Primer Parcial de Análisis Multivariado I

PRIMER CUATRIMESTRE DE 2017

- Justifique todas sus respuestas -

1	2	3	4	Calificación
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Nombre y Apellido: [redacted]

Carrera: [redacted]

Cantidad Total de Hojas: [redacted]

Firma: [redacted]

Criterio de ***aprobación:*** *Para aprobar este examen se debe totalizar al menos 50 puntos.*

Por favor, hacer cada ejercicio en hoja aparte, enumerar todas las hojas, colocar el nombre en ellas e indicar en la última hoja antes de la firma el número total de hojas.

---

1. Consideremos dos muestras independientes de vectores de dimensión d: $\mathbf{x}_1, \ldots, \mathbf{x}_m$ con distribución $N_d(\mu, \Sigma)$ e $\mathbf{y}_1, \ldots, \mathbf{y}_n$ con distribución $N_d(\nu, \Sigma)$. Se puede pensar que las coordenadas de cada vector representan la medición de la misma variable en d períodos de tiempo equiespaciados.

   a) Proponer un test de nivel exacto para la hipótesis
   $$H_0 : \nu_i = \mu_i + a_0 + b_0\, i \qquad 1 \leq i \leq d$$
   con a_0 y b_0 constantes conocidas.

   Describa claramente

   - *cuáles son los supuestos que está haciendo,*
   - *cuáles son los vectores en estudio y su distribución,*
   - *que estadístico está utilizando,*
   - *la distribución del estadístico bajo H_0 (ya sea la exacta o la asintótica, según corresponda) y la región de rechazo del test.*

   b) Se quiere ahora testear la hipótesis
   $$H_0 : \nu_i = \mu_i + a + b\, i \qquad 1 \leq i \leq d$$
   con a y b constantes *desconocidas*, es decir, se quiere ver que las diferencias entre las medias de las dos poblaciones crecen linealmente en el tiempo.

Para ello, se define la siguiente transformación lineal $\mathbf{A} : \mathbb{R}^d \to \mathbb{R}^{d-2}$. Dado un vector $\mathbf{x}$ de dimensión d, indicaremos por $\mathbf{u} = \mathbf{A}\,\mathbf{x}$ al vector $\mathbf{u} = (u_1, \dots, u_{d-2})^{\mathrm{T}}$ definido por $u_i = x_{i+2} - 2x_{i+1} + x_i$. De la misma manera, dado un vector $\mathbf{y}$ se define $\mathbf{v} = (v_1, \dots, v_{d-2})^{\mathrm{T}} = \mathbf{A}\,\mathbf{y}$ con $v_i = y_{i+2} - 2y_{i+1} + y_i$.

i. Si $\mathbf{x} \sim N_d(\mu, \mathbf{\Sigma})$ e $\mathbf{y} \sim N_d(\nu, \mathbf{\Sigma})$, ¿qué distribución tiene $\mathbf{u} = \mathbf{A}\,\mathbf{x}$ y $\mathbf{v} = \mathbf{A}\,\mathbf{y}$?

ii. Calcule $\mathbb{E}(\mathbf{v} - \mathbf{u})$ bajo H_0.

iii. Deduzca que H_0 se puede testear con un test de nivel exacto mediante los datos transformados $\mathbf{A}\,\mathbf{x}_1, \dots, \mathbf{A}\,\mathbf{x}_m$ y $\mathbf{A}\,\mathbf{y}_1, \dots, \mathbf{A}\,\mathbf{y}_n$.

Describa claramente

- *cuáles son los supuestos que está haciendo,*
- *cuáles son los vectores aleatorios en estudio y su distribución,*
- *que estadístico está utilizando,*
- *la distribución del estadístico bajo H_0 (ya sea la exacta o la asintótica, según corresponda) y la región de rechazo del test.*

2. El siguiente ejemplo fue tomado de Abruzzi (1950). El propósito del estudio era investigar cuánto tiempo les toma a varios operarios realizar determinadas tareas. La tarea total se dividió en seis partes. Sea $\mathbf{x}_i$ el vector de mediciones obtenidas sobre el operario i, siendo la j−ésima coordenada de $\mathbf{x}_i$ el tiempo que le lleva realizar la j−ésima parte de la tarea. Se eligieron 76 operarios y se obtuvieron el vector de medias $\bar{\mathbf{x}}$, la matriz de covarianzas $\mathbf{S}$ y la matriz de correlación $\mathbf{R}$ que se dan a continuación.

$$\bar{\mathbf{x}} = (9.47;\ 25.56;\ 13.25;\ 31.44;\ 27.29;\ 8.80)^{\mathrm{T}}$$

$$\mathbf{S} = \begin{pmatrix} 2.57 & 0.85 & 1.56 & 1.79 & 1.33 & 0.42 \\ 0.85 & 37.00 & 3.34 & 13.47 & 7.59 & 0.52 \\ 1.56 & 3.34 & 8.44 & 5.77 & 2.00 & 0.50 \\ 1.79 & 13.47 & 5.77 & 34.01 & 10.50 & 1.77 \\ 1.33 & 7.59 & 2.00 & 10.50 & 23.01 & 3.43 \\ 0.42 & 0.52 & 0.50 & 1.77 & 3.43 & 4.59 \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} 1 & 0.088 & 0.334 & 0.191 & 0.173 & 0.123 \\ 0.088 & 1 & 0.286 & 0.384 & 0.262 & 0.040 \\ 0.334 & 0.286 & 1 & 0.343 & 0.144 & 0.080 \\ 0.191 & 0.384 & 0.343 & 1 & 0.375 & 0.142 \\ 0.173 & 0.262 & 0.144 & 0.375 & 1 & 0.334 \\ 0.123 & 0.040 & 0.080 & 0.142 & 0.334 & 1 \end{pmatrix}$$

