# Lecture 5: Interpreting the Model Results

Dr. Logan Kelly

2024-09-11

## Overview:

In this lecture, we will interpret the results of a simple linear regression model. The model predicts **Energy Efficiency (MPG)** using **Horsepower** as the predictor variable. We will explain how to interpret the coefficients, R-squared value, p-values, and discuss what these metrics reveal about the relationship between **Horsepower** and **MPG**.

## Key Learning Outcomes:

By the end of this lecture, students will be able to: - Interpret the coefficients from a linear regression model. - Understand the significance of the R-squared value and residuals. - Evaluate p-values to assess statistical significance. - Connect the model output to real-world contexts, focusing on the trade-off between engine power and fuel efficiency.

## Interpreting the Model Output:

The regression output provides a summary of the model fitted between **Energy Efficiency (MPG)** and **Horsepower**:

```
Call:
lm(formula = `Energy Efficiency (MPG)` ~ Horsepower, data = car_data_clean)

Residuals:
   Min     1Q Median     3Q    Max
-9.968 -1.966  0.683  1.933  9.241

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept) 41.270812   1.864643  22.133  < 2e-16 ***
Horsepower  -0.053695   0.006956  -7.719 2.68e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.689 on 38 degrees of freedom
Multiple R-squared:  0.6106,    Adjusted R-squared:  0.6003
F-statistic: 59.58 on 1 and 38 DF,  p-value: 2.677e-09
```

#### Call:

```
lm(formula = `Energy Efficiency (MPG)` ~ Horsepower, data = car_data_clean)
```

#### Residuals:

- **Residuals** are the differences between the actual and predicted values of **MPG**.
  - The smallest residual is -9.968, and the largest is 9.241.
  - The median residual is close to zero (0.683), suggesting that most predictions are fairly accurate.
  - The first quartile (1Q) and third quartile (3Q) residuals are -1.966 and 1.933, showing that most residuals are within this range.

#### Coefficients:

The table below shows the relationship between **Horsepower** and **MPG**:

| Coefficient | Estimate | Std. Error | t value | Pr(> |
|---|---|---|---|---|
| Intercept | 41.270812 | 1.864643 | 22.133 | < 2e-16 *** |
| Horsepower | -0.053695 | 0.006956 | -7.719 | 2.68e-09 *** |

- **Intercept**:
  - The intercept is **41.27**, which represents the estimated **MPG** when **Horsepower** is zero. While this isn't realistic, it serves as a starting point for the model.
- **Slope (Horsepower)**:
  - The slope is **-0.0537**. This means that for every additional unit of **Horsepower**, **MPG** decreases by about 0.054.
  - A negative slope is expected, as higher horsepower generally results in lower fuel efficiency.

### Statistical Significance:

- **P-value for Horsepower**:
  - The p-value for **Horsepower** is **2.68e-09**, which is much smaller than 0.05. This means the relationship between **Horsepower** and **MPG** is statistically significant.
  - The three asterisks ('*'**) indicate high statistical significance ($p < 0.001$).

### Goodness of Fit:

The **Goodness of Fit** measures tell us how well the model explains the data:

```
Residual standard error: 3.689 on 38 degrees of freedom
Multiple R-squared:  0.6106,    Adjusted R-squared:  0.6003
F-statistic: 59.58 on 1 and 38 DF,  p-value: 2.677e-09
```

- **Residual Standard Error (RSE)**:
  - The residual standard error is **3.689**, meaning the typical prediction error is about 3.69 MPG.
  - A smaller RSE would mean the model is more accurate.
- **R-squared**:
  - The **R-squared** value is **0.6106**, meaning the model explains **61.06%** of the variation in **MPG**.
  - This shows that **Horsepower** is a fairly strong predictor of **MPG**, but 38.94% of the variation remains unexplained.
- **Adjusted R-squared**:
  - The **Adjusted R-squared** is **0.6003**, which is close to **R-squared** since we have only one predictor. This confirms the model isn't overfitting.
- **F-statistic**:
  - The F-statistic is **59.58**, with a p-value of **2.677e-09**. This shows the model as a whole is statistically significant, meaning **Horsepower** is a good predictor of **MPG**.

## Connecting the Output to the Real World:

- **Application of Results**:
  - The model confirms the trade-off between **Horsepower** and **MPG**. More powerful vehicles tend to be less fuel-efficient.

  - This insight is useful for both manufacturers and consumers. Manufacturers can design cars with this trade-off in mind, while consumers can choose based on their needs—either fuel efficiency or performance.

- **Trade-offs**:
  - The negative slope shows the trade-off between engine power and fuel efficiency. Manufacturers must balance **Horsepower** and **MPG** when designing vehicles.
  - Consumers must decide whether they want more horsepower at the cost of lower **MPG** or vice versa.

## Potential Limitations of the Model:

- **Linearity Assumption**:
  - The model assumes a linear relationship between **Horsepower** and **MPG**. However, this relationship might not be linear, especially for very high or very low horsepower values.
  - More complex models might better capture this relationship.

- **Unexplained Variance**:
  - While the model explains 61% of the variation in **MPG**, the remaining 39% is unexplained. Other factors, like vehicle weight or aerodynamics, might also affect **MPG**.

## Summary of Key Concepts:

- **Coefficients**: The slope (-0.0537) shows that **MPG** decreases by about 0.054 for each additional unit of **Horsepower**.
- **R-squared**: The model explains 61.06% of the variance in **MPG**, meaning **Horsepower** is a strong predictor, but other factors may still be at play.
- **P-values**: The p-value for **Horsepower** (2.68e-09) indicates that **Horsepower** significantly impacts **MPG**.
- **Real-World Application**: The model provides insights into the trade-offs between engine performance and fuel efficiency, important for both manufacturers and consumers.

## Next Steps:

In the next lecture, we will explore strategies for dealing with outliers and refining the model for better accuracy. We will discuss how to identify and address outliers and how doing so can improve model performance.

### Assignment for Students:

1. Use the model output from your own regression to interpret the coefficients and R-squared values.
2. Discuss the real-world implications of the relationship between **Horsepower** and **MPG**.
3. Identify any potential limitations of your model and suggest possible ways to refine it.