3. ANÁLISIS DE CORRESPONDENCIA

• PREGUNTA 1

(a) Plantea los problemas de independencia de caracteres y homogeneidad de poblaciones.

* INDEPENDENCIA DE CARACTERES

Consideramos dos variables categóricas (A, B), donde

$$\begin{cases} A \text{ con } n \text{ modalidades} : A_1, \ldots, A_n \\ B \text{ con } p \text{ modalidades} : B_1, \ldots, B_p . \end{cases}$$

El problema consiste en comprobar la hipótesis de independencia

$$H_0 : \mathbb{P}(A_i \cap B_j) = \mathbb{P}(A_i)\,\mathbb{P}(B_j), \quad \forall i, j$$

(Para ello, seleccionamos una muestra m de tamaño N y observamos (A,B) en cada elemento).

* HOMOGENEIDAD DE POBLACIONES

Consideramos una variable categórica B con p modalidades $B_1, \ldots, B_p$, y estudiamos la homogeneidad de B en n poblaciones $(P_1, \ldots, P_n)$.

El problema consiste en comprobar la hipótesis de homogeneidad

$$H_0 : P_1(B_j) = \ldots = P_n(B_j) \; (= P(B_j)), \; j = 1, \ldots, p .$$

(Para ello seleccionamos de forma independiente una muestra m_i de tamaño N_i de cada población).

(b) Demuestra que los estadísticos χ^2 asociados a ambos problemas y sus distribuciones coinciden.

Para resolver el problema de independencia, seleccionamos una muestra m de tamaño N y observamos (A,B) en cada elemento.

$N_{ij} \equiv$ nº de elementos de m que presentan $A_i \cap B_j$
$(\sim Bi(N, \mathbb{P}(A_i \cap B_j)))$

donde, la tabla de contingencia (matriz de datos experimentales) es:

$$(*)\quad \begin{array}{c|ccccc|c}
 & B_1 & \dots & B_j & \dots & B_p & \text{Marginal de } A \\
\hline
A_1 & N_{11} & \dots & N_{1j} & \dots & N_{1p} & N_{1\cdot} \\
\vdots & \vdots & & \vdots & & \vdots & \vdots \\
A_i & N_{i1} & \dots & N_{ij} & \dots & N_{ip} & N_{i\cdot} = \sum_{j=1}^{p} N_{ij} \\
\vdots & \vdots & & \vdots & & \vdots & \vdots \\
A_n & N_{n1} & \dots & N_{nj} & \dots & N_{np} & N_{n\cdot} \\
\hline
\text{Marginal de } B. & N_{\cdot 1} & \dots & N_{\cdot j} = \sum_{i=1}^{n} N_{ij} & \dots & N_{\cdot p} &
\end{array}$$

Luego, el estadístico chi-cuadrado para la hipótesis de independencia es:

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(N_{ij} - e_{ij})^2}{e_{ij}}, \quad \text{con } e_{ij} = \hat{E}_{H_0}(N_{ij})$$

Como $N_{ij} \underset{H_0}{\sim} Bi\left(N, \mathbb{P}(A_i)\mathbb{P}(B_j)\right)$, entonces:

$$e_{ij} = N \hat{P}_{H_0}(A_i) \hat{P}_{H_0}(B_j) = N \cdot \frac{N_{i\cdot}}{N} \frac{N_{\cdot j}}{N} = \frac{N_{i\cdot} N_{\cdot j}}{N}$$

Luego, $\chi^2 \xrightarrow{\mathcal{L}} \chi^2_{(n-1)(p-1)}$ $\quad [n(p-1) - (p-1)]$?

var. lin indep − lin. indep bajo H_0

Para resolver el pb de homogeneidad seleccionamos de forma indep. una muestra m_i de tamaño N_i de cada población. Ahora,

$$N_{ij} \equiv \text{nº de elementos de } m_i \text{ que presentan } B_j \quad (\sim Bi(N_i, P_i(B_j)))$$

