

Data Mining

Teacher: Annamaria Guolo

Practical examination: June 22, 2022

INSTRUCTIONS

Analyze the data described below according to the two questions listed. Prepare a brief report to summarize the main findings of your analysis. It is not necessary to include R commands in the report. Please indicate the seed you use for the analysis, if the seed is necessary.

Include your name, surname and enrollment number in the report. Reports without this information will not be corrected. Please, add the date. Print a copy of your report and return it to the teacher.

Dataset included in `travel.RData` refers to a study about the relationship between cost of trips (in dollars) and characteristics of the trip and of the respondent.

- `Q`: Calendar quarter of trip
- `Distance`: Distance traveled in trip (km)
- `Duration`: Number of nights spent on trip
- `Reason`: Reason of the trip (vacation/visit)
- `Age`: Age of the respondent (in classes of years)
- `Gender`: Gender of the respondent (Male/Female)
- `Others`: Number of persons accompanying the respondent during the trip
- `Mode`: Mode of transportation (car/plane)
- `Cost`: Total cost of the trip (logarithm of dollars)
- `Prov`: Trip province of origin
- `Income`: Income of the respondent (classes of thousands of dollars)

FIRST QUESTION.

Consider the dataset composed by `Cost`, `Duration`, `Reason`, `Income`, `Mode`. Construct the most appropriate model for the purpose of the analysis. Which variables are associated to the cost of trip?

SECOND QUESTION.

Consider all the variables in the dataset. Construct the most appropriate model for the purpose of the analysis. Which variables are associated to the cost of trip?