

COMPUTACIÓN EN ESTADÍSTICA Y OPTIMIZACIÓN

CURSO 2018-19

Parte SAS

ENUNCIADO DE LA PRÁCTICA

Conseguir una base de datos (ASCII, EXCEL, SPSS o SAS) en la que se midan un conjunto de variables sobre un mínimo de 5000 individuos. En concreto, de forma aproximada cada observación constará de una variable identificadora ID (puede ser un nombre de persona, una marca, un código, etc.), y estará acompañada de al menos siete variables más: Una variable categórica binaria, una variable categórica politómica (con más de 2 categorías) y cinco variables de tipo numérico. Se valorará la dificultad y la originalidad de los datos elegidos. Si la base de datos elegida no cumple con estos requerimientos, penalizará en la nota final.

A partir de dicho conjunto de datos, será necesario crear un script con nombre **Codigo_SAS.sas** que contendrá como primeras líneas y en forma de comentario:

- El nombre y apellidos del estudiante.
- Fuente de donde ha obtenido los datos con el enlace si lo hay.

Además el script contendrá las instrucciones necesarias para resolver los apartados que se indican a continuación:

1. Leer los datos que ha obtenido desde SAS para crear una base de datos en formato SAS.
2. Asignar etiquetas de variable y de valores a las variables de su archivo.
3. Realizar un descriptivo para comprobar que no tenga valores de las variables fuera de rango u observaciones extremas que podamos considerar outliers. Indicar en el propio script si todo ha salido bien o no.
4. Si encuentra algún tipo de error en el apartado anterior o bien observaciones muy extremas, decidir qué hacer con estas observaciones (eliminarlas? Convertirlas en missing? conservarlas?) Dejar constancia de lo que decida y hacerlo.
5. Obtener un descriptivo completo de entre tres y cinco de las variables que le parezcan más relevantes
6. Definir agrupamientos (realizar una categorización) de una o más de las variables numéricas originales.
7. Utilizar el lenguaje IML para calcular las correlaciones entre las variables numéricas.
8. Hacer algún tipo de análisis de algunas variables continuas, cruzándolas con alguna/s variables categóricas o bien con la/las variables agrupadas en el apartado anterior, debidamente etiquetadas.
 - a) Este análisis deberá incluir como mínimo un PROC estadístico y un PROC gráfico.
 - b) Indicar previamente a la ejecución de los PROCESOS, en forma de comentario, cuál es el objetivo del análisis que desea realizar.
9. Hacer otros análisis que os puedan interesar. Indicar previamente a la ejecución de los PROCESOS, en forma de comentario, cuál es el objetivo del análisis que desea realizar.
10. Crea alguna macro para repetir alguno de los procesos que has realizado, por ejemplo hacer un mismo análisis para distintos países/ciudades/etc, usando macros.

Documentos a entregar en un archivo llamado **Practica_SAS.ZIP**:

- a) Codigo_SAS.sas,
- b) **INFORME.pdf** (que contenga los resultados obtenidos en los diversos apartados, así como sus comentarios), y
- c) la base de datos original.