

Grau Interuniversitari (UB-UPC) d'Estadística

Data d'entrega: dimecres 25 de gener de 2017

Per a la resolució d'aquest exercici suposarem que tots tenim les dades en la carpeta sers: **SERS\EXERSAS**. El conjunt de dades a utilitzar pot ser el mateix que el que heu fet servir per a la pràctica d'R.

A partir del vostre conjunt de dades, caldrà crear un script amb nom **Codi\_SAS.sas** que contindrà com a primeres línies i en forma de comentari:

- Els nom i cognoms dels integrants del grup.
- Font d'on heu obtingut les dades amb l'enllaç si n'hi ha.
- Les instruccions necessàries per resoldre els apartats que s'indiquen tot seguit:
  1. Llegir les dades que heu obtingut des de SAS per crear una base de dades en format SAS.
  2. Assignar etiquetes de variable i de valors a les variables del vostre arxiu. Com a mínim heu de definir dos formats per a les etiquetes de valor. Feu servir un fitxer de formats que quedí gravat en disc dur.
  3. Caldrà realitzar un descriptiu per a comprovar que no tingueu valors de *missings*, fora de rang o observacions extremes que puguem considerar *outliers*. Indiqueu en el propi script si tot ha sortit bé.
  4. Si trobeu algun tipus d'errada en l'apartat anterior o bé observacions molt extremes, decidiu que feu amb aquestes observacions (les elimineu? Les passeu a *missing*? Les conserveu?) Deixeu constància del que decidiu i feu-lo.
  5. Obtenir un descriptiu d'alguna de les variables que us semblin més rellevants (almenys una categòrica i una numèrica).
  6. Definir agrupaments d'una o més de les variables numèriques originals. Feu servir formats per a l'agrupació.
  7. Feu algun tipus d'anàlisi d'algunes variables numèriques, creuant-les amb alguna/es variables categòriques o bé amb la/les variables agrupades en l'apartat anterior, degudament etiquetades.
    - Aquesta anàlisi haurà d'incloure com a mínim un PROC estadístic i un PROC gràfic.
    - Indiqueu prèviament a l'execució dels PROCS, en forma de comentari, quin és l'objectiu de l'anàlisi que voleu realitzar.
    - Haureu de fer tests estadístics (almenys un de la Chi-quadrat i un de la de la t-Student per una mostra) per analitzar inferencialment algunes relacions entre variables. Indiqueu clarament el test, l'estadístic, el p-valor, la decisió i la conclusió pràctica.
  8. Feu altres anàlisis que us puguin interessar. Indicar prèviament a l'execució dels PROCS, en forma de comentari, quin és l'objectiu de l'anàlisi que voleu realitzar.
  9. Afegeix a la teva base de dades dues noves variables fent simulació (utilitza com a valor de llavor el teu NIUB i determina els valors dels paràmetres):
    - a. Una normal amb probabilitat de valors negatius entre el 5% i el 9%
    - b. Una exponencial amb probabilitat de valors més grans que 3 entre el 90% i el 95%
  10. A partir de les variables anteriors, crea una nova variable que prengui el valor 1 si la suma d'ambdues variables és més gran que 3 i zero en cas contrari. Obté la distribució en percentatge d'aquesta variable.
  11. Crea una nova variable a partir d'una de les variables de la teva base de dades i assigna-li valors *missing* en aquelles posicions on el valor de la variable obtinguda a l'apartat anterior sigui 1. Quants *missings* presenta aquesta nova variable?
  12. Per a la variable transformada segons l'apartat anterior obté la seva mediana, desviació estàndard, el coeficient de variació, l'interval de confiança al 95% i el percentil 5.

En el vostre script haureu d'incloure instruccions pertinents per poder generar un fitxer en format pdf, anomenat **Llistat\_SAS.pdf** amb els resultats dels apartats 5, 7 i 8.

Documents a lliurar en un arxiu anomenat **Practica\_SAS.ZIP**

Script d'instruccions de SAS **Codi\_SAS.sas**

Arxiu pdf **Llistat\_SAS.pdf**

Arxiu de dades originals i arxiu de formats creats