**Cars fuel consumption in the 1970s**

**by Dávid Szabados (ID: 2302806)**

**Introduction**

Currently we are living in a world, where the fossil fuel consumptions (e.g.: coal, oil) are still really high in some countries and have a problem to adapt to newer resources, like in Germany[[1]](#footnote-1), where the fear of nuclear power plants lead to a negative trend, where instead of changing the fossil power plants to renewable ones, they are changing nuclear power plants instead. As fossil fuels, especially oil/petrol, are still one of the main resources that makes our economy running, it should be interesting to look at the cars of the past, and check, what influenced the fuel consumption for them.

**Data**

For my research, I will use the “Automotive Fuel Economy” dataset from Maven Analytics. This dataset includes the fuel economy in mpg (miles per gallon) for almost 400 cars, that have been sold in the US during the 70s and 80s, but we will only focus on one decade, the 70s’ sales. With this, we hopefully exclude time-series effects of the 80s, from our analysis.

The dataset includes one table with the following variables:

* mpg: The fuel economy of the car in terms of miles travelled per gallon of gasoline | *Dependent Variable*
* horsepower: Horsepower is a measure of power the engine produces | *Main Independent/Explanatory Variable*
* cylinders: The number of cylinders in the car's engine
* displacement: The volume of air displaced by all the pistons of a piston engine
* weight: The total weight of the car
* acceleration: The time in seconds it takes for the car to reach 60 miles per hour
* model year: The year (in the 20th century) the car model was released. For example: 75 means the car was released in 1975.
* origin: The region where the car was manufactured. 1 - USA. 2 - Europe. 3 - Japan
* car name: The name of the car model.

For the analysis, I will use mpg as the dependent variable and mainly horsepower as the independent variable to see, how they are related to each other, based on the given dataset. My first intuition is, that mpg and horsepower are negatively related, which means, that if a car in the 70s has more horsepower, its fuel consumption (mpg) will also increase (in other words, it can travel less distance with the same amount of fuel), because to achieve more power, more resources should be needed. On contrary, an engine which has more horsepower may mean, that it uses the fuel more efficiently, thus for the same amount of fuel, it can give back to use more power. In any case, I am sure there should be some kind of relationship between them and correlation of 0 (no relationship) should not happen!

If we look at **App. 1a**, the mean of the mpg variable (21.084) is higher than the 50th percentile, the median (20), due some high values, thus the distribution is right skewed, as it also can be seen on the **App. 1b** histogram. This variable will be changed into a logarithmic form, so on further analysis, we will see the relative change of it and on **App. 1c** it can be seen, that it has a more bell-shaped curve.

**Model**

**Unconditional Linear Regression**

As my first model, I choose a simple **log-level** regression between the ln(mpg) and horsepower variables, without any further conditionals at the moment, to see solely their relationship. This and further charts are made using ggplot2 in python. Looking at the **App. 2** scatter plot chart, we can see that there is a negative relationship between them. It means, that on average, if a car has more horsepower, it will be able to travel **less** miles, with the same amount of fuel in other words, my first intuition seems to be correct. On the right side of the graph, the dots are sparser, as it seems during the 70s, not many cars have been sold in the USA with a horsepower of more than 190-200. On **App. SUM**, horsepower is significant, as it has a p-value of less than 1%, **for every 10 more horsepower** a car has, **it can travel 6% less miles** with the same amount of fuel.

**Conditional Multiple Linear Regression**

Of course, the horsepower of the engine is not the only attribute a car has. It also has a weight (as it is the American market and the fuel consumption is also in mpg, we assume, that weight is measured in pounds (1pound ~ 0.45kg)) and the engine can have different numbers of cylinders and different air displacement value. Air displacement is based on the numbers of cylinders in the engine, so we only use the former one to avoid biases between the conditional variables. The next multiple linear regression will include horsepower (like previously), displacement and weight. On **App. SUM** the **R2 for this regression is 0.824**, which means that this regression approximates the actual data better, than our previous simple linear regression. We can see a big change, but still a significant value for horsepower at a -0.002 for every additional horsepower (**for every 10 more horsepower** a car has, **it can travel 2% less miles** with the same amount of fuel). While displacement is not that significant, weight is. **For every 100 more pounds** of weight, **the car can travel 2% less miles**, which is logical, as more power is needed to move heavier objects. This number can be seen more precisely on **App. 2b.**

**Conditional Regression Based on Origin**

The United States of America has free market, their cars aren’t solely from their own country, but the vehicles can be imported from Europe and from Japan as well. Taking a look at **App. 2c**, the previously shown scatterplot is now updated, where the origin of the vehicle can be seen as well. The non-USA cars are mostly grouped on the left side of the plot, which means these cars tends to have a better fuel consumption, but how much? Using USA origin as the reference (or left-out) category, on **App. SUM**, both European and Japanese import cars have significantly better fuel consumptions, than US cars. European and Japanese import cars are tend to be able to **travel 11.5% and 17.5% more miles** respectively, with the same amount of fuel, than American cars.

