Activity 7 - Salvin Chowdhury

Due Date: May 9, 2025

Problem 1:

Data for fuel economy, weight, and price for many popular 2015 models and their 2024 counterparts of vehicles is recorded in the file "Activity7CarData2015.csv" (which can be found in Canvas). All 2015 vehicles for which data was gathered are not hybrids, and use regular unleaded gasoline.

For this problem, we wish to investigate the relationship between fuel economy and weight of vehicles, and we'll stick with the 2015 models to do this.

In particular, we want to predict the fuel economy of the vehicle (combined city & highway - in miles per gallon) using the curb weight of the vehicle (the total weight of the vehicle, not including passengers or cargo - in pounds).

Be sure to run the code chunk below so that you can use the csv file within this markdown file.

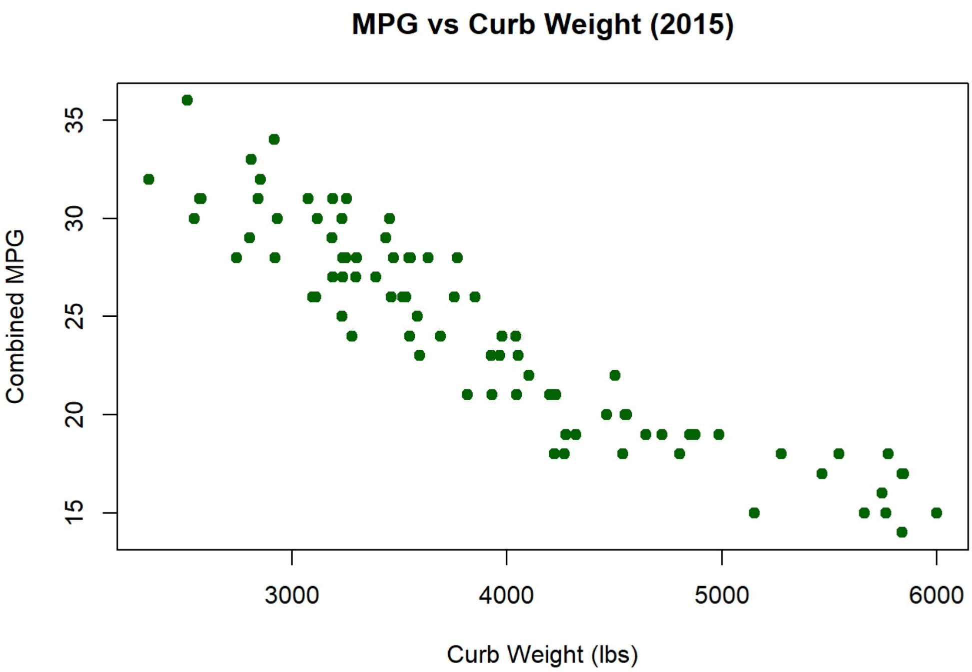
```
mpgdata <- read.csv("Activity7CarData2015.csv")
```

a. First, create a scatter plot of the data we are interested in. Include a title for the plot as well as clear labels for each of the axes. Describe the trend (if any) between the variables. Then, find the simple linear regression model and add the plot of this line to the scatter plot. Finally, take a look at the summary of the model, as this will be useful going forward.

```
# code for part (a)

# saving the plot
# png("C:/GitHub/StatisticsInR/wk_15/plots/mpg_vs_weight.png", width = 800, height = 600)

# combined mpg vs weight
plot(mpgdata$CURB.WEIGHT.2015, mpgdata$COMBINED.MPG.2015,
     main = "MPG vs Curb Weight (2015)",
     xlab = "Curb Weight (lbs)",
     ylab = "Combined MPG",
     pch = 19, col = "darkgreen")
```



