

Tarea No. 1 Curso: Introducción Al Análisis Multivariado-IAM UnalMed Semestre: 2022-01

Par esta tarea cada estudiante (o pareja de estudiantes) trabajará con una base de datos que contenga un mínimo de 4 variables cuantitativas y un máximo de 7, además la base debe contener una o dos variables cualitativas adicionales para realizar análisis descriptivos por grupos de las variables cualitativas asociadas a la base de datos. El número de observaciones de la base de datos deber ser máximo 500 y tengan presente que n sea mayor que p . Para la base de datos en cuestión deben responder las siguientes preguntas en un informe que será entregado en formato .pdf.

1. Realizar un resumen descriptivo de sus datos de la forma más adecuada posible para lo cuál se le sugiere que utilice ayudas como las siguientes:
 - (a) Ayudas gráficas como por ejemplo: Diagramas de dispersión de variables, Gráficas de cajas y bigotes o box-plot, histogramas, etc. etc.
 - (b) Para 3 de las variables de su base, comente los aspecto más relevantes de los datos. Comente sobre medidas descriptivas relevantes de esas variables.
 - (c) Para las mismas 3 variables consideradas en el punto (b) realice análisis descriptivos por categorías de una de sus variables cualitativas de sus datos. Igualmente pueda aca ayudarse de gráficos como box-plot.
2. Obtenga el vector de medias muestrales $\bar{\mathbf{x}}$, las matrices de varianzas-covarianzas muestrales \mathbf{S} y \mathbf{S}_n y la matriz de correlaciones muestrales \mathbf{R} de los datos. Comente algunas entradas relevantes de dichos resultados. Puede ayudarse con algunos gráficos apropiados.
3. Calcule la varianza total y la varianza generalizada de los datos. Comente.
4. Calcule la distancia de Mahalanobis entre cada dato u observación y el centroide de los datos (vector de medias). Qué observaciones están a mayor distancia? Qué Observaciones están a menor distancia? Se podrá usar para esta matriz de datos la distancia euclideana? Porqué sí o porqué no?.
5. Para dos de las variables de su tabla de datos realice una evaluación de normalidad univariada para cada una de ellas, siguiendo las distintas opciones que sean de su conocimiento, como por ejemplo ayudas gráficas o pruebas formales que conozca. Comente los resultados de cada evaluación.