Filippo Riva
MT4420

Assignment 3

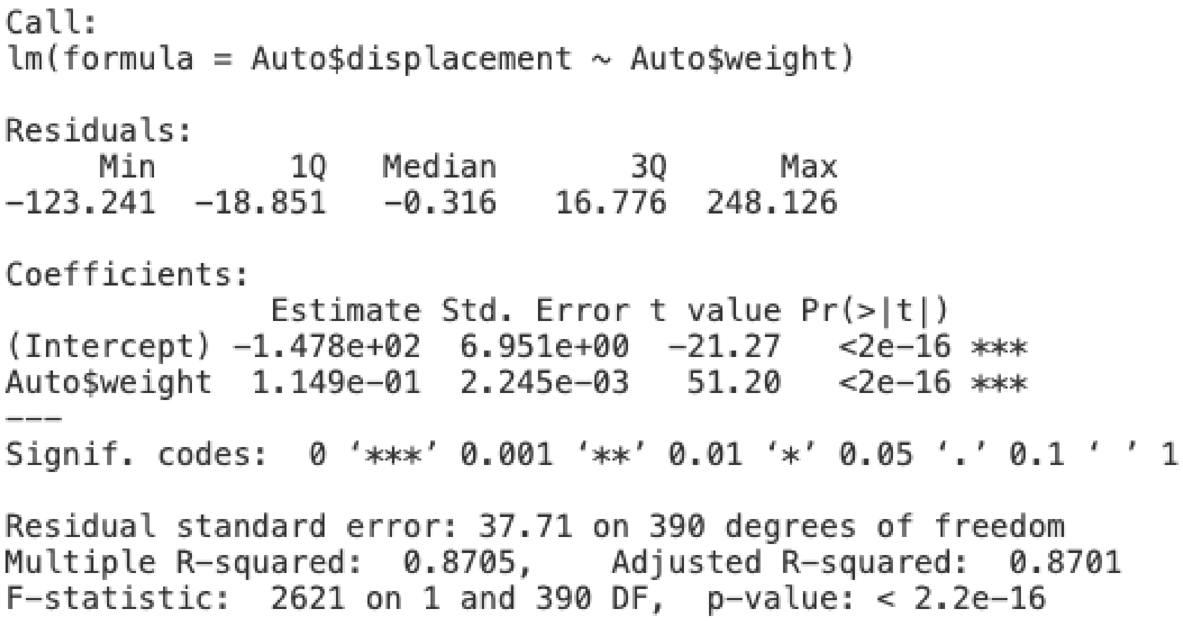
Let’s continue working with Auto data set studied in Assignment 2
1. Review your model for displacement as a function of weight.
a. In your report, include your commands, output, and plots. Find 95% confidence intervals for β0 and β1 . Follow the book’s approach on page 66. Based on your confidence interval for β1 , complete the following sentence: For each 1000 lb increase in weight, we should expect an increase (or is it a decrease?) in displacement between …
Explain why there is no practical interpretation for β0 (it may help to review page 67).

Code:

```
# Using the command on page 49:
Auto = read.csv(file="Auto.csv",na.strings="?", stringsAsFactors = T)
#print(Auto)
#dim(Auto)
#summary(Auto)

Auto= na.omit(Auto) #remove the rows that have missing entries now we have 392 cars remaining
dim(Auto)
#summary(Auto)

plot(Auto$weight, Auto$displacement,col="blue")
myline=lm(Auto$displacement ~ Auto$weight) #Using the lm command to find the least squares
line
summary(myline)
plot(Auto$weight,Auto$displacement,col="blue")
abline(myline,col="red")
```



```
Call:
lm(formula = Auto$displacement ~ Auto$weight)

Residuals:
     Min       1Q   Median       3Q      Max 
-123.241  -18.851   -0.316   16.776  248.126 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -1.478e+02  6.951e+00  -21.27   <2e-16 ***
Auto$weight  1.149e-01  2.245e-03   51.20   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 37.71 on 390 degrees of freedom
Multiple R-squared:  0.8705,	Adjusted R-squared:  0.8701 
F-statistic:  2621 on 1 and 390 DF,  p-value: < 2.2e-16
```

