



UNIVERSIDAD NACIONAL DE COLOMBIA

MAESTRÍA EN CIENCIAS ESTADÍSTICA

DEPARTAMENTO DE ESTADÍSTICA

FACULTAD DE CIENCIAS

— ANÁLISIS MULTIVARIADO DE DATOS —

Integrantes:

Luis David Hernández Pérez C.C. 1193549963
Daniel Felipe Villa Rengifo C.C. 1005087556

Medellín, Colombia
Semestre 2024-02

Medellín, Febrero 15 de 2025

Tabla de contenidos

Punto 01	2
Solución Punto-01	2
Literal a)	2
Literal b)	4
Literal c)	5
Punto 02	6
Solución Punto 02	7
Literal a)	7
Literal b)	8
Literal c)	9
Punto 03	11
Solución Punto 03	11
Literal a)	12
Literal b)	13
Literal c)	15
Punto 04	16
Solución Punto 04	16
Literal a)	17
Literal b)	19
Punto 05	20
Solución Punto 05	20
Conclusiones	21

Punto 01

Para el conjunto de datos asignado, realice los siguientes procesos de prueba de hipótesis (PH):

a)

$$\begin{cases} H_0 : \mu_{6 \times 1} = \mathbf{0}_{6 \times 1} \\ H_a : \mu_{6 \times 1} \neq \mathbf{0}_{6 \times 1} \end{cases}$$

b)

$$\begin{cases} H_0 : \Sigma_{6 \times 6} = \mathbf{I}_{6 \times 6} \\ H_a : \Sigma_{6 \times 6} \neq \mathbf{I}_{6 \times 6} \end{cases}$$

c)

$$\begin{cases} H_0 : \begin{bmatrix} \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix}_{3 \times 1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{3 \times 1} \\ H_a : \begin{bmatrix} \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix}_{3 \times 1} \neq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}_{3 \times 1} \end{cases}$$

Solución Punto-01

Para este punto se utilizaron los datos del equipo 07.

Literal a)

Primeramente verificaremos si la muestra de datos se distribuyen normal 6-variado por medio de la **Prueba de Mardia**.

Tabla 1: Prueba de Mardia

Test	Statistic	p value	Result
Mardia Skewness	56.3077170734103	0.46333230310962	YES
Mardia Kurtosis	-0.676469395950928	0.498742670347313	YES
MVN	NA	NA	YES

De la Tabla 1 podemos concluir que los datos si se ditribuyen normal 6-variado.

En esta literal nos piden probar la siguiente hipotesis

$$\begin{cases} H_0 : \underline{\mu} = \underline{\mu}_0 \\ H_a : \underline{\mu} \neq \underline{\mu}_0 \end{cases}$$

Dado que la muestra de datos se distribuye $\mathcal{N}_6(\underline{\mu}, \Sigma)$, con $\underline{\mu}$ -desconocida y Σ -desconocida.

El estadístico de prueba a utilizar es el estadístico de prueba T^2 de Hotelling y esta dado por:

$$T^2 = n(\bar{x} - \mu_0)' S^{-1}(\bar{x} - \mu_0)$$

Al nivel de significancia del $\alpha\%$, rechazamos $H_0 : \mu = \mu_0$, en favor de: $H_a : \mu \neq \mu_0$, si el valor de:

$$T_0^2 = n(\bar{x} - \mu_0)' \mathbf{S}^{-1}(\bar{x} - \mu_0) > kF = \frac{(n-1)p}{n-p} F_{\alpha; p, n-p},$$

o equivalentemente, rechazamos H_0 si:

$$F_0 = \frac{(n-p)}{(n-1)p} T^2 = \frac{1}{k} T_0^2 > F_{\alpha; p, n-p},$$

en caso contrario no rechazamos H_0 .

El vector μ_0 es: 0 0 0 0 0 0

T2	K	F0	df1	df2	F_Tabla	Valor_p
7.84267	7.42857	1.05574	6	21	2.57271	0.419112

Como $p\text{-valor} = 0.419112 > 0.05 = \alpha$, entonces **No se rechaza** H_0 , es decir el vector de medias para la muestra es el vector nulo.

Literal b)

Para este literal sabemos que la muestra de datos se distribuye $\mathcal{N}_6(\underline{\mu}, \Sigma)$, con $\underline{\mu}$ -desconocida y Σ -desconocida.

