Analyzing the Relationship Between Car Weight and Mileage Using Linear Regression

Anju Sambasivan

Research Question

Does the mileage of a car increase as its weight increases?

Methodology

To address the research question, data from the `mtcars` dataset was analyzed to understand the relationship between car weight (`wt`) and fuel efficiency (`mpg`). The fuel efficiency, originally measured in miles per gallon, was converted to kilometers per liter for better familiarity.

A **linear regression model** was used to analyze the relationship, with `kpl` (fuel efficiency) as the response variable and `wt` (car weight) as the predictor variable. The regression model was used to determine how car weight impacts mileage. The results were visualized using a scatterplot with a regression line, providing a clear depiction of the relationship.

Finally, the summary of the regression model and the correlation coefficient between the two variables were used to interpret the findings.

```
# Load necessary libraries

library(ggplot2)
library(dplyr)
```

```
# Convert mpg to km/l

mtcars <- mtcars %>%
  mutate(kpl = round(mpg * 0.425144, 2))
```

```r
# Fit the linear regression model for kpl and weight

model <- lm(kpl ~ wt, data = mtcars)
```

```r
# Get the summary of the model

summary(model)
```

```

Call:
lm(formula = kpl ~ wt, data = mtcars)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.92797 -1.00308 -0.05512  0.60049  2.91824 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  15.8500     0.7975  19.874  < 2e-16 ***
wt           -2.2719     0.2375  -9.567 1.27e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.294 on 30 degrees of freedom
Multiple R-squared:  0.7531,	Adjusted R-squared:  0.7449 
F-statistic: 91.52 on 1 and 30 DF,  p-value: 1.271e-10
```

Interpretation of Linear Regression Results

1. **Intercept (15.8500)**: When the car's weight is zero (hypothetically), the estimated fuel efficiency (mileage) is 15.85 km/l. While this is not realistic, it provides a reference point for the model.
2. **Slope (-2.2719)**: For every additional 1000 lbs of weight, the mileage decreases by approximately 2.27 km/l. This confirms a negative relationship between weight and mileage—heavier cars are less fuel-efficient.
3. **Statistical Significance**:
   - The **p-value for the slope (1.27e-10)** is extremely small (much less than 0.05), indicating that the relationship between weight and mileage is statistically significant.

- The three asterisks (`***`) also highlight high significance.

4. **Model Fit**:

   - The **Multiple R-squared (0.7531)** indicates that approximately 75.3% of the variability in mileage can be explained by the car's weight.
   - The **Adjusted R-squared (0.7449)** accounts for model complexity and shows a very similar value, confirming a good fit.

5. **Residual Standard Error (1.294)**: On average, the predictions of the model differ from the actual data by about 1.29 km/l.

6. **F-statistic (91.52)**: This value and its associated p-value (1.271e-10) show that the overall model is statistically significant.

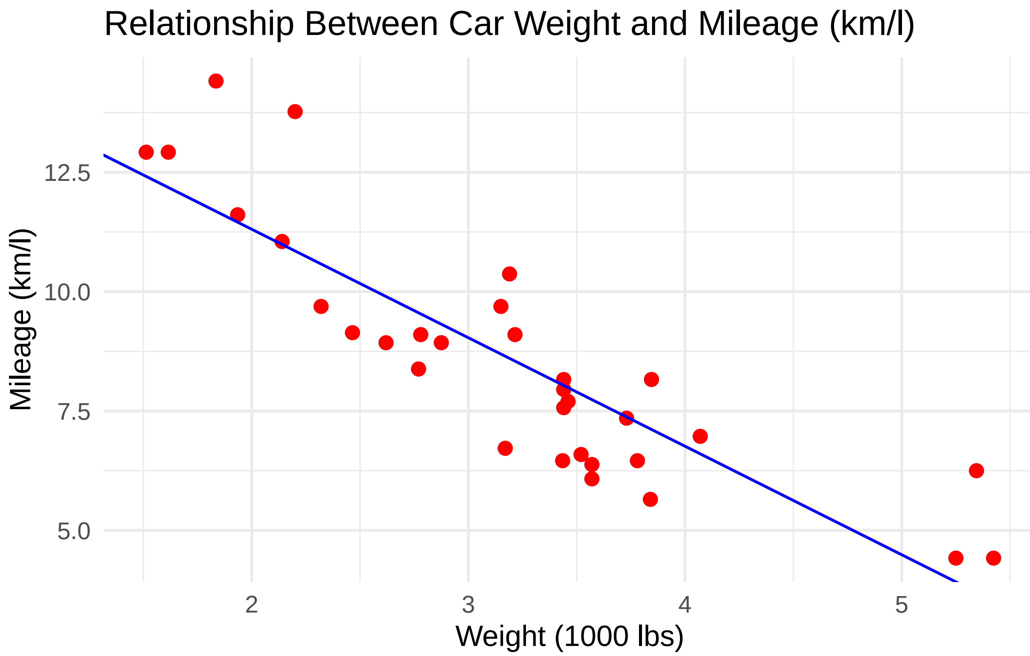
```r
# Create a scatter plot with the regression line for kpl

ggplot(mtcars, aes(x = wt, y = kpl)) +

  # Scatter plot of data points
  geom_point(color = "red", size = 2) +

  # Regression line
  geom_abline(intercept = coef(model)[1], slope = coef(model)[2],
              color = "blue") +

  labs(title = "Relationship Between Car Weight and Mileage (km/l)",
       x = "Weight (1000 lbs)",
       y = "Mileage (km/l)") +
  theme_minimal()
```



