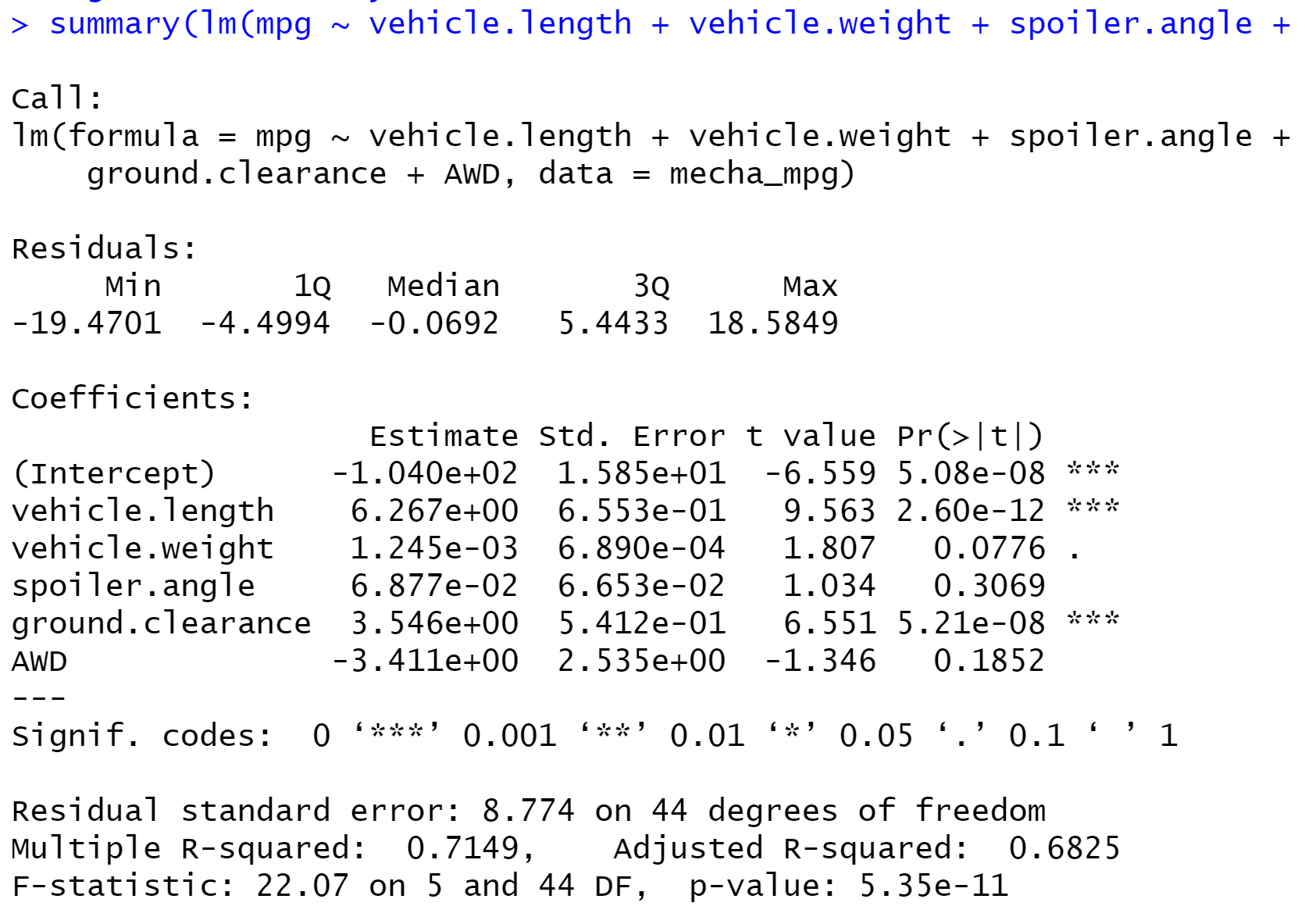
CHALLENGE

**MPG Regression—Interpretation of multiple linear regression results:**

To better predict the mpg dependent variable, we inspect how different vehicle length, vehicle weight, spoiler angle, ground clearance, and drivetrain contribute to mpg using multiple linear regression

* **Which variables/coefficients provided a non-random amount of variance to the mpg values in the dataset?** To determine which variables provide a significant contribution to the linear model, we look at the individual variable p-values. As shown in summary output, the Pr(>|t|) value represents the probability that each coefficient contributes a random amount of variance to the linear model. According to the results, vehicle length, ground clearance and intercept are unlikely to provide random amounts of variance to the linear model



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

No, the slope of the linear model is not 0. because the p-value (5.35e-11) is much smaller than assumed significance level of 0.05%. Therefore, there is sufficient evidence to reject the null hypothesis, which means that the slope linear model is not 0.

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

Yes. Multiple R-squared value (0.7149) is also known as the coefficient of determination and represents how well the regression model approximates real-world data points. In this case R-squared value is relatively high, thus we can conclude that the model predict mpg of MechaCar prototypes effectively

**Suspension Coil Summary**

By using groupby() and summarise() function within a dplyr pipe, we can create a summary statistics table for the suspension coil’s PSI produced in 3 different lots as shown in (insert table)

A screenshot of a cell phone

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 inch. **Does the current manufacturing data meet this design specification? Why or why not?** The variance of suspension coils is relatively small for Lot 1 and 2 (0.98 and 7.47) so the current manufacturing data meet design specification for Lot 1 and Lot 2. In Lot 3, the variance is 170.29, higher than 100, thus the manufacturing data does not meet design specification for Lot 3.

**Suspension Coil T-Test**

**Determine if the suspension coil’s pound-per-inch results are statistically different from the mean population.** As the mean of population distribution is known (1,500 PSI), we run independent one sample t-Test for the 3 Lot in this case. First, define the hypotheses:

Null hypothesis: There is no statistical difference between the observed sample mean and its presumed population mean.

Alternative hypothesis: There is a statistical difference between the observed sample mean and its presumed population mean.

To determine whether there is a statistical difference between the mean of sample distribution and population mean, we compare p-value with our assumed significance level 0.05%.

* Test 1, p-value=1 > 0.05, we do not have sufficient evidence to reject the null hypothesis. As a result, the two means are statistically similar
* Test 2, p-value=0.6072 > 0.05, we do not have sufficient evidence to reject the null hypothesis. As a result, the two means are statistically similar
* Test 3, p-value=0.04168 < 0.05, we have sufficient evidence to reject the null hypothesis. As a result, the two means are statistically different

**Design Your Own Study**

To design a statistical study that compares the performance of the MechaCar prototype vehicle to other comparable vehicles on the market. These are the questions I need to answer:

**What metrics would be of interest to a consumer**

* Cost, fuel efficiency, color options, reliability, horsepower, interior and exterior design, depreciation in value…

**Determine what question we would ask, what the null and alternative hypothesis would be to answer that question, and what statistical test could be used to test this hypothesis.**

The question: Can we predict customer’s interest using a linear model and values of the metrics listed above? And which variables provide significant contribution to the model.

The statistic test to use: Multiple Linear Regression

Null hypothesis: The slope of the linear model is 0

Alternative hypothesis: The slope of the linear model is not 0

**Knowing what test should be used, what data should be collected?**

Cost: car price

Fuel efficiency: mile per gallon

Color options: number of color options

Reliability: Consumer Reports' reliability ratings or customer survey on a scale of 1 to 5

Horsepower: car horsepower

Design: customer survey on a scale of 1 to 5

Depreciation in value: depreciation after 5 years