$$\begin{array}{c|ccccc|c}
 & B_1 & \dots & B_j & \dots & B_p & \\
\hline
P_1 & N_{11} & \dots & N_{1j} & \dots & N_{1p} & N_{1\cdot} = N_1 \\
\vdots & \vdots & & \vdots & & \vdots & \vdots \\
P_i & N_{i1} & \dots & N_{ij} & \dots & N_{ip} & N_{i\cdot} = N_i \\
\vdots & \vdots & & \vdots & & \vdots & \vdots \\
P_n & N_{n1} & \dots & N_{nj} & \dots & N_{np} & N_{n\cdot} = N_n \\
\hline
 & N_{\cdot 1} & \dots & N_{\cdot j} & \dots & N_{\cdot p} &
\end{array}$$

$$N_{\cdot j} = \sum_{i=1}^{n} N_{ij}$$

El estadístico chi-cuadrado para la hipótesis de homogeneidad es:

$$\chi^2 = \sum_{i=1}^{n} \sum_{j=1}^{p} \frac{(N_{ij} - e_{ij})^2}{e_{ij}}, \quad \text{con } e_{ij} = \hat{\mathbb{E}}_{H_0}(N_{ij})$$

Como $N_{ij} \underset{H_0}{\sim} Bi\left(N_i, \mathbb{P}_i(B_j)\right) \Rightarrow e_{ij} = N_i \cdot \hat{\mathbb{P}}_{H_0}(B_j) = N_i \cdot \frac{N_{\cdot j}}{N}$

Luego, $\chi^2 \xrightarrow{\mathcal{L}} \chi^2_{(n-1)(p-1)} \leftarrow [n(p-1)] - (p-1)$

- PREGUNTA 2

Sean $r_1', \dots, r_n'$ las distribuciones condicionadas por filas asociadas a una tabla de contingencia.

(*)

(a) Define el centro de gravedad de las filas : m_r'

Transformamos la tabla de contingencia para obtener las distribuciones condicionadas por filas:

$$M_r = \begin{bmatrix} r_1' \\ \vdots \\ r_i' \\ \vdots \\ r_n' \end{bmatrix} = \begin{bmatrix} \frac{N_{11}}{N_{1\cdot}} & \dots & \frac{N_{1j}}{N_{1\cdot}} & \dots & \frac{N_{1p}}{N_{1\cdot}} \\ \vdots \\ \frac{N_{i1}}{N_{i\cdot}} & \dots & \frac{N_{ij}}{N_{i\cdot}} & \dots & \frac{N_{ip}}{N_{i\cdot}} \\ \vdots \\ \frac{N_{n1}}{N_{n\cdot}} & \dots & \frac{N_{nj}}{N_{n\cdot}} & \dots & \frac{N_{np}}{N_{n\cdot}} \end{bmatrix} \begin{bmatrix} \frac{N_{1\cdot}}{N} \\ \vdots \\ \frac{N_{i\cdot}}{N} \\ \vdots \\ \frac{N_{n\cdot}}{N} \end{bmatrix}$$

$$m_r' = \begin{bmatrix} \frac{N_{\cdot 1}}{N} & \dots & \frac{N_{\cdot j}}{N} & \dots & \frac{N_{\cdot p}}{N} \end{bmatrix}$$

Llamamos $r_{ij} = \frac{N_{ij}}{N_{i\cdot}}$, $f_{i\cdot} = \frac{N_{i\cdot}}{N}$, $f_{\cdot j} = \frac{N_{\cdot j}}{N}$

Por tanto, definimos el centro de gravedad m_r de las filas de M_r como :

$$m_r' = \sum_{i=1}^{n} f_{i\cdot}\, r_i' = [\, f_{\cdot 1} \dots f_{\cdot j} \dots f_{\cdot p} \,]$$

(b) Define la distancia chi-cuadrado entre dos distribuciones.

La distancia chi-cuadrado es una distancia entre distribuciones de probabilidad.
(Distancia chi-cuadrado entre dos filas de M_r):

$$d_{\chi^2}(r_i, r_{i'}) = \sum_{j=1}^{p} \frac{(r_{ij} - r_{i'j})^2}{f_{\cdot j}} = (r_i - r_{i'})' D_p^{-1} (r_i - r_{i'})$$

