

COMPUTATION IN STATISTICS AND OPTIMIZATION

COURSE 2017-18

LANGUAGE: SAS

ENUNCIADO DE LA PRÁCTICA

Get a database (ASCII, EXCEL, SPSS o SAS) in which there is a series of variables on a minimum of 5000 individuals. Specifically, approximately every observation will consist of a variable Identifier ID (can be a person's name, a brand, a code, etc.), and will be accompanied by at least seven more variables: A categorical binary variable, a categorical polytomic variable (with more than 2 categories) and five variables of numerical type. The difficulty and originality of the chosen data will be assessed.

Here are some possible sources (links to the respective websites):

- 1) Survey of Family Budgets (Base 2006) of the National Institute of Statistics of Spain
- 2) Survey of Family Budgets - Welfare module of the National Institute of Statistics of Spain
- 3) Survey of Salary Structure of the National Statistics Institute of Spain
- 4) Hospital Morbidity Survey of the National Statistics Institute of Spain
- 5) Microdata from the Center for Sociological Research (CIS)
- 6) Microdata from the National Institute of Statistics and Geography of Mexico

Before starting the work the student must communicate which is the chosen database (description and / or link to your location) and wait for the approval of the subject's teaching staff. From this set of data, it will be necessary to create a script named Code_SAS.sas that will contain as first lines and in the form of a comment:

- The student's name and surnames.
- Source where you have obtained the data with the link if there is one.
- The instructions necessary to solve the sections indicated below:

- 1) Read the data you have obtained from SAS to create a database in SAS format.
- 2) Assign variable and value labels to the variables in your file
- 3) Make a descriptive to verify that it does not have values of the variables out of range or extreme observations that we can consider outliers. Indicate in the script itself if everything has gone good or not.
- 4) If you find any type of error in the previous section or very extreme observations, decide what to do with these observations (eliminate them? Make them missing? Keep them?) Leave record of what you decide and do it.
- 5) Get a complete descriptive of between three and five of the variables that seem most relevant
- 6) Define groupings of one or more of the original numeric variables
- 7) Do some kind of analysis of some continuous variables, crossing them with some variables categorical or with the variables grouped in the previous section, duly labeled.
 - a) This analysis should include at least one statistical PROC and one graphic PROC.

- b) Indicate beforehand the execution of the PROCESSES, in the form of a comment, which is the objective of the analysis you want to perform.
- 8) Make other analyses that may interest you. Indicate before the execution of the PROCESSES, in form of comment, what is the purpose of the analysis you want to perform.
- 9) Documents to be delivered in a file called Practica_SAS.ZIP:
 - a)Codigo_SAS.sas,
 - b) INFORME.pdf (containing the results obtained in the various sections, as well as their comments), and
 - c) the original database.