Los investigadores están interesados en testear la hipótesis de que sus variables son independientes.

a) ¿Qué estadístico utilizaría usted para testear esa hipótesis? Indique como construiría un test a partir de ese estadístico aclarando si el test es exacto o asintótico y qué supuestos involucra.

Especifique claramente el vector aleatorio en estudio, el modelo supuesto, la hipótesis nula y el estadístico utilizado así como su distribución bajo la hipótesis nula (ya sea la exacta o la asintótica, según corresponda).

b) ¿A qué conclusión llega teniendo en cuenta que $\det(\mathbf{R}) = 0.469$ con nivel 0.05? Calcular el $p-$valor.

3. Consideremos los datos y supuestos del ejercicio anterior. Sea $\boldsymbol{\mu}_0 = (10, 25, 13, 31, 28, 9)^{\mathrm{T}}$

a) Obtenga un test para

$$H_0 : \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{pmatrix} = \begin{pmatrix} 10 \\ 25 \\ 13 \\ 31 \\ 28 \\ 9 \end{pmatrix} \quad \text{versus} \quad H_1; \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \mu_5 \\ \mu_6 \end{pmatrix} \neq \begin{pmatrix} 10 \\ 25 \\ 13 \\ 31 \\ 28 \\ 9 \end{pmatrix}$$

¿A qué conclusión llega a nivel 0.05?

Para responder a la pregunta, describa claramente

- *cuáles son los supuestos que está haciendo,*
- *cuáles son los vectores aleatorias en estudio y su distribución,*
- *que estadístico está utilizando,*
- *la distribución del estadístico bajo H_0 (ya sea la exacta o la asintótica, según corresponda) y la región de rechazo del test*
- *el $p-$valor obtenido y la conclusión que saca.*

b) Considere ahora el problema de testear

$$H_{0,1} : \mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6 = 116 \quad \text{versus} \quad H_{1,1} : \mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6 \neq 116\,.$$

i. Muestre que $H_{0,1}$ puede escribirse como $\mathbf{1}^{\mathrm{T}}\boldsymbol{\mu} = \mathbf{1}^{\mathrm{T}}\boldsymbol{\mu}_0$ donde $\mathbf{1} = (1, 1, 1, 1, 1, 1)^{\mathrm{T}}$.

ii. ¿A qué conclusión llega a nivel 0.05?

Para responder a la pregunta, nuevamente describa claramente

- *cuáles son los supuestos que está haciendo,*
- *cuáles son los vectores o variables aleatorias en estudio y su distribución,*
- *que estadístico está utilizando,*
- *la distribución del estadístico bajo H_0 (ya sea la exacta o la asintótica, según corresponda) y la región de rechazo del test*
- *el $p-$valor obtenido y la conclusión que saca.*

iii. Compare con la conclusión obtenida en el item anterior y explique a qué se debe la diferencia si existe.

iv. Defina ahora $H_{0,\mathbf{a}}$: $\mathbf{a}^{\mathrm{T}}\mu = \mathbf{a}^{\mathrm{T}}\mu_0$ para $\mathbf{a} \neq \mathbf{0}$.
¿En que dirección $\mathbf{a}$ Ud. espera obtener la misma conclusión que en el item a)?

4. Supongamos que tenemos $\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i} \sim N_d(\mu_i, \mathbf{\Sigma}_i)$, $1 \leq i \leq k$ independientes.

a) Supongamos que

$$\mathbf{\Sigma}_1 = \sigma_1^2 \mathbf{\Sigma}_0 \,, \quad \mathbf{\Sigma}_2 = \sigma_2^2 \mathbf{\Sigma}_0, \quad \ldots, \quad \mathbf{\Sigma}_k = \sigma_k^2 \mathbf{\Sigma}_0 \,,$$

donde $\mathbf{\Sigma}_0$ es una matriz simétrica y definida positiva fija y conocida.

Hallar los estimadores de máxima verosimilitud de $\mu_1, \ldots, \mu_k, \sigma_1^2, \ldots, \sigma_k^2$.

b) Supongamos que se quiere testear

$$H_0 : \quad \mathbf{\Sigma}_j = \sigma_j^2 \mathbf{\Sigma}_0 \quad \text{para} \quad 1 \leq j \leq k$$

i. Hallar el estadístico del cociente de máxima verosimilitud γ^*.

ii. Expresar γ^* en función de los autovalores de $\mathbf{\Sigma}_0^{-1}\mathbf{Q}_i$, $1 \leq i \leq k$, donde

$$\mathbf{Q}_i = \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)(\mathbf{x}_{ij} - \overline{\mathbf{x}}_i)^{\mathrm{T}}$$

iii. ¿Cuál es la distribución asintótica de $-2\ln\gamma^*$ bajo H_0?

iv. Construir un test de nivel aproximado α para testear esta hipótesis.