La hipótesis a probar es:

$$\begin{cases} H_0 : \Sigma_{6 \times 6} = I_6 \\ H_a : \Sigma_{6 \times 6} \neq I_6 \end{cases}$$

En este caso, el estadístico de prueba de Razón de Verosimilitud está dado por:

$$\lambda = \left[\left(\frac{n-1}{n} \right)^p \frac{|S|}{|\Sigma_0|} \right]^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} [(n-1)\text{tr}(S\Sigma_0^{-1}) - np] \right\}$$

El cual se puede reescribir como:

$$\lambda = \frac{|S|^{\frac{v}{2}}}{|\Sigma_0|^{\frac{v}{2}}} \exp \left\{ -\frac{1}{2} [\text{tr}(S\Sigma_0^{-1}) - vp] \right\}, \quad \text{con: } v = n - 1 = n$$

y haciendo: $\lambda^* = -2 \log \lambda$ se tiene que el Estadístico de prueba es:

$$\lambda^* = v [\log |\Sigma_0| - \log |S| + \text{tr}(S\Sigma_0^{-1}) - p] \stackrel{\text{Bajo } H_0}{\sim} \chi_k^2 \text{ para: } (n-1) - \text{Grande}$$

$$\text{con: } k = \frac{p(p+1)}{2} - p = \frac{p(p+1)}{2}$$

Al nivel de significancia de α rechazamos $H_0 : \Sigma = \Sigma_0$ en favor de: $H_a : \Sigma \neq \Sigma_0$ si se cumple lo siguiente:

$$\lambda_0^* > \chi_{\alpha; k}^2$$

donde $\chi_{\alpha; k}^2$ denota el percentil superior α de la distribución chi-cuadrado con k -grados de libertad.

O equivalentemente, rechazamos H_0 si el valor-p es menor que α , es decir si: $\text{valor}_p < \alpha$.

Lamda1_est	c	df	Chi_Tabla	Valor_P
888.19	0.9185	21	32.671	0

Como $p\text{-valor} \approx 0 < 0.05 = \alpha$ entonces, **se rechaza** H_0 , es decir la matriz de varianzas y covarianzas de la muestra no es igual a la matriz identidad.

Literal c)

Para este literal sabemos que la muestra de datos se distribuye $\mathcal{N}_6(\underline{\mu}, \Sigma)$, con $\underline{\mu}$ -desconocida y Σ -desconocida.

La hipótesis a probar es:

$$\left\{ \begin{array}{l} H_0 : \begin{bmatrix} \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ H_a : \begin{bmatrix} \mu_1 \\ \mu_3 \\ \mu_5 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \end{array} \right.$$

El estadístico de prueba a utilizar es el estadístico de prueba T^2 de Hotelling y está dado por:

$$T^2 = n(\bar{x} - \mu_0)' S^{-1}(\bar{x} - \mu_0)$$

Al nivel de significancia del $\alpha\%$, rechazamos $H_0 : \mu = \mu_0$, en favor de: $H_a : \mu \neq \mu_0$, si el valor de:

$$T_0^2 = n(\bar{x} - \mu_0)^t \mathbf{S}^{-1}(\bar{x} - \mu_0) > kF = \frac{(n-1)p}{n-p} F_{\alpha; p, n-p},$$

o equivalentemente, rechazamos H_0 si:

$$F_0 = \frac{(n-p)}{(n-1)p} T^2 = \frac{1}{k} T_0^2 > F_{\alpha; p, n-p},$$

en caso contrario no rechazamos H_0 .

El vector μ_0 es: 1 1 1

T2	K	F0	df1	df2	F_Tabla	Valor_p
5.82151	3.25	1.79123	3	24	3.00879	0.175741

Como el p-valor = 0.175741 > 0.05 = α , entonces **No se rechaza** H_0 , es decir que la media para las variables 1, 3 y 5 es igual al vector $(1, 1, 1)'$.

