

Data Mining

Teacher: Annamaria Guolo

Practical assessment: June 6, 2018

INSTRUCTIONS: The examination takes 2 hours and 30 minutes. Write a brief report explaining the data analysis performed and the results. Provide a printed copy of the report. Remember to write your *i*) name and surname, *ii*) enrolment number, *iii*) date. Reports without the last three information will not be evaluated.

In the following you can find a description of the data and a brief guide for the analysis to be carried out. The use of the material of the course (slides, notes) is allowed. Internet is not allowed.

Dataset `cars`: data refer to the characteristics of some cars.

- `Type`: Type of car (Small, Sporty, Compact, Midsize, Large, Van)
- `Price`: Price in thousands of dollars
- `MPG.city`: City MPG (miles per US gallon)
- `MPG.highway`: Highway MPG
- `AirBags`: Air Bags presence (none, driver only, or driver & passenger)
- `DriveTrain`: Drive train type (rear wheel, front wheel or 4WD)
- `Cylinders`: Number of cylinders
- `EngineSize`: Engine size (litres)
- `Horsepower`: Horsepower
- `RPM`: RPM (revs per minute at maximum horsepower)
- `Rev.per.mile`: Engine revolutions per mile (in highest gear)
- `Man.trans.avail`: manual transmission version available? YES/NO
- `Fuel.tank.capacity`: Fuel tank capacity (US gallons)
- `Passengers`: capacity in terms of number of persons
- `Length`: Length (inches)
- `Wheelbase`: Wheelbase (inches)
- `Width`: Width (inches)
- `Turn.circle`: U-turn space (feet)
- `Rear.seat.room`: Rear seat room (inches)

- `Luggage.room`: Luggage capacity (cubic feet)
- `Weight`: Weight (pounds)
- `Origin`: Origin of the car? non-USA / USA

The aim of the analysis is the evaluation of the variables associated to the car price.

1. Consider the dataset composed by `Price`, `MPG.city`, `Horsepower`, `Origin`, `AirBags`. Construct the most appropriate model for the purpose of the analysis. Insert in the report the outputs from R and the graphical evaluation of the model/models that are considered most useful in order to explain the analysis and the results.
2. Consider all the variables in the dataset. Construct the most appropriate model for the purpose of the analysis. Insert in the report the outputs from R and the graphical evaluation of the model/models that are considered most useful in order to explain the analysis and the results.

If needed, please report the seed used in your analyses.