Para responder a lo pedido, nuevamente describa claramente

- *cuáles son los supuestos que está haciendo,*
- *cuáles son los vectores o variables aleatorias en estudio y su distribución,*
- *que estadístico está utilizando,*
- *la distribución del estadístico bajo H_0 (ya sea la exacta o la asintótica, según corresponda) y la región de rechazo del test.*

b) Ho: $\nu_i = \mu_i + a + b i$, $1 \leq i \leq d$, a y b ctes desconocidas

Se define $A : \mathbb{R}^d \to \mathbb{R}^{d-2}$

$$u = A.x = \begin{pmatrix} u_1 \\ \vdots \\ u_{d-2} \end{pmatrix} \quad \text{con } u_i = x_{i+2} - 2x_{i+1} + x_i$$

$$v = A.y = \begin{pmatrix} v_1 \\ \vdots \\ v_{d-2} \end{pmatrix} \quad \text{con } v_i = y_{i+2} - 2y_{i+1} + y_i$$

$$\begin{pmatrix} u_1 \\ \vdots \\ u_{d-2} \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_d \end{pmatrix} \rightarrow A = \begin{pmatrix} 1 & -2 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \ddots & & \vdots \\ \vdots & & & \ddots & & & 0 \\ 0 & \cdots & \cdots & 0 & 1 & -2 & 1 \end{pmatrix}$$

$A \in \mathbb{R}^{(d-2)\times d}$ y es triangular. Entonces

$rg(A) = d-2$ (Se puede observar que es de rango completo, sus filas son L.I)

i. $X \sim N_d(\mu, \Sigma) \to u = Ax \sim N_{d-2}(A\mu, A\Sigma A^T)$

$Y \sim N_d(\nu, \Sigma) \to v = A.Y \sim N_{d-2}(A.\nu, A\Sigma A^T)$

(la dimension de los nuevos vectores u y v será igual al $rg(A) = d-1$)

ii Bajo Ho, $\nu - \mu = a 1_d + b \begin{pmatrix} 1 \\ 2 \\ \vdots \\ d \end{pmatrix}$

$$E(v-u) = E(v) - E(u) = A\nu - A\mu$$
$$= A(\nu - \mu) \quad \Big\} \text{ Ho cierta}$$
$$= A . \begin{pmatrix} a+b \\ a+2b \\ \vdots \\ a+db \end{pmatrix}$$

$\longrightarrow$ sigue

$$\rightarrow E(v-u) = \underbrace{\begin{pmatrix} 1 & -2 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 & 0 \\ \vdots \\ 0 & \cdots & & & 1 & -2 & 1 \end{pmatrix}}_{(d-2)\times d} \underbrace{\begin{pmatrix} a+b \\ a+2b \\ a+3b \\ \vdots \\ a+db \end{pmatrix}}_{d\times 1}$$

$$= \underbrace{\begin{pmatrix} a+b-2(a+2b)+a+3b \\ a+2b-2(a+3b)+a+4b \\ \vdots \\ a+(d-2)b-2(a+(d-1)b)+a+db \end{pmatrix}}_{(d-2)\times 1} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

iii. Teniendo en cuenta los resultados de los items anteriores, podemos reescribir Ho de la siguiente manera:

$H_0: \quad E(v) - E(u) = 0$

es decir

$\mathbf{H_0}: \quad A.\nu = A.\mu \quad vs \quad H_1: A.\nu \neq A\mu$

Ahora el desarrollo es similar al del item ⓐ del ejercicio, pero los vectores en estudio son:

$Ay = v \sim N_{d-2}(A\nu, A\Sigma A^T)$

$Ax = u \sim N_{d-2}(A\mu, A\Sigma A^T)$

v y u son independientes ya que v es solo funcion de y, u es solo funcion de x (x e y indep)

Por lo tanto

$$S = \frac{(n-1)(A S_y A^T) + (m-1)(A S_x A^T)}{n+m-2}$$

$$= A \underset{\text{anterior}}{\tilde{S}} A^T$$

donde S_x y S_y son los definidos en el item ⓐ

[redacted] 1

1) $X_1, \dots X_m \quad N_d(\mu, \Sigma)$
$Y_1, \dots Y_n \quad N_d(\nu, \Sigma)$ muestras independientes

a)

$H_0: \nu_i = \mu_i + a_0 + b_0 i \quad , \; 1 \leq i \leq d, \; a_0$ y b_0 ctes conocidas

Voy a reescribir Ho para poder luego definir el test de nivel exacto

$H_0: \nu_i = \mu_i + a_0 + b_0 i \; , \; 1 \leq i \leq d$ es lo mismo que:

$$\begin{cases} \nu_1 - \mu_1 = a_0 + b_1 \cdot 1 \\ \nu_2 - \mu_2 = a_0 + b_1 \cdot 2 \\ \vdots \\ \nu_d - \mu_d = a_0 + b_1 \cdot d \end{cases} \Rightarrow \nu - \mu = a_0 \cdot 1_d + b_0 \begin{pmatrix} 1 \\ 2 \\ 3 \\ \vdots \\ d \end{pmatrix} = \eta_0 \checkmark$$

$\Rightarrow$ donde η_0 es un vector formado por constantes conocidas

Por lo tanto, la hipotesis resulta

$$H_0: \nu - \mu \overset{!}{=} \eta_0 \quad \text{vs} \quad H_1: \nu - \mu \neq \eta_0 \checkmark$$

Ahora busco un estadistico para poder construir el test pedido.

$X_1, \dots X_m$ muestra aleatoria $N_d(\mu, \Sigma) \Rightarrow \bar{X} \sim N_d(\mu, \frac{1}{m}\Sigma)$

$Y_1, \dots Y_n$ muestra aleatoria $N_d(\nu, \Sigma) \Rightarrow \bar{Y} \sim N_d(\nu, \frac{1}{n}\Sigma)$

$\bar{X}$ e $\bar{Y}$ son independientes (por ser funciones de muestras independientes) entonces

$$\bar{Y} - \bar{X} \sim N_d\left(\nu - \mu, \left(\tfrac{1}{m} + \tfrac{1}{n}\right)\Sigma\right)$$

$$(\bar{Y} - \bar{X}) - (\nu - \mu) \sim N_d\left(0, \left(\tfrac{1}{m} + \tfrac{1}{n}\right)\Sigma\right)$$

$$\frac{(\bar{Y} - \bar{X}) - (\nu - \mu)}{\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N_d(0, \Sigma) \checkmark$$

Por otro lado, defino

$$S = \frac{(n-1)S_Y + (m-1)S_X}{n+m-2}$$

$S_Y = \frac{Q_Y}{n-1}$, $Q_Y = \sum_{i=1}^{n} (Y_i - \bar{Y})(Y_i - \bar{Y})^T$, $Q_Y \sim \mathcal{W}(\Sigma, d, n-1)$

$S_X = \frac{Q_X}{m-1}$, $Q_X = \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T$, $Q_X \sim \mathcal{W}(\Sigma, d, m-1)$

S_X y S_Y indep., entonces por propiedad de las Wishart $(n+m-2) \cdot S \sim \mathcal{W}(\Sigma, d, n+m-2)$

Entonces si defino un estadístico (así como está no es el estadístico del test)

$$T^2 = \frac{n \cdot m}{n+m} \left(\bar{Y} - \bar{X} - (\nu - \mu)\right)^T S^{-1} \left(\bar{Y} - \bar{X} - (\nu - \mu)\right)$$

$$T^2 \sim \mathcal{H}(d, n+m-2)$$

Además $\frac{(n+m-2) - d + 1}{d(n+m-2)} \cdot T^2 \sim \mathcal{F}_{d;\, n+m-d-1}$

Volviendo a lo pedido

$H_0: \nu - \mu = \eta_0$ vs $H_1: \nu - \mu = \eta_0$

Entonces, bajo H_0

$$T_0^2 = \frac{n \cdot m}{n+m} (\bar{Y} - \bar{X} - \eta_0)^T S^{-1} (\bar{Y} - \bar{X} - \eta_0) \sim \mathcal{H}(d, n+m-2)$$

¡ojo!

y el test:

$$\phi(\underline{x}) = \begin{cases} 1 & \text{si} \quad T_0^2 > t^2_{d,\, n+m-2}(\alpha) \\ 0 & \text{si} \quad T_0^2 \leq t^2_{d,\, n+m-2}(\alpha) \end{cases}$$

donde $F_0 = \frac{n+m-d-1}{d(n+m-2)} T_0^2 \sim \mathcal{F}_{d,\, n+m-d-1}$
(bajo H_0)

y la región de rechazo será $F_0 > f_{d,\, n+m-d-1}(\alpha)$

(es decir: puedo plantear el test con el estadístico T_0^2 ó F_0, la conclusión será por supuesto la misma)

El estadístico a utilizar será

$$T^2 = \frac{n \cdot m}{n+m}\left(\left(A\bar{y} - A\bar{x}\right) - \left(A\nu - A\mu\right)\right)^T S^{-1}\left(\left(A\bar{y} - A\bar{x}\right) - \left(A\nu - A\mu\right)\right)$$

$$T^2 \sim \mathcal{H}(d-2, n+m-2)$$

Ho: $A\nu - A\mu = 0$ vs H_1 $A\nu - A\mu \neq 0$

entonces, Bajo Ho

$$T_0^2 = \frac{n \cdot m}{n+m}\left(\bar{y} - \bar{x}\right)^T A^T S^{-1} A\left(\bar{y} - \bar{x}\right) \sim \mathcal{H}(d-2, n+m-2)$$

y

$$F_0 = \frac{(n+m-2)-(d-2)+1}{(n+m-2)(d-2)} \cdot T_0^2 \sim \mathcal{F}_{d-2,\, n+m-d+1}$$

El test resulta

$$\phi(\underline{x}) = \begin{cases} 1 & \text{si} \quad F_0 > f_{d-2,\, n+m-d+1,}(\alpha) \\ 0 & \text{si} \quad F_0 \leq f_{d-2,\, n+m-d+1}(\alpha) \end{cases}$$

② x_{ij} : medición del operario i en la tarea j

$1 \leq i \leq 76$, $1 \leq j \leq 6$.

