

## **Trabajo 2**

**Ivan Santiago Rojas Martinez**

Estudiante de Pregrado en Estadística

Docente

**Rene Iral Palomino**

Asignatura

**Introducción al Análisis Multivariado**



Sede Medellín  
Octubre 21 de 2023

# Índice

<b>1</b>	<b>Parte A</b>	<b>2</b>
<b>2</b>	<b>Parte B</b>	<b>2</b>
2.1	Análisis discriminado sexo Masculino: . . . . .	4
2.2	Análisis discriminado sexo Femenino: . . . . .	5

## Índice de figuras

## Índice de cuadros

## Trabajo 2

### 1 Parte A

### 2 Parte B

Para todos los efectos el vector  $X = (P_1, P_7, P_{16}, P_{22}, P_{25}, P_{27}, P_{29}, P_{38})$  contiene las variables continuas de su base de datos. Por notación sea el respectivo  $\mu = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8)$  vector de medias y  $\Sigma$  su matriz de covarianzas.

Se procede a tomar la muestra aleatoria con la cedula **1020479466** y a seleccionar las variables numéricas.

```
library(splitstackshape)
uno <- read.table("Data/base.txt", header = TRUE)
genera <- function(cedula){
  set.seed(cedula)
  aux <- stratified(uno, "CAT_IMC", 200/2100)
  aux
}

datos <- genera(1020479466)
x <- datos %>% select(P1, P7, P16, P22, P25, P27, P29, P38)
```

- 1. (10 pts.)** Sea  $\mu_0 = (66.1, 58, 81.6, 37, 47, 25, 19.2, 167)'$ . Pruebe la hipótesis:  
 $H_0: \mu = \mu_0$ . Debe especificar todas las condiciones y elementos para probar esta hipótesis.  
 Anexe los códigos en R usados.

Primero procederemos a verificar si el vector de variables se distribuye normal por medio de la prueba estadística Shapiro-Wilk de normalidad multivariada.

```
library(mvnormtest)
mu_0 <- c(66.1, 58, 81.6, 37, 47, 25, 19.2, 167)
mshapiro.test(t(as.matrix(x)))
```

```
##
## Shapiro-Wilk normality test
##
## data: Z
## W = 0.9307, p-value = 4.09e-08
```

Observando un  $ValorP = 4.09 \times 10^{-8}$ , podemos rechazar la hipótesis nula con un nivel de significancia de  $\alpha = 0.05$  lo que nos permite concluir que  $X$  no cumple normalidad multivariada. Basado en el **Teorema del limite central** sabemos que si se tiene:

$$\bar{\mathbf{X}}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \text{ y } \mathbf{Z}_n = \sqrt{n} (\bar{\mathbf{X}}_n - \boldsymbol{\mu})$$

Luego:

$$\mathbf{Z}_n \xrightarrow{d} N_p(\mathbf{0}, \Sigma).$$

con  $\Sigma > 0$ :

$$\tilde{\mathbf{Z}}_n = \Sigma^{-\frac{1}{2}} \sqrt{n} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} N_p(\mathbf{0}, I_p)$$

y

$$n (\bar{\mathbf{X}}_n - \boldsymbol{\mu})' S^{-1} (\bar{\mathbf{X}}_n - \boldsymbol{\mu}) \xrightarrow{d} \chi^2(p)$$

Bajo  $H_0$  cierta. El estadístico de prueba es:

$$\chi_0^2 = n (\bar{X} - \mu_0)' S^{-1} (\bar{X} - \mu_0)$$

Se define las siguientes pruebas de hipótesis:

$$H_0 : \mu = \mu_0 \text{ vs } H_a : \mu \neq \mu_0$$

Se procede a hallar el vector de medias muestral  $\bar{X}$  y la matriz de covarianzas muestral  $S$  en R.

```
xbar <- colMeans(x)
s <- cov(x)
n <- nrow(x)
p <- length(x)
mu_0 <- as.matrix(c(mu_0))
chi_0 <- as.numeric(n*(t(xbar-mu_0)) %*%solve(s) %*%(xbar-mu_0))
chi_0
```

```
## [1] 1603.695
```