Punto 02

Para el conjunto de datos asignado, realice los siguientes procesos de PH:

a)

$$\begin{cases} H_0 : \mu_{6 \times 1} = \mathbf{1}_{6 \times 1} \\ H_a : \mu_{6 \times 1} \neq \mathbf{1}_{6 \times 1} \end{cases}$$

b)

$$\begin{cases} H_0 : \Sigma_{6 \times 6} = 2\mathbf{I}_{6 \times 6} \\ H_a : \Sigma_{6 \times 6} \neq 2\mathbf{I}_{6 \times 6} \end{cases}$$

c)

$$\begin{cases} H_0 : \begin{bmatrix} \mu_2 \\ \mu_4 \\ \mu_6 \end{bmatrix}_{3 \times 1} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}_{3 \times 1} \\ H_a : \begin{bmatrix} \mu_2 \\ \mu_4 \\ \mu_6 \end{bmatrix}_{3 \times 1} \neq \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}_{3 \times 1} \end{cases}$$

Solución Punto 02

Para este punto se utilizaron los datos del equipo 07.

Primeramente verificaremos si la muestra de datos se distribuyen normal seis-variado por medio de la **Prueba de Royston**.

Test	H	p value	MVN
Royston	7.842354	0.2543664	YES

Literal a)

Teniendo en cuenta que el resultado de la **Prueba de Royston**, por tanto la muestra de datos se distribuye $\mathcal{N}_6(\underline{\mu}, \Sigma)$, con $\underline{\mu}$ -desconocida y Σ -desconocida.

La hipótesis a probar es:

$$\begin{cases} H_0 : \underline{\mu} = \underline{\mu}_0 \\ H_a : \underline{\mu} \neq \underline{\mu}_0 \end{cases}, \quad \text{con } \underline{\mu}_0 = \mathbf{1}_{6 \times 1}$$

El estadístico de prueba a utilizar es el estadístico de prueba T^2 de Hotelling y está dado por:

$$T^2 = n(\bar{x} - \mu_0)' S^{-1}(\bar{x} - \mu_0)$$

Al nivel de significancia del $\alpha\%$, rechazamos $H_0 : \mu = \mu_0$, en favor de: $H_a : \mu \neq \mu_0$, si el valor de:

$$T_0^2 = n(\bar{x} - \mu_0)' \mathbf{S}^{-1}(\bar{x} - \mu_0) > kF = \frac{(n-1)p}{n-p} F_{\alpha; p, n-p},$$

o equivalentemente, rechazamos H_0 si:

$$F_0 = \frac{(n-p)}{(n-1)p} T^2 = \frac{1}{k} T_0^2 > F_{\alpha; p, n-p},$$

en caso contrario no rechazamos H_0 .

El vector μ_0 es: 1 1 1 1 1 1

T2	K	F0	df1	df2	F_Tabla	Valor_p
6.42079	6.69767	0.95866	6	43	2.3185	0.464322

Como $p\text{-valor} = 0.464322 > 0.05 = \alpha$, entonces No se rechaza H_0 , es decir el vector de medias asociado a la muestra de los datos es igual al vector de $(1, 1, 1, 1, 1, 1)'$.

Literal b)

Para este literal sabemos que la muestra de datos se distribuye $\mathcal{N}_6(\underline{\mu}, \Sigma)$, con $\underline{\mu}$ -desconocida y Σ -desconocida.

La hipótesis a probar es:

$$\begin{cases} H_0 : \Sigma_{6 \times 6} = 2I_6 \\ H_a : \Sigma_{6 \times 6} \neq 2I_6 \end{cases}$$

En este caso, es estadístico de prueba de Razón de Verosimilitud esta dado por:

$$\lambda = \left[\left(\frac{n-1}{n} \right)^p \frac{|S|}{|\Sigma_0|} \right]^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} [(n-1)\text{tr}(S\Sigma_0^{-1}) - np] \right\}$$

El cual se puede reescribir como:

$$\lambda = \frac{|S|^{\frac{v}{2}}}{|\Sigma_0|^{\frac{v}{2}}} \exp \left\{ -\frac{1}{2} [\text{tr}(S\Sigma_0^{-1}) - vp] \right\}, \quad \text{con: } v = n - 1 = n$$

y haciendo: $\lambda^* = -2 \log \lambda$ se tiene que el Estadístico de prueba es:

$$\lambda^* = v [\log |\Sigma_0| - \log |S| + \text{tr}(S\Sigma_0^{-1}) - p] \stackrel{\text{Bajo } H_0}{\sim} \chi_k^2 \text{ para: } (n-1) - \text{Grande}$$

$$\text{con: } k = \frac{p(p+1)}{2} - p = \frac{p(p+1)}{2}$$

Al nivel de significancia de α rechazamos $H_0 : \Sigma = \Sigma_0$ en favor de: $H_a : \Sigma \neq \Sigma_0$ si se cumple lo siguiente:

$$\lambda_0^* > \chi_{\alpha;k}^2$$

donde $\chi_{\alpha;k}^2$ denota el percentil superior α de la distribución chi-cuadrado con k-grados de libertad.

