Report Project 2

Data Analysis and Statistical Modeling

Prof Isabel Rodrigues

Uma imagem com captura de ecrã, design, círculo, texto

Descrição gerada automaticamente

**Grupo 1**

João Matos nº98949

Ana Pinto nº102949

Marina Nóbrega nº103880

Manuel Dias nº96056

Maria Freitas nº96757

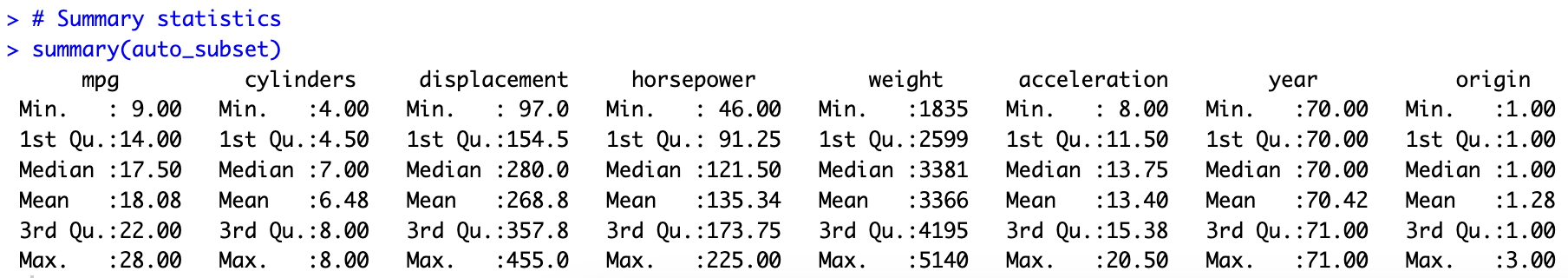
**Library**: ISLR

**Data frame**: Auto

**Subset**: observation 1 to 50

**Variables**: all except name:

* mpg – miles per gallon
* cylinders – number of cylinders between 4 and 8
* displacement – engine displacement (cu. Inches)
* horsepower – engine horsepower
* weight – vehicle weight (lbs.)
* acceleration – time to accelerate from 0 to 60 mph (sec.)
* year – model year (modulo 100)
* origin – (origin of the car (1. American, 2. European, 3. Japanese)

**Summary statistics**

**Uma imagem com texto, Tipo de letra, branco, algebra

Descrição gerada automaticamente**

Uma imagem com texto, captura de ecrã, Tipo de letra, branco

Descrição gerada automaticamente

**Total Variance**

Uma imagem com Tipo de letra, texto, branco, Gráficos

Descrição gerada automaticamente

**Summary Plots**







**Uma imagem com texto, Tipo de letra, captura de ecrã, branco

Descrição gerada automaticamente2.A)**

To find the best subset of regressors, we applied the regression model while until we got the ones that we considered useful.

Uma imagem com texto, Tipo de letra, captura de ecrã, branco

Descrição gerada automaticamente

Uma imagem com texto, Tipo de letra, branco, algebra

Descrição gerada automaticamenteUma imagem com texto, Tipo de letra, branco, captura de ecrã

Descrição gerada automaticamente

Uma imagem com texto, Tipo de letra, branco, captura de ecrã

Descrição gerada automaticamente

At the end, cylinders, weight, and acceleration are the selected ones. We chose these because it would allow us to work with less predictors. The adjusted r^2 value of the last two iterations is similar and lower in the last iteration.

After fitting a regression model to explain the mpg variable using the predictors we just selected we get the values:

r² = 0.891092607413735

r²adj = 0.8839899513755

**2.B)**

For this regression we are using p = 3 predictors for n = 50 observations

Searching for possible influential/leverage observations we get this:



For mpg, the possible leverage observations are 29, 20 and 14. The two observations with highest cook’s distance are 14 and 29, so the more possible influential observations are 29 and 14.

**2.C)**

Calculating the 97.5% Confidence Interval and Prediction Interval for the expected values of the responses for observations **14** and **31** we get:

