

Análisis Estadístico Multivariado | UTDT

Problem Set 4

Alejandra Clemente

Fiona Franco Churruarín

Segundo Trimestre 2023

Ejercicio 1

Construya un test de normalidad multivariada, y aplíquelo sobre los datos correspondientes al primer ejercicio de la práctica 2. ¿Puede concluir que el vector de medias de los datos es distinto del vector nulo?

Recordemos del ejercicio 1 de la práctica 2 que el archivo *ine_dta* es una base de datos que contiene los gastos promedio, en euros, de los hogares españoles, por grandes rubros y comunidad autónoma, correspondientes a los relevamientos de la encuesta de presupuestos familiares del año 2005 realizada por el Instituto Nacional de Estadísticas de España (INE).

```
ine <- haven::read_dta("C:/Users/fiona/Dropbox/Materias MAECO/Análisis Estadístico Multivariado/Practica
mardia(ine[,2:ncol(ine)])
```

```
## $mv.test
##           Test Statistic p-value Result
## 1      Skewness    302.0658  0.2459   YES
## 2      Kurtosis      0.2887  0.7728   YES
## 3  MV Normality      <NA>    <NA>   YES
##
```

```
## $uv.shapiro
##           W      p-value UV.Normality
## alybnh    0.7665 5e-04    No
## vestcal    0.9284 0.1819   Yes
## vivagelo    0.9725 0.8428   Yes
## mobymant    0.982  0.9682   Yes
## salud      0.9499 0.4232   Yes
## transp     0.9292 0.1881   Yes
## comu       0.9715 0.8249   Yes
## ocio       0.9851 0.9872   Yes
## educ       0.9489 0.4083   Yes
## esparc     0.9473 0.3837   Yes
## otros      0.9468 0.3765   Yes
```

```
k <- 11
N <- 18
(6*((k+1)*(N+1)-6))*302.0658/(k+1)*(N+1)*(N+3)
```

```
## [1] 13378192
```

```
0.2887*sqrt(8*k*(k+2)/N)+k*(k+2)*(N-1)/(N+1) #para la igualdad con STATA
```

```
## [1] 130.2489
```

```
qchisq((1-0.2459),(11*12*13)/6)
```

```
## [1] 302.0691
```

Ahora realizamos un test de significatividad sobre el vector de medias:

```
ICSNP::HotellingsT2(ine[,2:ncol(ine)])
```

```
##  
## Hotelling's one sample T2-test  
##  
## data: ine[, 2:ncol(ine)]  
## T.2 = 70.918, df1 = 11, df2 = 7, p-value = 4.258e-06  
## alternative hypothesis: true location is not equal to c(0,0,0,0,0,0,0,0,0,0,0)
```

Rechazo la hipotesis nula de que el vector de medias sea el vector nulo.

Ejercicio 2

Generar una muestra de 100 observaciones de una población normal multivariada teniendo en cuenta los siguientes vectores de parámetros:

$$\mu = \begin{bmatrix} 2 \\ 0.45 \\ 0.23 \\ 0.54 \\ 0.12 \\ 0.63 \\ 0.66 \\ 0.32 \end{bmatrix}, \quad \sigma = \begin{bmatrix} 0.67 \\ 0.89 \\ 0.56 \\ 0.90 \\ 0.56 \\ 0.34 \\ 0.76 \\ 0.13 \end{bmatrix}$$

- Efectúe un test para verificar la normalidad de la base de datos construída.
- ¿Existe evidencia empírica suficiente que permita concluir que el vector de medias es estadísticamente distinto del vector nulo?
- Efectúe un test para contrastar la hipótesis acerca de la igualdad de las medias de las variables que componen la matriz de datos.
- Contraste la esfericidad de la distribución de los datos.
- ¿Se verifica estadísticamente la diagonalidad de la matriz de varianzas y covarianzas en los datos?

Ejercicio 3

Repetir el ejercicio anterior, considerando en este caso la siguiente estructura:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \quad X_1 \sim N(\mu_1, \Sigma_1), \quad X_2 \sim N(\mu_2, \Sigma_2), \quad X_3 \sim N(\mu_3, \Sigma_3).$$

siendo

$$\mu_1 = \begin{bmatrix} 2 \\ 0.45 \\ 0.23 \\ 0.54 \\ 0.12 \\ 0.63 \\ 0.66 \\ 0.32 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0.67 \\ 0.89 \\ 0.56 \\ 0.90 \\ 0.56 \\ 0.34 \\ 0.76 \\ 0.13 \end{bmatrix}, \quad \pi_1 = 40\%$$

$$\mu_2 = \begin{bmatrix} 3 \\ 1.45 \\ 1.23 \\ 1.54 \\ 1.12 \\ 1.63 \\ 1.66 \\ 1.32 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 1.67 \\ 1.89 \\ 1.56 \\ 1.90 \\ 1.56 \\ 1.34 \\ 1.76 \\ 1.13 \end{bmatrix}, \quad \pi_2 = 50\%$$

$$\mu_3 = \begin{bmatrix} 4 \\ 2.45 \\ 2.23 \\ 2.54 \\ 2.12 \\ 2.63 \\ 2.66 \\ 2.32 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 2.67 \\ 2.89 \\ 2.56 \\ 2.90 \\ 2.56 \\ 2.34 \\ 2.76 \\ 2.13 \end{bmatrix}, \quad \pi_3 = 10\%$$

- Comente lo observado en contraste con el ejercicio anterior.
- Efectúe similares contrastes pero considerando cada partición de X por separado.
- Proponga soluciones al problema observado, incluso considerando la eventualidad de que existan valores atípicos en la muestra que conlleven al mismo. Aplica la/s soluciones sufridas y comentar brevemente el resultado encontrado.