O equivalentemente, rechazamos H_0 si el valor-p es menor que α , es decir si: $\text{valor}_p < \alpha$.

Lamda1_est	c	df	Chi_Tabla	Valor_P
392.87	0.95585	21	32.671	0

Como p-valor $\approx 0 < 0.05 = \alpha$ entonces, **se rechaza** H_0 , es decir la matriz de varianzas y covarianzas de la muestra de datos es diferente a 2 veces la matriz identidad.

Literal c)

La hipótesis a probar es:

$$\left\{ \begin{array}{l} H_0 : \begin{bmatrix} \mu_2 \\ \mu_4 \\ \mu_6 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \\ H_a : \begin{bmatrix} \mu_2 \\ \mu_4 \\ \mu_6 \end{bmatrix} \neq \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} \end{array} \right.$$

En este caso, es estadístico de prueba es:

$$T^2 = (\bar{x} - \mu_0)^t \left(\frac{1}{n} \mathbf{S} \right)^{-1} (\bar{x} - \mu_0) = n(\bar{x} - \mu_0)^t \mathbf{S}^{-1} (\bar{x} - \mu_0)$$

Se utiliza el siguiente resultado:

$$T^2 \stackrel{\text{Bajo } H_0}{\sim} \frac{(n-1)p}{n-p} F_{p,n-p} = kF, \quad \text{con: } k = \frac{(n-1)p}{n-p}$$

o equivalentemente:

$$F = \frac{1}{k} T^2 = \frac{(n-p)}{(n-1)p} T^2 \stackrel{\text{Bajo } H_0}{\sim} F_{p,n-p},$$

donde $F_{p,n-p}$ -denota una v.a con distribución F con p y (n-p) grados de libertad respectivamente.

Al nivel de significancia de α rechazamos $H_0 : \mu = \mu_0$ en favor de: $H_a : \mu \neq \mu_0$ si se cumple lo siguiente:

$$T_0^2 = n(\bar{x} - \mu_0)^t \mathbf{S}^{-1}(\bar{x} - \mu_0) > kF = \frac{(n-1)p}{n-p} F_{\alpha;p,n-p},$$

donde $F_{\alpha;p,n-p}$ -denota el percentil superior α distribución F con p y (n-p) grados de libertad respectivamente.

O equivalentemente, rechazamos H_0 si:

$$F_0 = \frac{1}{k} T_0^2 = \frac{(n-p)}{(n-1)p} T_0^2 > F_{\text{tabla}} = F_{\alpha;p,n-p}$$

O equivalentemente, rechazamos H_0 si el valor-p es menor que: α es decir si: $\text{valor}_p < \alpha$.

El vector mu0 es: 2 2 2

T2	K	F0	df1	df2	F_Tabla	Valor_p
14.7165	3.13043	4.70109	3	46	2.80684	0.00604357

Como p-valor = 0.00604357 < 0.05 = α , entonces, se rechaza H_0 , es decir que la media para las variables 2, 4 y 6 no es igual al vector (2, 2, 2)'.

Punto 03

Para el conjunto de datos asignado, realice los siguientes procesos de PH:

a)

$$\begin{cases} H_0 : \Sigma_1 = \Sigma_2 \\ H_a : \Sigma_1 \neq \Sigma_2 \end{cases}$$

b) Asumiendo independencia entre las muestras:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 \neq \mu_2 \end{cases}$$

c) Asumiendo no-independencia entre las muestras:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 \neq \mu_2 \end{cases}$$

Solución Punto 03

Para este punto se utilizaron los datos del equipo 07.

Primeramente verifiquemos si las dos muestras provienen de una distribución normal 3-variada por medio de la **Prueba de Royston**.

Tabla 9: Prueba de Royston para la Muestra 1

Test	H	p value	MVN
Royston	1.170216	0.7650185	YES

De la Tabla 9 podemos concluir que los datos asociados a la muestra 1 se distribuyen normal 3-variado.

Tabla 10: Prueba de Royston para la Muestra 2

Test	H	p value	MVN
Royston	1.746243	0.6335564	YES

De la Tabla 10 podemos concluir que los datos asociados a la muestra 2 se distribuyen normal 3-variado.

Literal a)

Las hipótesis a probar son

$$\begin{cases} H_0 : \Sigma_1 = \Sigma_2 \\ H_a : \Sigma_1 \neq \Sigma_2 \end{cases}$$

En este caso, el estadístico de prueba de Razón de Verosimilitud está dado por:

$$\lambda = \prod_{i=1}^g \left(\frac{|S_i|}{|S_p|} \right)^{\frac{n_i-1}{2}}$$

donde S_i es la matriz de Var-Cov muestral asociada a la m.a de la i -ésima población, para $i = 1, 2 = g$, y

$$\mathbf{S}_p = \hat{\Sigma} = \frac{1}{\sum_{i=1}^{g=2} (n_i - 1)} \left[\sum_{i=1}^{g=2} (n_i - 1) \mathbf{S}_i \right] = \frac{(n-1)\mathbf{S}_1 + (m-1)\mathbf{S}_2}{n+m-2}$$

es la matriz de var-cov ponderada.

La **Estadística M de Box** se define como $M = -2 \log \lambda$ y se puede escribir como sigue:

$$M = \left[\sum_{i=1}^{g=2} (n_i - 1) \right] \log |\mathbf{S}_p| - \sum_{i=1}^{g=2} (n_i - 1) \log |\mathbf{S}_i| = v \log |\mathbf{S}_p| - \sum_{i=1}^{g=2} v_i \log |\mathbf{S}_i|$$

Se utiliza la siguiente aproximación:

$$C = (1 - u)M \xrightarrow{d} \chi_k^2$$

con:

$$u = \left[\sum_{i=1}^{g=2} \left(\frac{1}{n_i - 1} \right) - \frac{1}{\sum_{i=1}^{g=2} (n_i - 1)} \right] \left(\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right)$$

$$= \left[\sum_{i=1}^{g=2} \frac{1}{v_i} - \frac{1}{v} \right] \left(\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right)$$

$$\text{con: } k = \left[gp + g\frac{1}{2}p(p+1) \right] - \left[gp + \frac{1}{2}p(p+1) \right] = \frac{p(p+1)(g-1)}{2} - \text{grados de libertad, } g = 2$$

Al nivel de significancia de α rechazamos $H_0 : \Sigma_1 = \Sigma_2$ en favor de: $H_a : \Sigma_1 \neq \Sigma_2$ si se cumple lo siguiente:

$$C_0 > \chi_{\alpha; k}^2$$

donde $\chi_{\alpha; k}$ denota el percentil superior α de la distribución chi-cuadrado con k-grados de libertad.

O equivalentemente, rechazamos H_0 si el valor-p es menor que: α es decir si: $\text{valor}_p < \alpha$.

M	U	C	df	Chi_Tabla	Valor_p
4.79866	0.065	4.48675	6	12.5916	0.611107

Como $p\text{-valor} = 0.611107 > 0.05 = \alpha$ entonces, No se rechaza H_0 , es decir que la matriz de varianzas y covarianzas de la muestra 1 y 2 asociada a los datos son iguales.

Literal b)

Teniendo en cuenta que las dos muestras se distribuyen normal 3-variada con vector de medias $\underline{\mu}_1$ y $\underline{\mu}_2$ desconocidos y $\Sigma_1 = \Sigma_2 = \Sigma$ -Desconocida. Asumiendo que las dos muestras son independientes las hipótesis contrastar son:

$$\begin{cases} H_0 : \underline{\mu}_1 - \underline{\mu}_2 = \underline{\delta}_0 \\ H_a : \underline{\mu}_1 - \underline{\mu}_2 \neq \underline{\delta}_0 \end{cases}, \quad \text{con } \underline{\delta}_0 = \underline{\mathbf{0}}$$

En este caso, se usa como estimador de Σ a la matriz de varianzas-covarianzas ponderada dada por:

$$S_p = \hat{\Sigma} = \frac{(n-1)S_1 + (m-1)S_2}{n+m-2}$$

En este caso el estadístico de prueba es:

$$T_0^2 = \frac{nm}{n+m} (\bar{x} - \bar{y} - \delta_0)^t S_p^{-1} (\bar{x} - \bar{y} - \delta_0) \stackrel{\text{Bajo } H_0}{\sim} \frac{(n+m-2)p}{n+m-p-1} F_{p;n+m-p-1} = kF$$

$$\text{con: } k = \frac{(n+m-2)p}{n+m-p-1}$$

Rechazamos H_0 si:

$$T_0^2 > kF = \frac{(n+m-2)p}{n+m-p-1} F_{\alpha;p,n+m-p-1}$$

O equivalentemente, rechazamos H_0 si:

$$F_0 = \frac{n+m-p-1}{(n+m-2)p} T_0^2 = \frac{1}{k} T_0^2 > F_{\text{tabla}} = F_{\alpha;p,n+m-p-1}$$

O equivalentemente, rechazamos H_0 si el valor-p es menor que: es decir si: $\text{valor}_p < \alpha$.

El vector μ_0 es: 0 0 0

T2	k	F0	df1	df2	F_Tabla	Valor_p
0.925175	3.125	0.296056	3	48	2.79806	0.828051

Como $p\text{-valor} = 0.828051 > 0.05 = \alpha$ entonces, **No se rechaza** H_0 , es decir el vector de medias para la muestra 1 y 2 son iguales.

Literal c)

Teniendo en cuenta que las dos muestras se distribuyen normal 3-variada con vector de medias $\underline{\mu_1}$ y $\underline{\mu_2}$ desconocidos y Σ -Desconocida. Asumiendo que las dos muestras **no** son independientes las hipótesis contrastar son:

$$\begin{cases} H_0 : \underline{\mu_1} - \underline{\mu_2} = \underline{\mathbf{0}} \\ H_a : \underline{\mu_1} - \underline{\mu_2} \neq \underline{\mathbf{0}} \end{cases}$$

y el estadístico de prueba es:

$$T^2 = n\bar{D}^t S_D^{-1} \bar{D} \approx \frac{(n-1)p}{n-p} F_{p, n-p}, \quad \text{donde:}$$

$$\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i \quad \text{y} \quad S_D = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})(D_i - \bar{D})^t,$$

son el vector de medias y la matriz de Var-Cov de las diferencias muestrales.

$$\text{Se rechaza } H_0 \text{ si: } T_0^2 > \frac{(n-1)p}{n-p} F_{\alpha; p, n-p}.$$

O equivalentemente, rechazamos H_0 si el valor-p es menor que: α , es decir si: $\text{valor}_p < \alpha$.

T2	K	F0	df1	df2	F_Tabla	Valor_p
1.08748	3.26087	0.333494	3	23	3.028	0.801211

Como el p-valor = 0.801211 > 0.05 = α entonces, **No se rechaza** H_0 , es decir el vector de medias para la muestra 1 y 2 son iguales.

Punto 04

Para el conjunto de datos asignado, realice los siguientes procesos de PH:

a)

$$\begin{cases} H_0 : \Sigma_1 = \Sigma_2 \\ H_a : \Sigma_1 \neq \Sigma_2 \end{cases}$$

b) Asumiendo independencia entre las muestras:

$$\begin{cases} H_0 : \mu_1 = \mu_2 \\ H_a : \mu_1 \neq \mu_2 \end{cases}$$

Solución Punto 04

Para este punto se utilizaron los datos del equipo 07.

Primeramente verifiquemos si las dos muestras provienen de una distribución normal 3-variada por medio de la **Prueba de Mardia**.

Tabla 14: Prueba de Mardia para la Muestra 1

Test	Statistic	p value	Result
Mardia Skewness	16.2134834125481	0.0936821029634656	YES
Mardia Kurtosis	1.42555894737023	0.153995693344217	YES
MVN	NA	NA	YES

De la Tabla 14 podemos concluir que los datos asociados a la muestra 1 se distribuyen normal 3-variado

Tabla 15: Prueba de Mardia para la Muestra 2

Test	Statistic	p value	Result
Mardia Skewness	6.94205472340896	0.73090378798055	YES
Mardia Kurtosis	0.724287679038595	0.46888913955506	YES
MVN	NA	NA	YES

De la Tabla 15 podemos concluir que los datos asociados a la muestra 2 se distribuyen normal 3-variado

Literal a)

Teniendo en cuenta que el resultado de la **Prueba de Mardia** para las dos muestras podemos decir que ambas muestras se distribuye $\mathcal{N}_3(\underline{\mu}_i, \Sigma_i)$, para $i=1,2$, con $\underline{\mu}_1$ y $\underline{\mu}_2$ desconocidos y Σ_1 y Σ_2 desconocidas.

La hipótesis a probar es

$$\begin{cases} H_0 : \Sigma_1 = \Sigma_2 \\ H_a : \Sigma_1 \neq \Sigma_2 \end{cases}$$

En este caso, el estadístico de prueba de Razón de Verosimilitud está dado por:

$$\lambda = \prod_{i=1}^g \left(\frac{|S_i|}{|S_p|} \right)^{\frac{n_i-1}{2}}$$

donde S_i es la matriz de Var-Cov muestral asociada a la m.a de la i -ésima población, para $i = 1, 2 = g$, y

$$\mathbf{S}_p = \hat{\Sigma} = \frac{1}{\sum_{i=1}^{g=2} (n_i - 1)} \left[\sum_{i=1}^{g=2} (n_i - 1) \mathbf{S}_i \right] = \frac{(n-1)\mathbf{S}_1 + (m-1)\mathbf{S}_2}{n+m-2}$$

es la matriz de var-cov ponderada.

La **Estadística M de Box** se define como $M = -2 \log \lambda$ y se puede escribir como sigue:

$$M = \left[\sum_{i=1}^{g=2} (n_i - 1) \right] \log |\mathbf{S}_p| - \sum_{i=1}^{g=2} (n_i - 1) \log |\mathbf{S}_i| = v \log |\mathbf{S}_p| - \sum_{i=1}^{g=2} v_i \log |\mathbf{S}_i|$$

Se utiliza la siguiente aproximación:

$$C = (1 - u)M \xrightarrow{d} \chi_k^2$$

con:

$$\begin{aligned}
u &= \left[\sum_{i=1}^{g=2} \left(\frac{1}{n_i - 1} \right) - \frac{1}{\sum_{i=1}^{g=2} (n_i - 1)} \right] \left(\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right) \\
&= \left[\sum_{i=1}^{g=2} \frac{1}{v_i} - \frac{1}{v} \right] \left(\frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right)
\end{aligned}$$

$$\text{con: } k = \left[gp + g\frac{1}{2}p(p+1) \right] - \left[gp + \frac{1}{2}p(p+1) \right] = \frac{p(p+1)(g-1)}{2} - \text{grados de libertad, } g = 2$$

Al nivel de significancia de α rechazamos $H_0 : \Sigma_1 = \Sigma_2$ en favor de: $H_a : \Sigma_1 \neq \Sigma_2$ si se cumple lo siguiente:

$$C_0 > \chi_{\alpha; k}^2$$

donde $\chi_{\alpha; k}$ denota el percentil superior α de la distribución chi-cuadrado con k-grados de libertad.

O equivalentemente, rechazamos H_0 si el valor-p es menor que: α es decir si: $\text{valor}_p < \alpha$.

M	U	C	df	Chi_Tabla	Valor_p
13.2937	0.0295455	12.9009	6	12.5916	0.0446368

Como $p\text{-valor} = 0.0446368 < 0.05 = \alpha$ entonces, **se rechaza** H_0 , es decir que la matriz de varianzas y covarianza de la muestra 1 y 2 asociadas a los datos son diferentes.

Literal b)

Asumiendo que las dos muestras son independientes y de la hipótesis anterior $\Sigma_1 \neq \Sigma_2$, las hipótesis contrastar son:

$$\begin{cases} H_0 : \underline{\mu_1} - \underline{\mu_2} = \underline{\delta_0} \\ H_a : \underline{\mu_1} - \underline{\mu_2} \neq \underline{\delta_0} \end{cases}, \quad \text{con} \quad \underline{\delta_0} = \mathbf{0}$$

En este caso el estadístico de prueba es:

$$T^2 = (\bar{x} - \bar{y} - \delta_0)^t \left[\frac{S_1}{n} + \frac{S_2}{m} \right]^{-1} (\bar{x} - \bar{y} - \delta_0) \stackrel{\text{Bajo } H_0}{\sim} \frac{vp}{v-p+1} F_{p;v-p+1} = kF$$

$$\text{con: } k = \frac{vp}{v-p+1} \quad \text{y} \quad v = \frac{\text{tr}(S_e) + [\text{tr}(S_e)]^2}{\sum_{i=1}^2 \frac{1}{n_i-1} \{ \text{tr}(V_i) + [\text{tr}(V_i)]^2 \}}$$

$$V_i = \frac{S_i}{n_i} \quad \text{y} \quad S_e = V_1 + V_2 = \frac{S_1}{n} + \frac{S_2}{m}.$$

$$\text{Rechazamos } H_0 \quad \text{si:} \quad T_0^2 > kF \quad \text{ó} \quad F_0 = \frac{1}{k} T_0^2 > F_{\text{tabla}}$$

El vector mu0 es: 0 0 0

T_2	v	k	F0	df1	df2	F_Tabla	Valor_p
3.57569	107	3.05714	1.16962	3	105	2.69113	0.324959

Como p-valor = 0.324959 > 0.05 = α entonces, **No se rechaza** H_0 , es decir los vectores de medias asociados a la muestra 1 y 2 de los datos son iguales.

Punto 05

- a) Usando los datos asignados para el Punto 01, construya elipses de confianza del 90% y del 95% para el vector de medias poblacional:

$$\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}_{2 \times 1}$$

a partir de los datos muestrales.

- b) A partir de las regiones de confianza del inciso (a), determine la decisión sobre la siguiente PH:

$$\begin{cases} H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 2.5 \\ 2.2 \end{bmatrix} \\ H_a : \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \neq \begin{bmatrix} 2.5 \\ 2.2 \end{bmatrix} \end{cases}$$

- c) A partir de las regiones de confianza del inciso (a), determine la decisión sobre la siguiente PH:

$$\begin{cases} H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 1.3 \\ 1.2 \end{bmatrix} \\ H_a : \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \neq \begin{bmatrix} 1.3 \\ 1.2 \end{bmatrix} \end{cases}$$

Solución Punto 05

Dado que los puntos purpura $\mu_0 = (1.3, 1.2)'$ y naranjado $\mu_0 = (1.3, 1.2)'$ se encuentran dentro de las regiones de confianza por tanto se apoyan estas hipótesis nulas

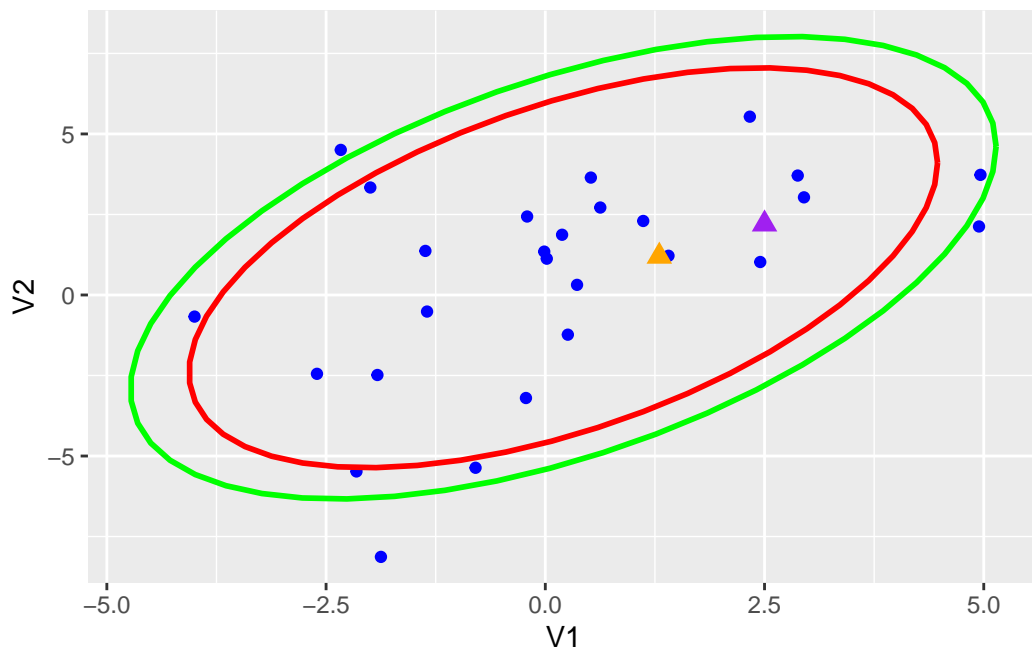


Figura 1: Elipses de confianza del 90% y 95% para el vector de medias con puntos H_{01} y H_{02}

Conclusiones

Los resultados obtenidos proporcionan información relevante sobre la estructura de medias y covarianzas del conjunto de datos analizado. La combinación de pruebas multivariadas y herramientas gráficas permite una interpretación más completa y detallada de la estructura subyacente de los datos.