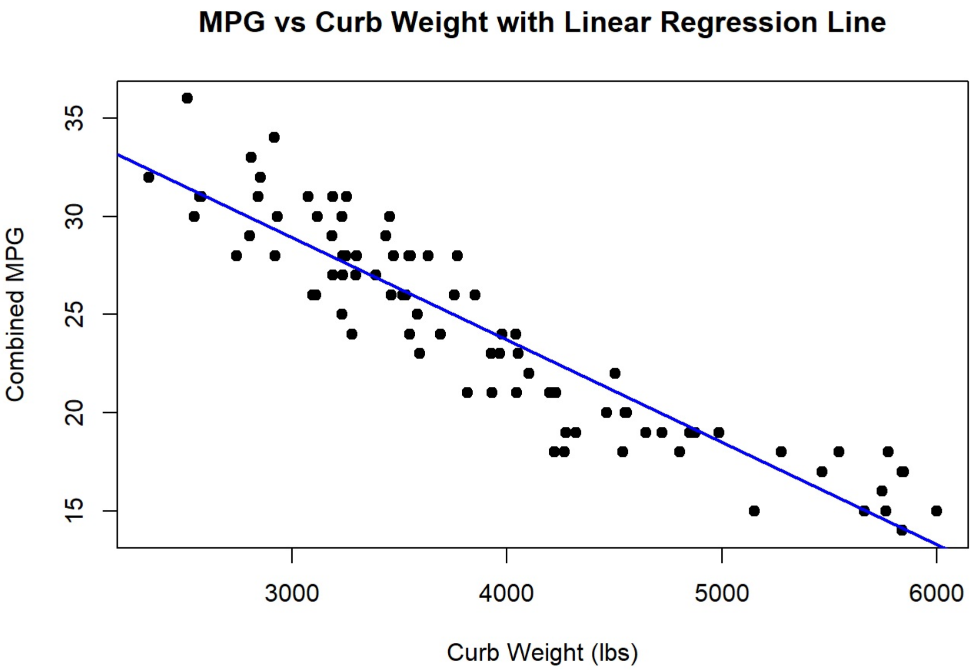

```
# saving the plot
# dev.off()
```

```
# fitting the linear regression model
model <- lm(COMBINED.MPG.2015 ~ CURB.WEIGHT.2015, data = mpgdata)

# saving the plot
# png("C:/GitHub/StatisticsInR/wk_15/plots/mpg_vs_weight_regression.png", width = 800, height = 600)

# producing a new scatter plot
plot(mpgdata$CURB.WEIGHT.2015, mpgdata$COMBINED.MPG.2015,
    main = "MPG vs Curb Weight with Linear Regression Line",
    xlab = "Curb Weight (lbs)",
    ylab = "Combined MPG",
    pch=19
)

# adding the regression line
abline(model, col='blue', lwd=2)
```




```
# closing the device
# dev.off()

# obtaining summary of the model
summary(model)
```

```
##
## Call:
## lm(formula = COMBINED.MPG.2015 ~ CURB.WEIGHT.2015, data = mpgdata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.5631 -1.3168 -0.1471  1.5229  4.6466
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)      44.5716348  0.9690848   45.99   <2e-16 ***
## CURB.WEIGHT.2015 -0.0052153  0.0002401  -21.72   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.093 on 83 degrees of freedom
## Multiple R-squared:  0.8504, Adjusted R-squared:  0.8486
## F-statistic: 471.9 on 1 and 83 DF,  p-value: < 2.2e-16
```

Answer for (a) Looking at the slop of the line, we can say that as the curb weight increases, the combined MPG decreases.

b. What is the slope of this regression? Interpret this value in the context of the problem (it may make sense to multiply your values by 1000 for the interpretation).

Answer for (b) Since the slope of the regression line is -0.0052, which means for every 1,000 pound increase in a vehicle's curb weight,, the combined fuel economy is expected to decrease by about 5.22 MPG. This simply means that the combined fuel economy is expected to decrease by about 5.22 MPG.

c. What is the value of the y-intercept of this regression? Interpret this value in the context of the problem (regardless of whether this logically makes sense or not). Does the interpretation of the intercept have a logical interpretation in this example? Why or why not?

Answer for (c) Looking at the value of the intercept, we see that it is 44.57. This is simply the predicted value when the curb weight is 0 pounds. The interpretation of the intercept doesn't have a logical interpretation in this example because a vehicle can't have zero weight.

d. Suppose the curb weight of a vehicle in 2015 was 4000 pounds. What would our model predict the fuel economy to be?

Answer for (d) To calculated the MPG, we use the formula y = mx + c, where m = -0.0052153, x = 4000 and c = 44.5716348. Using the following values, we get our value to be 23.71. To concluce, we say that if a vehicle has a curb weight of 4000 pounds, the model predicts the fuel economy to be 23.71 MPG.

e. What is the residual for the 2015 Honda CR-V?

```
# fetching the data
honda_crv = mpgdata[mpgdata$MAKE == 'Honda'  & mpgdata$MODEL == "CR-V",]

# retrieving the actual and predicted MPG
actual_mpg <- honda_crv$COMBINED.MPG.2015
predicted_mpg <- predict(model, newdata = honda_crv)

# calculating the residual
residual_value <- actual_mpg - predicted_mpg
print(residual_value)
```

```
##       35
## 1.953051
```

Answer for (e) The residual is 1.953051

f. What is the r^2 value? Interpret this value.

Answer for (f) The R^2 value is 0.8504. This means that about 85.04% of the variation in 2015 combined MPG across vehicles can be explained by their curb weight.

g. Calculate a 95% confidence interval for the slope coefficient, and provide an interpretation in the context of the problem (it may make more sense to multiply your values by 1000 for your interpretation). You can use the information from our model summary to do this, or there is a confint() function that can accomplish this as well - you can choose which you prefer to use.

```
# code for part (g)

# finding the confidence interval
confint(model)
```

```
##                         2.5 %      97.5 %
## (Intercept)      42.644164193 46.49910549
## CURB.WEIGHT.2015 -0.005692797 -0.00473777
```

Answer for (g) The 65% confidence interval for the slope is [-0.005692797, -0.00473777] We multiply the values by a 1,000 which means [−5.6928,−4.7378]. This means based on the 2015 vehicle data, we are 95% confident that for every 1,000 pound increase in curb weight, the combined fuel economy decreases by between 4.74 and 5.69 MPG.