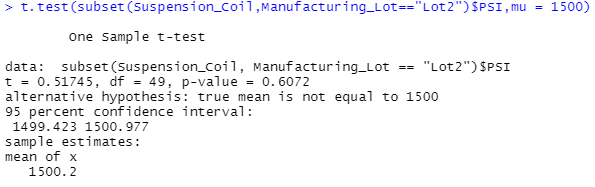
The following three t-tests show how each lot performed.

**Lot 1 T-Test**

[](https://user-images.githubusercontent.com/104606589/189551626-cfce176f-74bb-47a9-b838-5462ef0e9359.png)

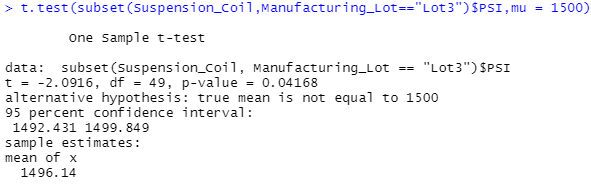
Lot 1 had a sample mean of 1,500 a p-value of 1. This means there is no statistical difference between Lot 1 and the greater population. We would fail to reject the null hypothesis for lot 1.

**Lot 2 T-Test**

[](https://user-images.githubusercontent.com/104606589/189551628-35cccd7e-77d9-4c52-908f-bd59045ab0df.png)

Lot 2 had a sample mean of 1,500.2 and a p-value of 0.61. With the p-value so much greater than the assumed significance level of 0.05, we will again fail to reject the null hypothesis.

**Lot 3 T-Test**

[](https://user-images.githubusercontent.com/104606589/189551629-69575736-51c3-4c59-bfb2-cbfc9f33e186.png)

Lot 3 had a sample mean of 1,496.14 and a p-value of 0.04. This p-value is below the assumed statistical significance of 0.05. We would be inclined to reject the null hypothesis and deem this sample to be statistically different then the greater population of coils.

**Study Design: MechaCar vs Competition**

A lot of people will look at the total cost of the car and not consider what the total cost will be for the life of the vehicle. The second value has a lot more to do with what the cost to maintain the vehicle will be compared to cheaper or more expensive vehicles.

To test this, we could compare the average maintenance cost for MechaCar vs its closest 2 competitors considering how different metrics (horsepower, weight, mpg, clearance, and AWD). The closest 2 competitors will be determined by total units (cars) sold in a fiscal year.

H0(Null Hypothesis): Based off the performance metrics of horsepower, weight, mpg, clearance, and AWD the average cost of maintenance for a car from MechaCars is no different than the cost of its 2 closest competitors.

Ha(Alternative Hypothesis): Based off the performance metrics of horsepower, weight, mpg, clearance, and AWD the average cost of maintenance for a car from MechaCars is different than the cost of its 2 closest competitors.

To test this, we would need to collect repair records as well as build specs for vehicles from the two competing car companies. This data may be available to the public through Carfax or other consumer protection/facing companies. We could conduct a linear regression on each companies’ vehicles to build a prediction of total maintenance cost of vehicle based the combination of performance metrics.