Project III
Linear regression

The given dataset `vw_data.csv` provides information on used cars that were advertised on the e-commerce platform Exchange and Mart in the UK in 2020. The dataset is restricted to models by the manufacturer Volkswagen (VW). It is an extract of a larger data set available on `kaggle.com`.

Additionally to the model name, the data set contains the following eight variables:

- `price`—the price of the cars, in 1000 GBP (£);
- `year`—the year that the car was first registered in;
- `mileage`—the total distance (in 1000 miles) the car has been driven;
- `mpg`—the distance (in miles) the car can travel with one gallon (imperial) of fuel;
- `fuelType`—the type of fuel the car consumes;
- `engineSize`—the size of the car's engine in litres;
- `tax`—the amount of the annual tax (Vehicle Excise Duty) to be paid for the car;
- `transmission`—the type of gearbox the car has.

Tasks

1. In this project, the unit for measuring fuel consumption shall be l/(100 km) (litres per 100 km). Convert the variable `mpg` to this new unit.

2. Use the variable `year` to calculate the cars' age, and use the obtained values to define the new variable `age`. Replace the variable `year` with `age` in the data set.

3. Build a linear regression model to predict `price` from the other covariates. Decide whether it is better to use `price` or log(`price`) as the response variable. Apply appropriate model selection techniques to choose a good set of explanatory variables, e.g. best subset selection based on Akaikes Information Criterion (AIC) or Mallow's C_p statistic.

4. Interpret the coefficients of the model and their statistical significance, provide confidence intervals for the regression parameters, and evaluate the goodness of fit.

Submission

Submission of the report and the corresponding program code (executable and commented) is due on **Friday, 17 June 2022, at 08:30 a.m.** All relevant files must be uploaded to Moodle.