Se quiere testear la hipótesis de que las variables son independientes.

Si $X_1, \ldots X_{76}$ son una muestra aleatoria de manera que $X_i \sim N_6(\mu, \Sigma)$ (cada uno de los 6 elementos del vector X_i indica la medición del operario i en cada una de las 6 tareas). Entonces, para probar que las variables x_{ij} son independientes, podría realizar un test para la matriz de covarianza. Porque si Σ fuera una matriz diagonal, entonces las variables serían independientes (por ser normales)

$$H_0: \Sigma = diag(\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{44}, \sigma_{55}, \sigma_{66})$$

Aplicando el test de cociente de verosimilitud resulta el

$$\gamma^* = \left(\frac{|Q|}{\prod_{i=1}^{6} |q_{ii}|} \right)^{n/2}$$

La distribución asintótica bajo H_0 resulta

$$-2\ln(\gamma^*) \xrightarrow{D} \chi^2_D$$

donde $D = \dim \Omega - \dim \Omega_0$.

$\Omega = \{ (\mu, \Sigma), \mu \in \mathbb{R}^6, \Sigma > 0 \}$ $\dim \Omega = 6 + \frac{6 \cdot (6+1)}{2} = 27$

$\Omega_0 = \{ (\mu, \Sigma_0), \mu \in \mathbb{R}^6, \Sigma_0 = diag(\sigma_{11} \ldots \sigma_{66}), \sigma_{ii} > 0 \}$

$\dim \Omega_0 = 6 + 6 = 12$

$\rightarrow D = 27 - 12 = 15$

El test resulta

$$\phi(\underline{X}) = \begin{cases} 1 & \text{si} \quad -2\ln(\gamma^*) \geq \chi^2_{15,\alpha} \\ 0 & \text{si} \quad -2\ln(\gamma^*) < \chi^2_{15,\alpha} \end{cases}$$

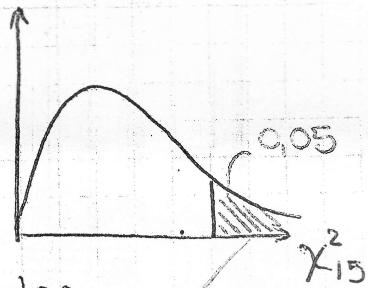
Es un test asintótico de nivel α.

b. Como se probó en clase y en la práctica 3-parte 2, $(\det(R))^{n/2} = \gamma^*$

Por lo tanto, puedo usar el resultado $\det(R) = 0{,}469$ para concluir.

$n = 76$, $\alpha = 0{,}05$

$$-2\ln\left[\det(R)^{76/2}\right] = 57{,}54359$$



$$\chi^2_{15,0{,}05} = 24{,}99$$

$\rightarrow$ Rechazo Ho, las variables no son independientes

Si defino $U \sim \chi^2_{15}$, calculo el p-valor

$$p\text{-}V = P(U > 57{,}54359) = 6{,}63 \times 10^{-7}$$

③ Continuación del ejercicio anterior.

$X_i \sim N_6(\mu, \Sigma)$, $1 \leq i \leq 76$

$H_0: \mu = \mu_0$ Vs $H_1: \mu \neq \mu_0$ $\alpha = 0{,}05$

donde $\mu_0 = (10, 25, 13, 31, 28, 9)^T$

$$\phi(\underline{x}) = \begin{cases} 1 & \text{si } T_0^2 > t^2_{6;76-1}(0{,}05) \\ 0 & \text{si } T_0^2 \leq t^2_{6;75}(0{,}05) \end{cases}$$

donde $T_0^2 = 76(\bar{x} - \mu_0)^T S^{-1}(\bar{x} - \mu_0)$

Bajo H_0 $T_0^2 \sim t^2_{6,75}$

y $\frac{76-6}{6(76-1)} \cdot T_0^2 \sim \mathcal{F}_{6;76-6}$

Con los datos de la muestra (calculos en R)

$$F_0 = \frac{70}{6 \cdot 75} \cdot 76(\bar{x} - \mu_0)^T S^{-1}(\bar{x} - \mu_0) = 2{,}24$$

$$f_{6,70,0{,}05} = 2{,}23$$

$$\text{p-Valor} = P(\mathcal{F}_{6,70} > 2{,}24) = 0{,}0489$$

Conclusión: $F_0 = 2{,}24 > 2{,}23$, $P\text{-}V = 0{,}0489 < 0{,}05$

→ Rechazo H_0 y con un nivel 0,05 puedo decir que el vector μ no es $(10, 25, 13, 31, 28, 9)^T$

▸ Supuestos: $X_1, \ldots X_{76}$ son una muestra aleatoria $N_6(\mu, \Sigma)$, $S = \frac{Q}{76-1}$ donde $Q = \sum_{i=1}^{76} (x_i - \bar{x})(x_i - \bar{x})^T$

y $Q \sim W(\Sigma, \underset{P}{\underset{\downarrow}{6}}, \underset{n-1}{\underset{\downarrow}{75}})$, $\bar{x} \sim N_6(\mu, \frac{1}{76}\Sigma)$

indep.

b)