FIGURE 1

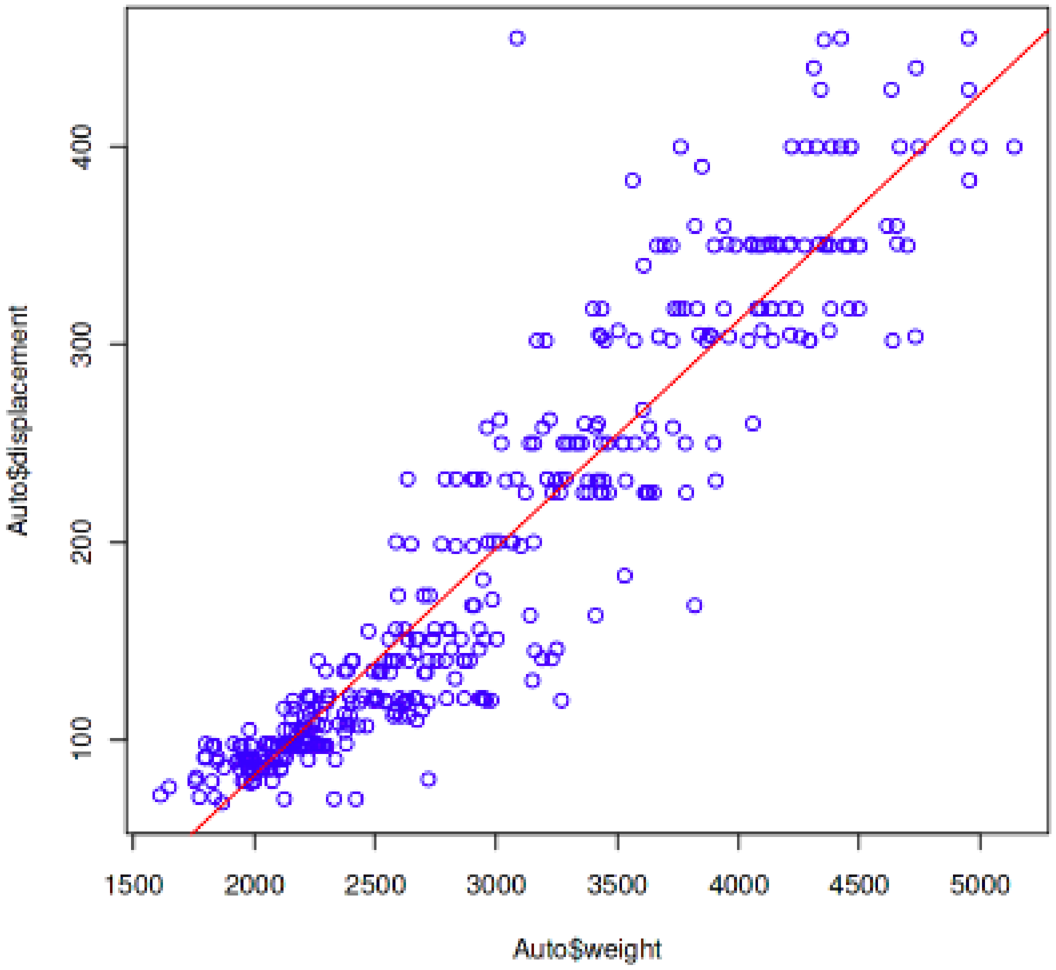


FIGURE 2

In order to calculate the 95% interval for β0 and β1 we need to use the following formula:

[βˆ1 − 2 · SE(βˆ1), βˆ1 + 2 · SE(βˆ1)]
[βˆ0 − 2 · SE(βˆ0), βˆ0 + 2 · SE(βˆ0)]

According to summary table (Figure 1) obtained above:
βˆ0= - 147.8
SE(βˆ0)= 6.951
βˆ1 = 0.1149
SE(βˆ1)= 0.0022

By substituting these numbers in the formula above, the following intervals are obtained

βˆ0 (-161.702 , -133.898)
βˆ1 (0.1105, 0.1193)

