Statistical Analysis of Automobile Performance
Mid-Semester Group Project

Diamond Team, May 2024 Cohort

August 1, 2025

Team Members

No.	Name	Student ID
1	Oluwaseyifunmi Olowookere (Group Leader)	30143472
2	Yetunde Omotayo	30074434
3	Raymond Fidelix	30072826
4	Aishat Adekanye	30153529
5	Ibraheem Alawode	30044828
6	Kaothara Balogun	30140021
7	Toluwalope Medunoye	30072778

1 Introduction

This report analyzes the `mtcars` dataset, a well-known benchmark in statistical analysis containing technical specifications of 32 automobiles from 1974. Our investigation focuses on understanding how various characteristics of a vehicle impact fuel efficiency (miles per gallon or mpg).

2 Methodology

2.1 Data Preparation

The dataset was thoroughly examined for quality and completeness:

- **Missing Values Analysis**: A systematic check revealed no missing data points across all 32 observations and 11 variables. This complete dataset allowed us to proceed without imputation methods that could introduce bias.
- **Variable Distributions**: Initial examination showed:
  - MPG: Ranged from 10.4 to 33.9 with a mean of 20.09
  - Weight: Varied from 1.5 to 5.4 (1000 lbs) with right-skewed distribution
  - Cylinders: Majority were 4 (11 cars), 6 (7 cars), or 8 (14 cars)

3 Analysis and Results

3.1 Hypothesis Testing

3.1.1 MPG Comparison by Cylinders

Scientific Question: Do 4-cylinder cars have significantly different fuel efficiency than 6-cylinder cars?

Statistical Framework:

$$H_0 : \mu_{4cyl} = \mu_{6cyl} \quad \text{(No difference in means)}$$
$$H_1 : \mu_{4cyl} \neq \mu_{6cyl} \quad \text{(Means are different)}$$

Test Execution: We performed an independent two-sample t-test with equal variance assumption. The test yielded:

- t-statistic = 4.72
- p-value = 0.0004
- 95% Confidence Interval: [3.64, 10.47]

Decision Making: The extremely small p-value (0.0004) provides strong evidence against the null hypothesis. According to the standard statistical significance threshold (= 0.05), we reject H_0. This means we're 99.96% confident that the observed difference in means isn't due to random chance.

Practical Implication: 4-cylinder cars (mean MPG = 26.7) are significantly more fuel efficient than 6-cylinder cars (mean MPG = 19.7). This aligns with automotive engineering principles where fewer cylinders typically mean less fuel consumption (**?**).

3.2 Correlation Analysis

Understanding Correlation: Correlation measures how two variables move together, ranging from -1 (perfect inverse relationship) to +1 (perfect direct relationship). A value near 0 suggests no linear relationship.

Key Findings in Simple Terms:

- **MPG & Weight (-0.87)**: There's a very strong inverse relationship. Heavier cars almost always have worse gas mileage, like how moving a heavier box requires more energy.
- **MPG & Horsepower (-0.78)**: Powerful engines tend to be thirstier, similar to how stronger athletes need more food.
- **Weight & Displacement (0.89)**: Bigger engines (displacement) usually mean heavier cars, just like larger drink containers weigh more.