$H_{01}: \mu_1+\mu_2+\mu_3+\mu_4+\mu_5+\mu_6 = 116$ Vs $H_{11}: \sum_{i=1}^{6} \mu_i \neq 116$

$$1^T\mu = (1\ 1\ 1\ 1\ 1\ 1)\begin{pmatrix}\mu_1\\ \vdots \\ \mu_6\end{pmatrix} = \mu_1+\mu_2+\mu_3+\mu_4+\mu_5+\mu_6.$$

$$1^T\mu_0 = (1\,1\,1\,1\,1\,1)\begin{pmatrix}10\\25\\13\\31\\28\\9\end{pmatrix} = 10+25+13+31+28+9 = 116.$$

$\rightarrow H_{0,1}: 1^T\mu = 1^T\mu_0$

Para armar el test puedo definir el vector

$$y = 1^T X \sim N(1^T\mu,\ 1^T\Sigma 1)$$

de esta manera puedo usar el estadístico

$$T^2 = n(1^T\bar{x} - 1^T\mu)^T(1^T S 1)^{-1}(1^T\bar{x} - 1^T\mu) \sim \mathcal{H}(1, 76-1)$$

Con $\bar{x}$ y S los variables planteados en el ítem anterior

(También se puede plantear un test univariado pero lo planteo de esta manera para poder usar los resultados anteriores)

ok

Por lo tanto, Bajo Ho:

$$T_0^2 = 76(1^T\bar{x} - 1^T\mu_0)^T(1^T S 1)^{-1}(1^T\bar{x} - 1^T\mu_0) \sim \mathcal{H}(1,75)$$

$$F_0 = \frac{76-1}{76-1}\cdot T^2 \sim F_{1,75}$$

$$\phi(\underline{x}) = \begin{cases} 1 & \text{si} \quad F_0 > f_{1,75}(\alpha) \\ 0 & \text{si} \quad F_0 \leq f_{1,75}(\alpha)\end{cases}$$

$F_0 = 0,012$

$f_{1,75(0,05)} = 3,96$

$p\text{-}V = P(F_{1,75} > 0,012) = 0,9113$

En base a los resultados no Rechazo Ho, y no tengo evidencia para decir que la suma no es 116.

iii En el item a decidí que μ no era μ_0 pero en el ii no pude rechazar que la suma de los μ_i era 116 (resultado que se obtiene de sumar los elementos de μ_0). Pero que la suma de los μ_i de 116 da lugar a infinitos vectores μ_0, no significa que justo va a ser el μ_0 del item a, por eso tiene sentido que en uno rechazo y en el otro no. ok

IV. $H_{0,a}: a^T\mu = a^T\mu_0 \;,\; a \neq 0$

Para obtener la misma conclusion que en a, es decir, rechazar Ho, a debe ser la dirección en la que la diferencia se hace máxima, y eso ocurre en la dirección

$$\hat{a} = S^{-1}(\bar{X} - \mu_0)$$

Igual se esperaba que hallaran el $\hat{a}$ con la computadora

~~SEMINA~~

4) $x_{i1}, \dots x_{in_i} \sim N_d(\mu_i, \Sigma_i)$; $1 \leq i \leq k$ indep.

a) Sup: $\Sigma_1 = \sigma_1^2 \Sigma_0$, $\Sigma_2 = \sigma_2^2 \Sigma_0$, $\dots$ $\Sigma_K = \sigma_K^2 \Sigma_0$

con $\Sigma_0 > 0$, simétrica y conocida.

EMV de $\mu_1, \dots, \mu_K, \sigma_1^2, \dots \sigma_K^2$

La función a maximizar es

$$f_{\underline{X}}(\underline{x}, \mu, \sigma^2) = \prod_{i=1}^{k} \prod_{j=1}^{n_i} \frac{1}{(2\pi)^{d/2}} \frac{1}{|\Sigma_i|^{1/2}} e^{-\frac{1}{2}(x_{ij}-\mu_i)^T \Sigma_i^{-1} (x_{ij}-\mu_i)}$$

$$= \prod_{i=1}^{k} \frac{1}{(2\pi)^{\frac{dn_i}{2}}} \frac{1}{|\Sigma_i|^{\frac{n_i}{2}}} e^{-\frac{1}{2}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^T \Sigma_i^{-1} (x_{ij}-\mu_i)}$$

busco $\mu_1, \dots \mu_K, \sigma_1^2, \dots \sigma_K^2$ tal que $f_{\underline{X}}$ sea máxima

$$\max_{\substack{\mu_1, \dots \mu_K, \sigma_1^2, \dots \sigma_K^2 \\ >0 \quad >0}} \prod_{i=1}^{k} \frac{1}{(2\pi)^{\frac{dn_i}{2}}} \frac{1}{|\Sigma_i|^{\frac{n_i}{2}}} e^{-\frac{1}{2}\sum_{j=1}^{n_i}(x_{ij}-\mu_i)^T \Sigma_i^{-1} (x_{ij}-\mu_j)}$$

Tengo que maximizar la productoria, pero cada ~~elemento~~ (termino) i, $1 \leq i \leq k$, de la productoria depende solo de μ_i y σ_i^2, entonces para maximizar toda la función puedo maximizar por separado cada término y el producto será máximo cuando todos los términos sean máximos.