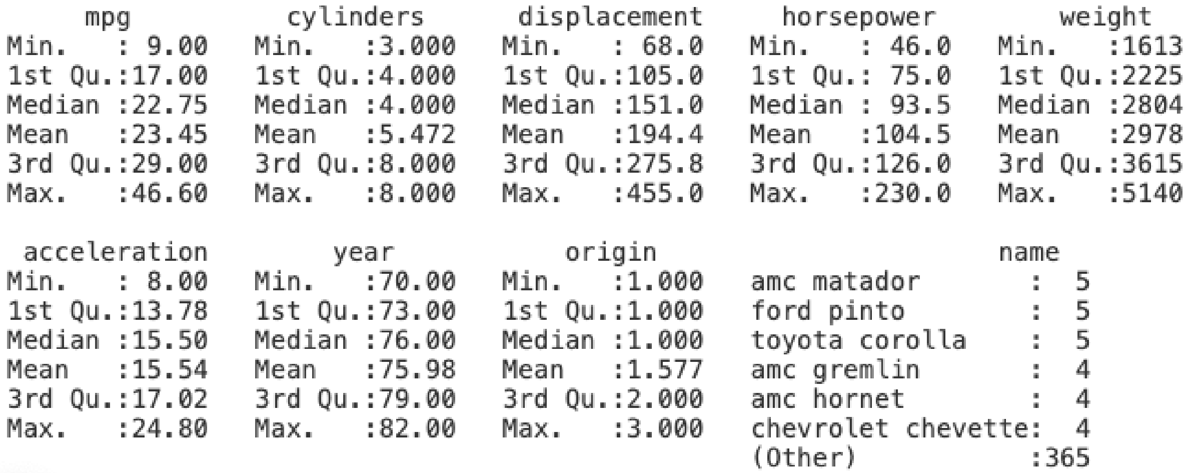
Based on the intervals reported above for an increase of 1000 lbs in weight, we should expect an increase in displacement of between 110.5 and 119.3 feet.
There is not any practical interpretation for βˆ0 because there will not be any auto-existing weighting 0 or fewer lbs

b. Explain why you know there is a clear relationship between weight and displacement. Refer to the R output in your answer.

In order to understand if there is a relationship between the two variables and how strong is the relationship it is possible to look at the R-value present in the summary table above (Figure 1).
The R^2 value will be always a number between 0 and 1. A value close to 0 says there is very little connection between the 2 variables whilst a value near 1 says there is a strong connection between the two variables.
R^2 = 0.8705
By taking the square root of this number we obtain R = 0.9330 which is a number really close to 1 showing us that there is a strong relationship between weight and displacement.
It is also possible to run a Hypothesis test to understand if there is a clear relationship between the two variables, where the null hypothesis Ho = β1 = 0 meaning that there is no relationship between the two variables and H1 = β1≠0
According to the summary table reported above (Figure 1) the the p-value < 2e-16 which is a really small number that allows us to reject the null Hypothesis explaining that there is a relationship between the two variables.

c. What is the percentage error in RSE? Follow the book’s example on p. 69.

According to the summary table above (Figure 1) the RSE = 37.71



```
      mpg          cylinders      displacement     horsepower        weight    
 Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0   Min.   :1613  
 1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0   1st Qu.:2225  
 Median :22.75   Median :4.000   Median :151.0   Median : 93.5   Median :2804  
 Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5   Mean   :2978  
 3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0   3rd Qu.:3615  
 Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0   Max.   :5140  
                                                                               
  acceleration        year           origin                      name    
 Min.   : 8.00   Min.   :70.00   Min.   :1.000   amc matador       :  5  
 1st Qu.:13.78   1st Qu.:73.00   1st Qu.:1.000   ford pinto        :  5  
 Median :15.50   Median :76.00   Median :1.000   toyota corolla    :  5  
 Mean   :15.54   Mean   :75.98   Mean   :1.577   amc gremlin       :  4  
 3rd Qu.:17.02   3rd Qu.:79.00   3rd Qu.:2.000   amc hornet        :  4  
 Max.   :24.80   Max.   :82.00   Max.   :3.000   chevrolet chevette:  4  
                                                 (Other)           :365  
```

FIGURE 3

By looking a this descriptive statistics table (Figure 3) it is possible to see that the mean value for displacement is equal to 194.4

By taking the ratio between these two values and multiplying it by 100 it is possible to find the percentage error in RSE : (37.71 / 194.4) x 100 = 19.40%

d. From viewing the plot of the model with the data, is the model more accurate for cars with low weight (say 2000 lb) or high weight (say 4000 lb)? Explain.

