



Pontificia Universidad Javeriana
Tarea No.1 de Análisis Multivariado
Fecha de entrega (física): Jueves 15 de Agosto de 2024

1. **(2.0 pts)** En el archivo **anexo1.csv** contiene los datos de una muestra de vectores aleatorios de 3 componentes, X_1, X_2, \dots, X_{40} .
 - a. Evalúe la normalidad multivariada de la muestra dada.
 - b. Si el vector de medias poblacionales es $\mu = [0.1, -0.2, 0.05]^t$ y S es la matrix de varianzas-covarianzas muestrales. ¿Cuál es la distribución aproximada de $40\left(\bar{X} - [0.1, -0.2, 0.05]^t\right)^t S^{-1}\left(\bar{X} - [0.1, -0.2, 0.05]^t\right)$?
 - c. Usando la **distancia cuadrada generalizada** establezca si existen valores atípicos.
2. **(3.0 pts)** Los datos Protein del paquete MultBiplotR contiene información sobre datos nutricionales de 9 diferentes fuentes de proteínas para los habitantes de 25 países europeos alrededor de 1970:
 - **RedMeat**: Consumo de proteínas provenientes de carnes rojas
 - **WhiteMeat**: Consumo de proteínas provenientes de carnes blancas;
 - **Eggs**: Consumo de proteínas del huevo;
 - **Milk**: Consumo de proteínas de la leche;
 - **Fish**: Consumo de proteínas provenientes del pescado;
 - **Cereals**: Consumo de proteínas procedentes de cereales;
 - **Starch**: Consumo de proteínas provenientes de carbohidratos;
 - **Nuts**: Consumo de proteínas procedentes de cereales, frutos secos y semillas oleaginosas;
 - **FruitVeg**: Consumo de proteínas procedentes de frutas y verduras.

Estos datos fueron colectados inicialmente para entender las diferencias nutricionales entre los países europeos.

- a. Determine y analice el vector de medias y la matriz de covarianzas muestrales para las diferentes regiones.
- b. Calcule la media de las variables por regiones ¿qué puede decir al respecto?
- c. Intente construir grupos de países usando representaciones pictóricas (gráficos de estrellas o caras de Chernoff).
- d. Utilice las herramientas de gráficas más adecuadas para verificar normalidad multivariada.
- e. Realice la prueba de Mardia para verificar las hipótesis:

H_0 : Los datos provienen de una población Normal Multivariada

H_1 : Los datos NO provienen de una población Normal Multivariada

- f. Verifique si hay outliers (multivariados) e identifíquelos.
- g. Pruebe las hipótesis

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$$

$$H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$$

donde $\boldsymbol{\mu}_0 = [9, 7, 2, 15, 5, 30, 4, 3, 4]^\top$

- i. de forma univariada
- ii. de forma multivariada.

comente los resultados