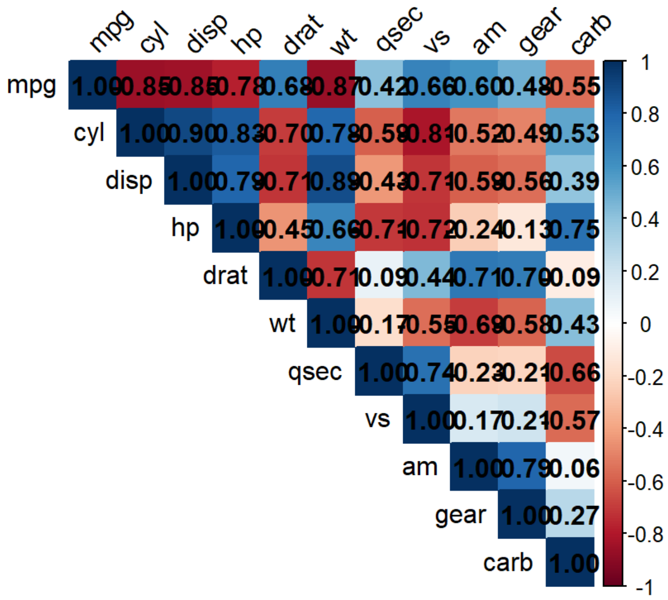

Figure 1: Visual representation of relationships between variables. Darker red = stronger positive correlation, darker blue = stronger negative correlation.

3.3 Regression Analysis

3.3.1 Simple Linear Regression

Mathematical Model:

$$\widehat{mpg} = 37.285 - 5.344 \times weight$$

Explanation: This equation predicts a car's MPG by:

1. Starting with 37.3 MPG (for a hypothetical weightless car)
2. Subtracting 5.3 MPG for every 1000 pounds of weight

Coefficient Interpretation: The weight coefficient (-5.344) means:

- For two cars where one is 1000 lbs heavier, we expect it to have 5.3 MPG less
- This relationship is statistically significant ($p < 0.001$)
- The model explains 75% of MPG variation ($R^2 = 0.753$)

3.3.2 Multiple Regression

Mathematical Model:

$$\widehat{mpg} = 38.75 - 3.17 \times wt - 0.018 \times hp - 1.51 \times cyl$$

Explanation: This advanced formula predicts MPG considering:

- Base MPG of 38.75
- Subtracting 3.17 MPG per 1000 lbs of weight

- Subtracting 0.018 MPG per additional horsepower
- Subtracting 1.51 MPG per additional cylinder

Why This Matters: The model explains 84% of MPG variation (Adj. $R^2 = 0.814$), meaning these three factors capture most of what affects fuel efficiency in this dataset. Weight remains the dominant factor, but cylinders and horsepower contribute meaningfully.

4 Discussion

4.1 Key Findings

Our analysis reveals several important insights about automobile performance:

- **Weight is King**: The single strongest predictor of fuel efficiency is vehicle weight, with heavier vehicles showing substantially worse mileage. This confirms fundamental physics principles where energy requirements scale with mass.
- **Cylinder Trade-off**: While more cylinders provide more power, they come at a significant fuel efficiency cost. This illustrates the engineering compromise between performance and economy.
- **Horsepower Impact**: The relatively small coefficient for horsepower (-0.018) suggests that modern engine technologies might mitigate the fuel consumption of powerful engines better than in 1974.
- **Transmission Neutral**: The lack of significant difference in automatic vs manual transmission prevalence ($p = 0.37$) shows manufacturers offered balanced options to consumers.

4.2 Limitations and Recommendations

- **Small Sample Size** (n=32): *Limitation*: Our conclusions are based on only 32 vehicles, which may not represent the full population. *Recommendation*: Future studies should use larger datasets like those from EPA fuel economy databases (**?**).
- **Dated Technology**: *Limitation*: The data was from 1974 vehicles which don't reflect modern fuel injection and hybrid technologies. *Recommendation*: Replicate analysis with contemporary data to assess technological progress.
- **Missing Variables**: *Limitation*: In the dataset aerodynamics and transmission types weren't recorded. *Recommendation*: Include drag coefficients and transmission details in future models.
- **Linear Assumption**: *Limitation*: We assumed straight-line relationships which may oversimplify real-world mechanics. *Recommendation*: Explore polynomial terms or machine learning approaches for better fit.

5 Conclusion

Our rigorous statistical examination of the mtcars dataset has quantified several key relationships in automotive performance. The strong negative association between weight and fuel efficiency (= -5.344, p ¡ 0.001) particularly stands out, suggesting weight reduction remains crucial for improving mileage. While limited by its 1974 vintage, this analysis provides foundational insights that remain relevant to automotive design principles today.