According to the plot above, the model will be more accurate for cars with low weight because there is a higher concentration of data closer to the regression line in comparison to the very spread data for cars with high weight

2. Repeat the four parts above for acceleration as a function of horsepower.

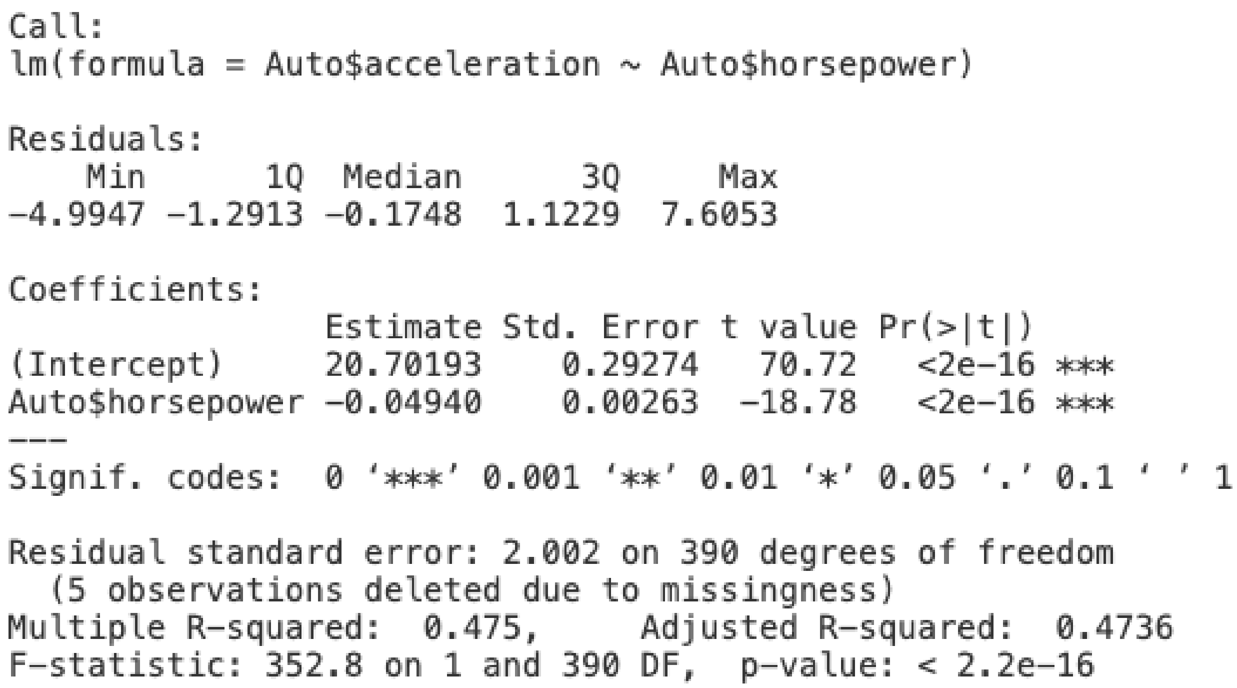
• For part a: For each 100 increase in horsepower, we should expect an increase (or is it a decrease?) in acceleration between …

a. In your report, include your commands, output, and plots. Find 95% confidence intervals for $\beta0$ and $\beta1$. Follow the book's approach on page 66. Based on your confidence interval for $\beta1$, complete the following sentence: For each 100 increase in horsepower, we should expect an increase (or is it a decrease?) in acceleration between …

Explain why there is no practical interpretation for $\beta0$ (it may help to review page 67).