The function `geom_abline()` is used in a graph to draw a straight line. This line is based on the equation: y=a+bx

Where:

- `a` **(intercept)**: The starting value of y when x=0. It tells where the line crosses the y-axis.
- `b` **(slope)**: This shows how much y changes when x increases by 1 unit. A positive slope means the line goes up, and a negative slope means it goes down.

The term **"abline"** is short for "a + b line," meaning the straight line defined by the intercept and slope. In the code, we use **`coef(model)[1]`** to get the intercept (a). **`coef(model)[2]`** to get the slope (b). This line is often used in graphs to show the relationship between two variables, like a regression line, making it easier to see trends in the data.

The graph shows the relationship between a car's weight (in 1000 lbs) and its mileage (in km/l). The blue line represents a linear regression model, which shows a clear negative relationship between the two variables. This means that as the weight of a car increases, its mileage decreases. The data points are close to the line, indicating a strong relationship. For example, lighter cars (around 2 on the weight scale) have higher mileage (above 12 km/l), while heavier cars (around 5) have much lower mileage (below 5 km/l). This graph demonstrates that heavier cars are less fuel-efficient, providing a clear answer to the research question: mileage does not increase with weight.

```r
# Calculate Pearson correlation between car weight (wt) and mileage (kpl)

pearson_correlation <- cor(mtcars$wt, mtcars$kpl)
pearson_correlation
```

```
[1] -0.8678309
```

The `cor()` function calculates the **Pearson correlation coefficient**, which measures the strength and direction of the linear relationship between two variables. Here, it calculates the correlation between `wt` (car weight) and `kpl` (mileage in kilometers per liter). We get the correlation coefficient is **-0.8678**. This indicates a **strong negative relationship**: as car weight increases, mileage decreases.

```r
# Calculate the square of the Pearson correlation

r_squared <- pearson_correlation^2
r_squared
```

```
[1] 0.7531305
```

The Multiple R-squared value from linear regression and the square of the Pearson correlation coefficient are the same when there is only one independent variable in the regression model.The square of the Pearson correlation tells us how much of the change in one variable is related to the change in the other variable. Here r^2 = 0.7531, this means 75.31% of the changes in car mileage (kpl) can be explained by the car's weight (wt). This shows that the relationship between weight and mileage is strong, and weight is a significant factor affecting mileage.

Findings

The scatterplot and regression line in the analysis show that the data points for the 32 automobiles are closely aligned with the regression line. This indicates a **strong linear relationship** between car weight and mileage. The negative slope of the regression line demonstrates an **inverse relationship**, meaning as the weight of the car increases, its fuel efficiency decreases.

The negative Pearson correlation coefficient further supports this finding, showing a **strong negative correlation** between the two variables. With an r value of -0.8678, the square (r^2=0.7531) indicates that approximately **75.31% of the variation in mileage (kpl)** is explained by car weight.

From this analysis, it is evident that heavier cars have lower fuel efficiency. Therefore, the answer to the research question is **"NO."** Mileage does not increase as the weight of the car increases.