Se plantea una región de rechazo de  $H_0$  dada por:

$$X_0^2 > X_\alpha^2(p)$$

```
qchisq(0.05, p, lower.tail = F)
```

```
## [1] 15.50731
```

Dado que la prueba nos arroja  $X_0^2 = 1603.695 > X_{0.05}^2(8) = 15.50731$  con un nivel de significancia de  $\alpha = 0.05$ . Se rechaza  $H_0$ , lo que nos permite concluir que existen diferencia entre el vector  $\mu$  y  $\mu_0$ .

- 2. (15 pts.)** Repita la hipótesis anterior, pero discriminando por Género. ¿Observa algún cambio en la conclusión? Debe especificar todas las condiciones y elementos para probar esta hipótesis. Anexe los códigos en R usados.

## 2.1 Análisis discriminado sexo Masculino:

Se procede a filtra los datos por el genero masculino.

```
hom <- datos %>% filter(SEXO == "Hom") %>% select(P1, P7, P16, P22, P25, P27, P29, P38)
```

Se plantea sus hipótesis de normalidad multivariada discriminando por el genero de hombres.

$$H_o : X_H \sim N_p(\mu, \Sigma) \text{ VS } H_a : X_H \not\sim N_p(\mu, \Sigma)$$

```
mshapiro.test(t(as.matrix(hom)))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Z
## W = 0.92315, p-value = 1.91e-06
```

Observando un  $ValorP = 1.91 \times 10^{-6}$ , podemos rechazar la hipótesis nula con un nivel de significancia de  $\alpha = 0.05$  lo que nos permite concluir que el vector de variables  $X$  para los hombres no se distribuye normal, por tanto basado en el teorema del limite central tenemos que el estadístico de prueba es:

$$\chi_0^2 = n (\bar{X}_{hom} - \mu_0)' S_{hom}^{-1} (\bar{X}_{hom} - \mu_0) \xrightarrow{d} \chi_\alpha^2(P)$$

Con sus respectivas hipotesis:

$$H_0 : \mu_{homb} = \mu_0 \text{ vs } H_a : \mu_{homb} \neq \mu_0$$

```
xbar <- colMeans(hom)
s <- cov(hom)
n <- nrow(hom)
p <- length(hom)
mu_0 <- as.matrix(c(mu_0))
chi_0 <- as.numeric(n*(t(xbar-mu_0)) %*%solve(s) %*%(xbar-mu_0))
chi_0
```

```
## [1] 1403.088
```

Se plantea una región de rechazo de  $H_0$  dada por:

$$X_0^2 > X_\alpha^2(p)$$

```
qchisq(0.05, p, lower.tail = F)
```

```
## [1] 15.50731
```

Dado que la prueba nos arroja  $X_0^2 = 1403.088 > X_{0.05}^2(8) = 15.50731$  con un nivel de significancia de  $\alpha = 0.05$ . Se rechaza  $H_0$ , lo que nos permite concluir que existen diferencia entre el vector  $\mu_{homb}$  y  $\mu_0$ .

## 2.2 Análisis discriminado sexo Femenino:

Se procede a filtra los datos por el genero masculino.

```
muj <- datos %>% filter(SEX0 == "Muj") %>% select(P1, P7, P16, P22, P25, P27, P29, P38)
```

Se plantea sus hipótesis de normalidad multivariada discriminando por el genero de mujeres.

$$H_o : X_M \sim N_p(\mu, \Sigma) \text{ VS } H_a : X_M \not\sim N_p(\mu, \Sigma)$$

```
mshapiro.test(t(as.matrix(muj)))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  Z
## W = 0.93582, p-value = 0.001285
```