Code:

```
Auto = read.csv(file="Auto.csv",na.strings="?", stringsAsFactors = T)
myline=lm(Auto$acceleration ~ Auto$horsepower) #Using the lm command to find the least squares line
summary(myline)
plot(Auto$horsepower,Auto$acceleration,col="blue")
abline(myline,col="red")
```



```
Call:
lm(formula = Auto$acceleration ~ Auto$horsepower)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.9947 -1.2913 -0.1748  1.1229  7.6053 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)     20.70193    0.29274   70.72   <2e-16 ***
Auto$horsepower -0.04940    0.00263  -18.78   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.002 on 390 degrees of freedom
  (5 observations deleted due to missingness)
Multiple R-squared:  0.475,	Adjusted R-squared:  0.4736 
F-statistic: 352.8 on 1 and 390 DF,  p-value: < 2.2e-16
```

FIGURE 4

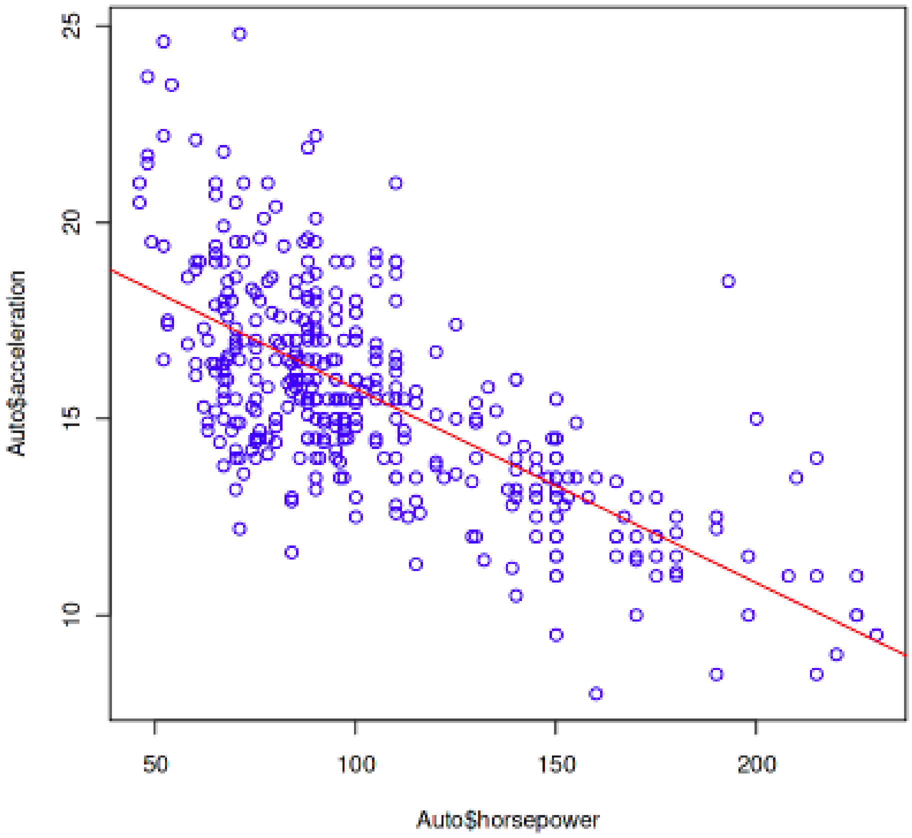


FIGURE 5

In order to calculate the 95% interval for β0 and β1 we need to use the following formula:

[β^1 − 2 · SE(β^1), β^1 + 2 · SE(β^1)]
[β^0 − 2 · SE(β^0), β^0 + 2 · SE(β^0)]

According to descriptive statistics obtained above:
βˆ0= 20.70
SE(βˆ0)= 0.2927
βˆ1 = -0.0494
SE(βˆ1)= 0.0026

By substituting these numbers in the formula above, the following intervals are obtained

