Experimental Design Proposal

Effects of Cylinders, Horsepower, Weight, and Model Year on Fuel Efficiency: A 2⁴ Factorial Design Using Car Dataset

1. Introduction and Research Question

Fuel efficiency remains one of the most important performance indicators in the automotive industry, influencing both environmental impact and consumer cost. Previous studies (Ahmad et al., 2020; Greene & Welch, 2017) have shown that physical and mechanical factors such as engine size, vehicle weight, and horsepower are closely related to fuel consumption. However, few studies have analyzed their combined effects within a controlled factorial design framework.

This project aims to examine how a combination of four controllable car characteristics—cylinders, horsepower, weight, and model year—affects miles per gallon (mpg), a direct measure of fuel efficiency.

Research Question: How do the combination of number of cylinders, horsepower, vehicle weight, and model year jointly influence car fuel efficiency (mpg), and which interactions are most significant?

2. Hypotheses

Main Effects:

H₀₁: The number of cylinders does not significantly affect mpg.

H₀₂: Horsepower does not significantly affect mpg.

H₀₃: Weight does not significantly affect mpg.

H₀₄: Model year does not significantly affect mpg.

Interaction Effects:

H₀₅: There are no interaction effects among cylinders, horsepower, weight, and model year.

Hₐ: One or more main or interaction effects significantly affect mpg.

3. Overview of Experimental Design

3.1 Design Type

A 2⁴ factorial design (no replication) will be used.

This design evaluates four independent factors, each at two levels:

|  |  |  |
| --- | --- | --- |
| Factor | Level | Type |
| Cylinders | 4 vs 8 | Discrete |
| Horsepower | Low vs High  (based on median split) | Continuous (categorized) |
| Weight | Low vs High  (based on median split) | Continuous (categorized) |
| Model Year | Old vs New  (e.g., pre-1980 vs post-1980) | Categorical |

Dependent Variable: Miles per Gallon (mpg)— continuous measure of fuel efficiency.

3.2 Experimental Units

Each car in the dataset represents one experimental unit.

Cars are assigned into treatment combinations according to the four factors. With a 2⁴ factorial design, there are 16 treatment combinations (e.g., 4-cylinder, low horsepower, low weight, old model year).

3.3 Randomization and Control

Randomization: Cars are randomly selected from the dataset for each treatment combination to minimize selection bias.

Control Variables: Transmission type and engine displacement will be monitored to reduce confounding effects.

4. Statistical Analysis Plan

4.1 Data Preparation

Data from the uploaded car dataset will be preprocessed in R:

Handle missing values using mean or median imputation.

Categorize continuous predictors (horsepower and weight) into “low” and “high” based on median splits.

Code categorical variables (e.g., cylinders = {4, 8}; model\_year = {old, new}).

4.2 Analysis Method

A four-factor ANOVA (2⁴ design) will be conducted to test main and interaction effects.

Model in Python:

If assumptions are not met:

Apply transformation (e.g., log(mpg)).

Use multiple regression as an alternative model:

4.3 Statistical Tools

Normality test: Shapiro–Wilk test

Variance homogeneity: Levene’s test

Effect size: Partial η² using `effect size` package

Visualization: Interaction plots and residual diagnostics (`ggplot2`, `interactions` package)

5. Brief Literature Review

Studies in automotive engineering and energy policy emphasize that vehicle weight and engine design strongly influence fuel efficiency.

Ahmad et al. (2020) found that a 10% reduction in weight improves fuel economy by up to 8%.

Greene and Welch (2017) reported that higher cylinder counts and engine power typically reduce mpg due to increased energy requirements.

Li and Zhao (2022) observed that newer model years integrate improved engine management and aerodynamics, often enhancing fuel efficiency despite higher horsepower.

While these studies analyzed individual factors, few examined their interaction effects through factorial experiments. This project addresses that gap by using a 2⁴ factorial design to quantify both individual and combined influences of mechanical and temporal factors on mpg.

6. Ethical Considerations

Although the dataset is non-human and publicly available, ethical integrity is maintained by:

Data source: https://www.kaggle.com/datasets/uciml/autompg-dataset.

Privacy and Compliance: Ensuring no personally identifiable information is used.

Reproducibility: Providing python code and random seeds for full replication.

Fair Representation: Avoiding selective reporting or p-hacking.

Academic Integrity: Properly citing all sources and acknowledging dataset creators.

7. Expected Outcomes

Main Effects: Cars with fewer cylinders, lower horsepower, lighter weight, and newer models are expected to have higher mpg.

Interaction Effects:

The negative effect of weight may be stronger for older cars.

The effect of horsepower may depend on the number of cylinders.

These outcomes could provide insights into energy-efficient automotive design and environmental sustainability initiatives.

References

Ahmad, N., et al. (2020). Weight reduction and fuel efficiency in automotive design: An integrated assessment. Transportation Research Part D: Transport and Environment, 86, 102446.

Greene, D. L., & Welch, T. (2017). Impact of engine size and vehicle weight on fuel economy: Policy implications. Energy Policy, 108, 273–282.

Li, F., & Zhao, J. (2022). Advances in automotive efficiency: The role of model year innovation. Applied Energy, 325, 119823.

U.S. Department of Energy. (2021). Fuel Economy Trends Report. Retrieved from [https://www.energy.gov](https://www.energy.gov)