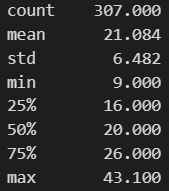
**Generalization and external validity**

**Causal interpretation**

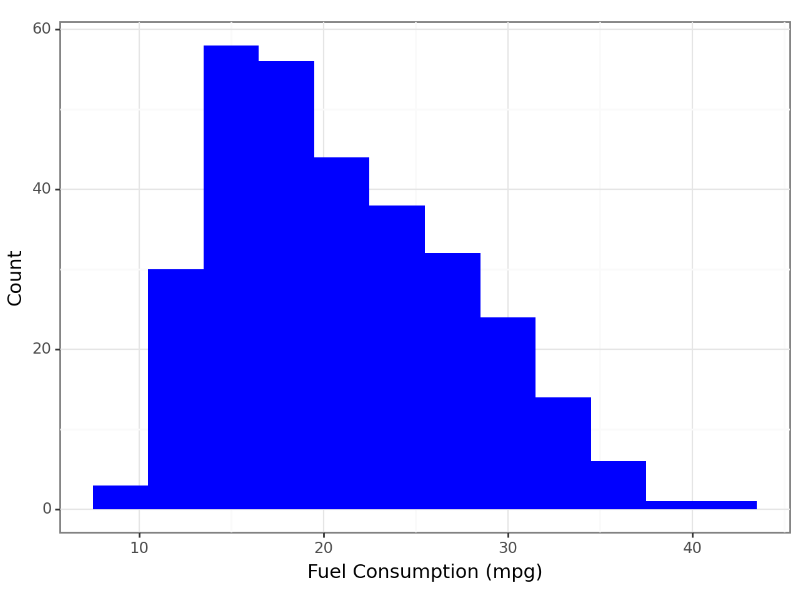
**Conclusion**

**Appendix**

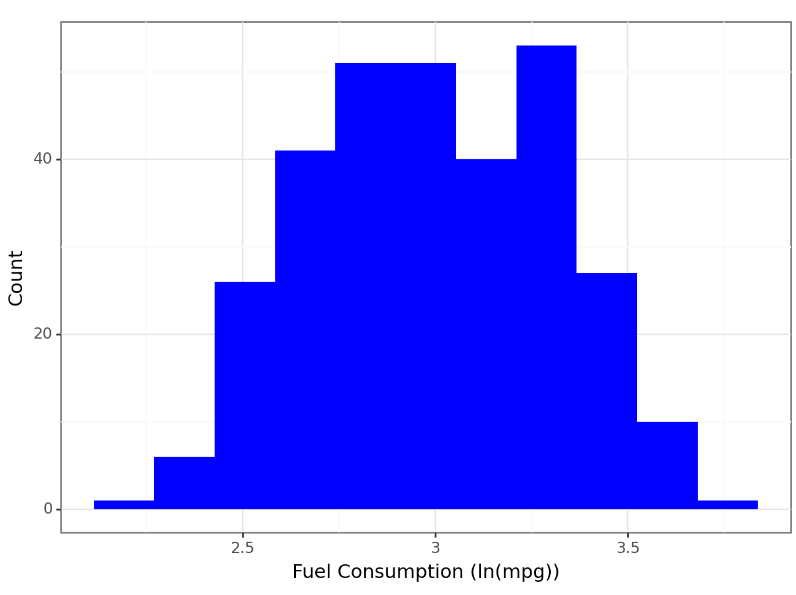
**App. 1a**



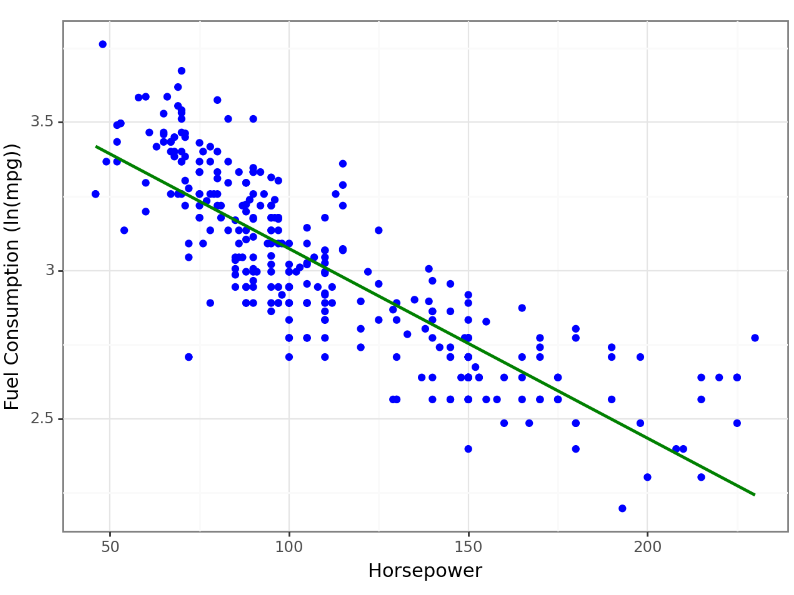
**App. 1b**



**App. 1c**



**App. 2a**



**App. 2b**

A képen szöveg, képernyőkép, menü, Betűtípus látható

Automatikusan generált leírás

**App. 2c**

A képen diagram, sor, szöveg, Diagram látható

Automatikusan generált leírás

1. <https://www.voanews.com/a/german-finance-minister-casts-doubt-on-2030-coal-exit/7337035.html> [Accessed: 12/18/2023] [↑](#footnote-ref-1)