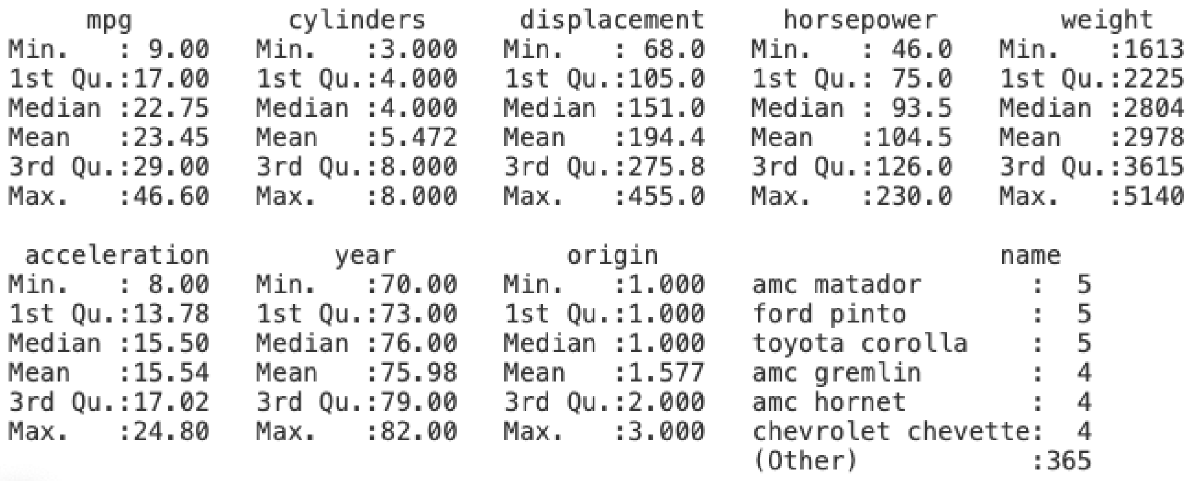
βˆ0 (20.114, 21.285)
βˆ1 (-0.0546, -0.0442)

For each 100 increase in horsepower, we should expect a decrease in acceleration between – 5.46 m/s^2 and – 4.42 m/s^2
β0 will have no practical interpretation because there are no cars with 0 horsepowers

b. Explain why you know there is a clear relationship between weight and displacement. Refer to the R output in your answer.

In order to understand if there is a relationship between the two variables and how strong is the relationship it is possible to look at the R-value present in the summary table above (Figure 4).
The R^2 value will be always a number between 0 and 1. A value close to 0 says there is very little connection between the 2 variables whilst a value near 1 says there is a strong connection between the two variables.
R^2 = 0.475
By taking the square root of this number we obtain R = 0.689 which is a number closer to 1 than 0 showing us that there is a relationship between horsepower and acceleration, however, this relationship is not as strong as the previous one between weight and displacement which had an higher R-value

c. What is the percentage error in RSE? Follow the book’s example on p. 69.



```
      mpg          cylinders      displacement     horsepower        weight    
 Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0   Min.   :1613  
 1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0   1st Qu.:2225  
 Median :22.75   Median :4.000   Median :151.0   Median : 93.5   Median :2804  
 Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5   Mean   :2978  
 3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0   3rd Qu.:3615  
 Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0   Max.   :5140  

  acceleration        year           origin                      name    
 Min.   : 8.00   Min.   :70.00   Min.   :1.000   amc matador       :  5  
 1st Qu.:13.78   1st Qu.:73.00   1st Qu.:1.000   ford pinto        :  5  
 Median :15.50   Median :76.00   Median :1.000   toyota corolla    :  5  
 Mean   :15.54   Mean   :75.98   Mean   :1.577   amc gremlin       :  4  
 3rd Qu.:17.02   3rd Qu.:79.00   3rd Qu.:2.000   amc hornet        :  4  
 Max.   :24.80   Max.   :82.00   Max.   :3.000   chevrolet chevette:  4  
                                                 (Other)           :365  
```

FIGURE 6
According to the summary table created for horsepower (Figure 4) and acceleration and reported above, the RSE = 2.002, and according to the descriptive table (Figure 6) above the mean for acceleration is 15.54

By taking the ratio between those two values and multiplying the result by 100 it is possible to find the percentage error in RSE
(2.002 / 15.54) x 100 = 12.88 %

• For part d, is the model more accurate for cars with low horsepower (say 100) or high horsepower (say 400)?

According to the plot above (Figure 5) the model is more accurate for cars with low horsepower because of the higher number of data in that region and the concentration of the data around the regression line. The data for high horsepower are all spread around with an high standard deviation not giving us a right estimate of the trend behavior