Además el desarrollo será igual para todos, así que desarrollo para 1 ok

$$\max_{\mu_1, \sigma_1^2 > 0} \frac{1}{(2\pi)^{\frac{n_1 d}{2}}} \cdot \frac{1}{|\sigma_1^2 \Sigma_0|^{n_1/2}} e^{-\frac{1}{2} \underbrace{\sum_{j=1}^{n_1} (x_{1j}-\mu_1)^T \Sigma_1^{-1} (x_{1j}-\mu_1)}_{\in \mathbb{R}, \text{ Aplico la Traza y da lo mismo}}}$$

$$= \max_{\mu_1, \sigma_1^2 > 0} \frac{1}{(2\pi)^{\frac{n_1 d}{2}}} \frac{1}{|\sigma_1^2 \Sigma_0|^{\frac{n_1}{2}}} e^{-\frac{1}{2} Tr\left(\Sigma_1^{-1} \cdot \underbrace{\sum_{j=1}^{n_1} (x_{1j}-\mu_1)(x_{1j}-\mu_1)^T}_{*}\right)}$$

(*) Desarrollo la sumatoria

$$\sum_{j=1}^{n_1} (x_{1j} - \bar{x}_1 + \bar{x}_1 - \mu_1)(x_{1j} - \bar{x}_1 + \bar{x}_1 - \mu_1)^T$$

$$= \sum_{j=1}^{n_1} (x_{1j}-\bar{x}_1)(x_{1j}-\bar{x}_1)^T + \cancel{\sum_{j=1}^{n_1} (x_{1j}-\bar{x}_1)(\bar{x}_1-\mu_1)^T}_{0}$$

$$+ \cancel{\sum_{j=1}^{n_1} (\bar{x}_1-\mu_1)(x_{1j}-\bar{x}_1)}_{0} + n_1(\bar{x}_1-\mu_1)(\bar{x}_1-\mu_1)^T$$

$$= Q_1 + n_1(\bar{x}_1-\mu_1)(\bar{x}_1-\mu_1)^T$$

$$= \max_{\mu_1, \sigma_1^2 > 0} \frac{1}{(2\pi)^{\frac{n_1 d}{2}}} \cdot \frac{1}{|\sigma_1^2 \Sigma_0|^{n_1/2}} e^{-\frac{1}{2\sigma_1^2} Tr\left(\Sigma_0^{-1} Q_1 + \Sigma_0^{-1} n_1 (\bar{x}_1-\mu_1)(\bar{x}_1-\mu_1)^T\right)}$$

$$= \frac{1}{(2\pi)^{\frac{n_1 d}{2}}} \max_{\sigma_1^2 > 0} \left\{ \frac{1}{|\sigma_1^2 \Sigma_0|^{n_1/2}} e^{-\frac{1}{2\sigma_1^2} Tr(\Sigma_0^{-1} Q_1)} \cdot \max_{\mu_1} e^{-\frac{1}{2\sigma_1^2} Tr\left(\Sigma_0^{-1} n_1 (\bar{x}_1-\mu_1)(\bar{x}-\mu_1)^T\right)} \right.$$

lo separé de esa forma para simplificar el cálculo, de esta manera, para cada σ_1^2 fijo, el maximo para μ se encuentra cuando minimizo el exponente.

$$Tr\left((\bar{x}_1-\mu_1)^T \Sigma_0^{-1} (\bar{x}-\mu_1) \right).$$

~~Jornina~~

Ahora, como Σ_0 es definida positiva,

$$(\bar{x}_1-\mu_1)^T \Sigma_0^{-1} (\bar{x}_1-\mu_1) \geq 0 \text{ siempre},$$

los autovalores seran entonces no nulos,

al igual que la traza (que es la suma λ_i)

Valdrá 0 cuando $\mu_1 = \bar{x}_1$

Por lo tanto los estimadores de MV para los μ_i seran $\hat{\mu}_i = \bar{X}_i \quad 1 \leq i \leq K$

Ahora busco el maximo para σ_1^2, es decir

$$\underbrace{\max_{\sigma_1^2>0} \frac{1}{(\sigma_1^2)^{\frac{n_1 d}{2}} |\Sigma_0|^{n/2}} e^{-\frac{1}{2\sigma_1^2} Tr(\Sigma_0^{-1} Q_1)}}_{L}$$

(alcanza con maximizar esa parte).

$$\ln(L) = -\frac{n_1 d}{2} \ln(|\Sigma_0|) - \frac{n_1}{2} \ln(\sigma_1^2) - \frac{1}{2\sigma_1^2} Tr(\Sigma_0^{-1} Q_1)$$

$$\frac{\partial \ln(L)}{\partial \sigma_1^2} = -\frac{n_1 d}{2} \cdot \frac{1}{\sigma_1^2} + \frac{1}{2(\sigma_1^2)^2} Tr(\Sigma_0^{-1} Q_1) = 0$$

$$\longrightarrow \hat{\sigma}_1^2 = \frac{Tr(\Sigma_0^{-1} Q_1)}{n_1 \cdot d}$$

de la misma forma resultará

$$\hat{\sigma}_i^2 = \frac{Tr(\Sigma_0^{-1} Q_i)}{n_i \cdot d}$$

$i = 1, 2, \ldots K$

$$\hat{\mu}_i = \bar{X}_i$$

b) Queremos testear

$$H_0: \Sigma_j = \sigma_j^2 \Sigma_0 \qquad 1 \leq j \leq k.$$

$$\lambda^* = \frac{\max_{\mu, \Sigma \in H_0} f_{\underline{X}}(\underline{x}, \mu, \Sigma)}{\max_{\mu, \Sigma > 0} f_{\underline{X}}(\underline{x}, \mu, \Sigma)} = \frac{L_0}{L_1}$$

