Analysis of MechaCar Data

Technical Report

Background

The results of an mpg testing dataset of 50 potential 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.

MechaCar suspension coil test results from multiple production lots. In this dataset, the weight capacity of multiple suspension coils was tested to determine if the manufacturing process is consistent across lots.

Observations

Multi Linear Regression

Multi Linear Regression observations focused on below questions:

1. Which variables/coefficients provided a non-random amount of variance to the mpg values in the dataset?
2. Is the slope of the linear model considered to be zero? Why or why not?
3. Does this linear model predict mpg of MechaCar prototypes effectively? Why or why not?

In the summary output, each Pr(>|t|) value represents the probability that each coefficient contributes a random amount of variance to the linear model. According to our results, vehicle length and ground clearance (as well as intercept) are statistically unlikely to provide random amounts of variance to the linear model. In other words, the vehicle length and ground clearance have a significant impact on mpg. When an intercept is statistically significant, it means there are other variables and factors that contribute to the variation in mpg that have not been included in our model. These variables may or may not be within our dataset and may still need to be collected or observed.

Despite the number of significant variables, the multiple linear regression model outperformed the simple linear regression. According to the summary output, the r-squared value has increased from 0.37 in the simple linear regression model to 0.71 in our multiple linear regression model and the p-value changed from 2.636e-06 to 5.35e-11.

The slope of the linear model is not considered to be zero. Through this data, we fail to reject the null hypothesis based on the coefficient values, the dependent variable mpg has a linear relationship to few independent variables.

```
Call:
lm(formula = mpg ~ vehicle_length, data = MechaCar_table)

Residuals:
    Min      1Q  Median      3Q     Max
-26.303  -7.160  -1.231   9.374  26.670

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -25.0622    13.2960  -1.885   0.0655 .
vehicle_length   4.6733     0.8774   5.326 2.63e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.47 on 48 degrees of freedom
Multiple R-squared:  0.3715,    Adjusted R-squared:  0.3584
F-statistic: 28.37 on 1 and 48 DF,  p-value: 2.632e-06
```

Fig1: Linear Regression Summary

```
> # perform multiple linear regression
> mecha_lm <- lm(mpg ~ vehicle_length + vehicle_weight + spoiler_angle + ground_clearance + AWD,data=MechaCar_table)
> summary(mecha_lm)

Call:
lm(formula = mpg ~ vehicle_length + vehicle_weight + spoiler_angle +
    ground_clearance + AWD, data = MechaCar_table)

Residuals:
     Min       1Q   Median       3Q      Max
-19.4701  -4.4994  -0.0692   5.4433  18.5849

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -1.040e+02  1.585e+01  -6.559 5.08e-08 ***
vehicle_length   6.267e+00  6.553e-01   9.563 2.60e-12 ***
vehicle_weight   1.245e-03  6.890e-04   1.807   0.0776 .
spoiler_angle    6.877e-02  6.653e-02   1.034   0.3069
ground_clearance 3.546e+00  5.412e-01   6.551 5.21e-08 ***
AWD             -3.411e+00  2.535e+00  -1.346   0.1852
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.774 on 44 degrees of freedom
Multiple R-squared:  0.7149,    Adjusted R-squared:  0.6825
F-statistic: 22.07 on 5 and 44 DF,  p-value: 5.35e-11

> summary(mecha_lm)$r.squared
[1] 0.7149033
```

Fig2: Multiple Linear Regression

Suspension Coil Summary

Suspension Coil manufacturing data summary:

- Mean: **1499.531**
- Median: **1499.747**
- Variance: **76.234**
- Standard Deviation: **8.731**

```
> summary(suspension_data)
  VehicleID         Manufacturing_Lot       PSI
 Length:150         Length:150         Min.   :1463
 Class :character   Class :character   1st Qu.:1497
 Mode  :character   Mode  :character   Median :1500
                                       Mean   :1500
                                       3rd Qu.:1501
                                       Max.   :1536
>
```

Fig3: Summary on full suspension data

```
> # Perform Summary Analysis on suspension coil dataset
> # Since we have to perform two sample t-test to compare Mean values, group the data.
> lot_summary <- suspension_data %>% group_by(Manufacturing_Lot) %>%
+   summarise(Mean=mean(PSI),Median=median(PSI),Variance=var(PSI),SD=sd(PSI))
`summarise()` ungrouping output (override with `.groups` argument)
> lot_summary
# A tibble: 3 x 5
  Manufacturing_Lot  Mean Median Variance    SD
  <chr>             <dbl>  <dbl>    <dbl> <dbl>
1 Lot1              1500.  1500.     1.15  1.07
2 Lot2              1500.  1499.    10.1   3.18
3 Lot3              1499.  1498.   220.   14.8
> # Total Summary
> total_summary <- suspension_data %>%
+   summarise(Mean=mean(PSI),Median=median(PSI),Variance=var(PSI),SD=sd(PSI))
> total_summary
      Mean   Median Variance       SD
1 1499.531 1499.747 76.23459 8.731242
```

Fig4: Summary of the Total suspension data by grouping

Summary Interpretation:

The design specifications for the MechaCar suspension coils dictate that the variance of the suspension coils must not exceed 100 pounds per inch. Since suspension data variance (76.23) is under 100PSI, we can predict that the current manufacturing data meets the design specifications.

T-Test

Performing the t-test on the suspension coil data determines the suspension coil's pound-per-inch results are statistically different from the mean population results of 1,500 pounds per inch.

```
> # Across all Lots
> t.test(suspension_data$PSI,mu = 1500)

        One Sample t-test

data:  suspension_data$PSI
t = -0.65784, df = 149, p-value = 0.5117
alternative hypothesis: true mean is not equal to 1500
95 percent confidence interval:
 1498.122 1500.940
sample estimates:
mean of x
 1499.531
```

Fig5: T-test on suspension data

Here true mean is not equal to 1500PSI and there is 95% confidence interval.

Design Study:

Upper management is looking for your expertise and wants you to design a study that compares the performance of the MechaCar prototype vehicle to other comparable vehicles on the market. In the MechaCarWriteUp.txt text file, write a short description of a statistical study that can quantify how the MechaCar outperforms the competition. In your study design, be sure to write about the following considerations:

For further design study on MechaCar performance with competitors, we can perform statistical analysis on other customer focused aspects such as cost feature of the car and its significance relations with other variables: mpg, color/specific finish, hp. With more data, we can analyze if cost is predictive of the variables by performing Regression. We can even perform correlation to identify how variables are related to each other (positively/negatively). If you have diff makes of cars, we could perform t-test based on cost analysis between different makes to observe significance difference between variables. We may also focus on the sales data to identify MechaCar outperforms with competitors.