

## SAS

### Practical exercise

You should get a database (ASCII, EXCEL, SPSS or SAS) with a set of variables and 5000 observations as a minimum. Namely, approximately every observation will have an identification variable ID (it can be the name of a person, a brand, a code, etc.), and 7 additional variables: a binary categorical variable, a polyatomic categorical variable (with two or more categories) and 5 numerical variables. It will be appreciated the difficulty and originality of the dataset. If the chosen dataset does not meet the requirements, it will be punished in the mark.

With this data set, you will create a SAS script with name **Codigo\_SAS.sas** which will the following information at the beginning as a commentary:

- Name/s and surname/s of student/s.
- Original dataset source and link where it is available if possible.

Moreover, the script will have the program code necessary to solve the following items:

1. Import the dataset to SAS and create the corresponding SAS dataset.
2. Give labels and formats to the variables of the dataset.
3. Carry out a descriptive analysis to check that there are no values of any variable out of range or extreme observations which could be considered outliers. Indicate in the script if everything was ok or not.
4. If you find any error in the previous section or extreme observations, you should decide what to do with them (remove? Convert them to missing? Keep them?) Write down what you did and why.
5. Obtain a complete description of three to five of the variables that seem most relevant
6. Define groupings (make a categorization) of one or more of the original numerical variables.
7. Use the IML language to calculate the correlations between the numerical variables.
8. Carry out some type of analysis of the continuous variables crossing them with some categorical variable or with some of the categorizations defined in the previous section, after giving them the corresponding format.
  - a. This analysis must include a statistical proc and a graphical proc as minimum.
  - b. Indicate before the execution of the PROCs, as a commentary, what is the purpose of the analysis you wish to perform.
9. Carry out other analyzes that may interest you. Indicate prior to the execution of the PROCESSES, in the form of a comment, what is the purpose of the analysis you wish to perform
10. Create a macro to repeat some of the processes you have done, for example, do the same analysis for different countries / cities / etc, using macros.

Files that you should include in a zip file called **Practica\_SAS.ZIP**:

- a) Codigo\_SAS.sas, the SAS script
- b) **INFORME.pdf** (with the results obtained in the various sections, as well as your comments), and
- c) the original dataset