Los valores que maximizan el numerador los calcule en el item anterior:

$$L_0 = f_{\underline{X}}(\underline{x}, \hat{\mu}_1, \dots \hat{\mu}_k, \hat{\sigma}_1^2, \dots, \hat{\sigma}_k^2)$$

$$= \prod_{i=1}^{k} \frac{1}{(2\pi)^{\frac{d n_i}{2}}} \cdot \frac{1}{\left| \frac{Tr(\Sigma_0^{-1} Q_i)}{n_i d} \cdot \Sigma_0 \right|^{n_i/2}} e^{-\frac{n_i d}{2}}$$

Si $n = n_1 + n_2 + \dots n_k$

$$L_0 = \frac{1}{(2\pi)^{\frac{dn}{2}}} e^{-\frac{nd}{2}} \cdot \frac{1}{\prod_{i=1}^{k} \left| \frac{Tr(\Sigma_0^{-1} Q_i)}{n_i d} \cdot \Sigma_0 \right|^{n_i/2}}$$

$$L_1 = \max_{\mu, \Sigma_i > 0} f_{\underline{X}}(\underline{x}, \mu, \Sigma)$$

tengo que buscar el máximo para $\mu_1 \dots \mu_k, \Sigma_1, \dots \Sigma_k$

Al desarrollarlo pasará lo mismo que antes, me quedará un producto que para maximizarlo puedo maximizar cada elemento o término de la productoria, y de cada uno ya se cuanto vale el máximo (el desarrollo es similar al anterior, pero sin descomponer la matriz Σ_i)

~~[illegible]~~

resulta que el max, se alcanza en $\mu_i = \bar{x}_i$ y $\Sigma_i = \frac{Q_i}{n_i}$

para $i = 1, 2, \dots k$

$$L_1 = \prod_{i=1}^{k} \frac{1}{(2\pi)^{\frac{n_i d}{2}}} \cdot \frac{1}{\left|\frac{Q_i}{n_i}\right|^{n_i/2}} \cdot e^{-\frac{n_i d}{2}} = \frac{1}{(2\pi)^{\frac{nd}{2}}} \cdot e^{-nd/2} \cdot \frac{1}{\prod_{i=1}^{k} \left|\frac{Q_i}{n_i}\right|^{n_i/2}}$$

Finalmente resulta

$$\gamma^* = \frac{L_0}{L_1} = \frac{\prod_{i=1}^{k} \left|\frac{Q_i}{n_i}\right|^{n_i/2}}{\prod_{i=1}^{k} \left|\frac{Tr(\Sigma_0^{-1} Q_i)}{n_i \cdot d} \Sigma_0\right|^{n_i/2}} = \prod_{i=1}^{k} \left(\frac{\left|\frac{Q_i}{n_i}\right|}{\left|\frac{Tr(\Sigma_0^{-1} Q_i)}{n_i \cdot d} \cdot \Sigma_0\right|} \right)^{n_i/2}$$

ii) Para expresar a γ^* en función de los autovalores de $\Sigma_0^{-1} Q_i$ simplifico y reordeno el resultado anterior

$$\gamma^* = \prod_{i=1}^{k} \left(\frac{|\Sigma_0^{-1} Q_i|}{\left(\frac{Tr(\Sigma_0^{-1} Q_i)}{d}\right)^d} \right)^{n_i/2} =$$

si $\lambda_1, \dots \lambda_d$ son los autovalores de $\Sigma_0^{-1} Q_i$ entonces

$$|\Sigma_0^{-1} Q_i| = \prod_{j=1}^{d} \lambda_j$$

$$tr(\Sigma_0^{-1} Q_i) = \sum_{j=1}^{d} \lambda_j$$

$$\rightarrow \gamma^* = \prod_{i=1}^{k} \left(\frac{\left(\prod_{j=1}^{d} \lambda_j\right)^{1/d}}{\sum \lambda_j / d} \right)^{\frac{n_i d}{2}}$$

iii. Bajo Ho, $-2\ln(\gamma^*) \xrightarrow{D} \chi^2_D$

$$D = \dim \Omega - \dim \Omega_0 = k \cdot d + \frac{d(d+1)}{2} \cdot k - (k \cdot d + k)$$

$$D = d(d+1)\frac{k}{2} - k$$

IV. Un test de nivel aproximado α para testear Ho será

$$\phi(\underline{x}) = \begin{cases} 1 & \text{si } -2\ln(\gamma^*) \geq \chi^2_{d(d+1)\frac{k}{2}-k,\,\alpha} \\ 0 & \text{si } -2\ln(\gamma^*) < \chi^2_{d(d+1)\frac{k}{2}-k,\,\alpha} \end{cases}$$