Observando un  $ValorP = 0.001285$ , podemos rechazar la hipótesis nula con un nivel de significancia de  $\alpha = 0.05$  lo que nos permite concluir que el vector de variables  $X$  para las mujeres no se distribuye normal, por tanto basado en el teorema del limite central tenemos que el estadístico de prueba es:

$$\chi_0^2 = n \left( \bar{X}_{muj} - \mu_0 \right)' S_{muj}^{-1} \left( \bar{X}_{hom} - \mu_0 \right) \xrightarrow{d} \chi_\alpha^2(\mathbf{P})$$

Con sus respectivas hipotesis:

$$H_0 : \mu_{muj} = \mu_0 \text{ vs } H_a : \mu_{muj} \neq \mu_o$$

```
xbar <- colMeans(muj)
s <- cov(muj)
n <- nrow(muj)
p <- length(muj)
mu_0 <- as.matrix(c(mu_0))
chi_0 <- as.numeric(n*(t(xbar-mu_0))%*%solve(s)%*%(xbar-mu_0))
chi_0
```

```
## [1] 2773.412
```

Se plantea una región de rechazo de  $H_0$  dada por:

$$X_0^2 > X_{\alpha}^2(p)$$

```
qchisq(0.05, p, lower.tail = F)
```

```
## [1] 15.50731
```

Dado que la prueba nos arroja  $X_0^2 = 2773.412 > X_{0.05}^2(8) = 15.50731$  con un nivel de significancia de  $\alpha = 0.05$ . Se rechaza  $H_0$ , lo que nos permite concluir que existen diferencia entre el vector  $\mu_{muj}$  y  $\mu_0$ .

- 3. (15 pts.)** Para el mismo vector  $\mathbf{X}$ , definido anteriormente, se sabe que  $\mathbf{X} \sim N_8(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Pruebe la hipótesis:  $H_0: 2\mu_1 - \mu_2 + 3\mu_4 + \mu_7 - \mu_6 = \mu_8$  y  $3\mu_2 - 4\mu_3 + 2\mu_5 + 2\mu_6 = 0$ . Debe especificar todas las condiciones y elementos para probar esta hipótesis. Anexe los códigos en R usados.

Primero se procede a plantear las hipótesis para los contrastes, como:

$$H_0 : C\boldsymbol{\mu} = \boldsymbol{\gamma} \quad VS \quad C\boldsymbol{\mu} \neq \boldsymbol{\gamma}$$

Escritas de otra manera como:

$$H_0 : 2\mu_1 - \mu_2 + 3\mu_4 + \mu_7 - \mu_6 = \mu_8 \quad y \quad 3\mu_2 - 4\mu_3 + 2\mu_5 + 2\mu_6 = 0$$

Donde:

$$C = \begin{pmatrix} 2 & -1 & 0 & 3 & 0 & -1 & 1 & -1 \\ 0 & 3 & -4 & 0 & 2 & 2 & 0 & 0 \end{pmatrix}$$

Estadístico de prueba:

$$T_0^2 = n(C\bar{\mathbf{X}} - \boldsymbol{\gamma})' (CSC')^{-1} (C\bar{\mathbf{X}} - \boldsymbol{\gamma})$$

Bajo  $H_0$  cierto. Se rechaza si  $\frac{n-k}{(n-1)k}T_0^2 > f_\alpha(k, n-k)$

Luego tenemos un  $\frac{199-2}{2(199-1)}T_0^2 = 50.57493 > f_{0.05}(2, 197) = 3.041753$  y con una significancia de  $\alpha = 0.05$  lo que nos permite rechazar  $H_0$  concluyendo que  $2\mu_1 - \mu_2 + 3\mu_4 + \mu_7 - \mu_6 \neq \mu_8$  y  $3\mu_2 - 4\mu_3 + 2\mu_5 + 2\mu_6 \neq 0$