donde, $D_p = \operatorname{diag}(f_{\cdot 1}, \dots, f_{\cdot p})$

(c) Demuestra que $\chi^2 = N \sum_{i=1}^{n} f_{i\cdot}\, d_{\chi^2}(r_i, m_r)$

$$\chi^2 = \sum_{i=1}^{n}\sum_{j=1}^{p} \frac{(N_{ij}-e_{ij})^2}{e_{ij}} \underset{\substack{\uparrow \\ e_{ij} \text{ bajo } H_0}}{=} \sum_{i=1}^{n}\sum_{j=1}^{p} \frac{\left(N_{ij} - \frac{N_{i\cdot}\, N_{\cdot j}}{N}\right)^2}{\frac{N_{i\cdot}\, N_{\cdot j}}{N}} =$$

$$\underset{\substack{\uparrow \\ \text{sacamos} \\ N_{i\cdot}}}{=} \sum_{i=1}^{n}\sum_{j=1}^{p} \underbrace{N_{i\cdot}^{2}}_{= f_{i\cdot} N} \frac{\Big(\underbrace{\frac{N_{ij}}{N_{i\cdot}}}_{r_{ij}} - \underbrace{\frac{N_{\cdot j}}{N}}_{f_{\cdot j}}\Big)^2}{\cancel{N_{i\cdot}}\,\underbrace{\frac{N_{\cdot j}}{N}}_{= f_{\cdot j}}} =$$

$$= N \sum_{i=1}^{n} f_{i\cdot} \underbrace{\sum_{j=1}^{p} \frac{1}{f_{\cdot j}} (r_{ij} - f_{\cdot j})^2}_{d_{\chi^2}(r_i, m_r)} = N \sum_{i=1}^{n} f_{i\cdot}\, d_{\chi^2}(r_i, m_r)$$

- PREGUNTA 3

Describe el objetivo del análisis de correspondencia (por filas).

El objetivo del análisis de correspondencia por filas es representar los perfiles filas en un espacio de menor dimensión (generalmente 2) de forma que si

$$M_r = \begin{pmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{n1} & \cdots & r_{np} \end{pmatrix} \xrightarrow{\text{tranf.}} \begin{pmatrix} P_{r_{11}} & \cdots & P_{r_{1p}} \\ \vdots & & \\ P_{r_{n1}} & \cdots & P_{r_{np}} \end{pmatrix} = PM_r$$

buscamos la base de $\mathbb{R}^p$, $\{\underline{u}_1, \ldots, \underline{u}_p\}$ de esa transformación de forma que se verifique:

* $d_{\chi^2}(r_i, r_j) = d_e^2(P_{r_i}, P_{r_j})$
  Es decir, los perfiles próximos en la métrica chi-cuadrado tienen representaciones próximas en la distancia euclídea.

* Sean $v_1, \ldots, v_p$ las varianzas de las columnas de PM_r.
  Debe cumplirse que $v_1 \geq v_2 \geq \ldots \geq v_p$ y que las últimas varianzas sean próximas a 0.

(*) NOTA : $P_{r_i}^t = (r_i^t D_p^{-1} u_1, \ldots, r_i^t D_p^{-1} u_p)$

• PREGUNTA 4

Analogías y diferencias entre el ACP y el AC.

* Ambos procedimientos (tanto ACP como AC) buscan reducir la dimensión del espacio original de estudio. Para ello, se realizan transformaciones (cambios de base) que mantengan la máxima varianza en las primeras componentes o coordenadas.

* Sin embargo, en el ACP se verifica que la distancia euclídea de los individuos originales y transformados coinciden, mientras que en el AC, la distancia chi-cuadrado de las coordenadas originales (distribuciones) coincide con el cuadrado de la distancia euclídea de las coordenadas en la nueva base.
  Además, en el ACP, las coordenadas principales deben estar incorreladas. (componentes)

• PREGUNTA 5

Plantea y resuelve el problema de optimización al que conduce el AC.

El problema que se plantea es determinar los ejes $u_1, \dots, u_p$ unitarios y ortogonales, que maximicen la varianza (ponderada) de las proyecciones.

(1) Primer eje (u_1)

$$\max_{u_1} \quad u_1^t D_p^{-1} S_r u_1$$
$$\text{s.a.} \quad u_1^t D_p^{-1} u_1 = 1$$

(2) segundo eje (u_2)

$$\max_{u_2} \quad u_2^t D_p^{-1} S_r u_2$$
$$\text{s.a:} \quad u_2^t D_p^{-1} u_2 = 1$$
$$u_2^t D_p^{-1} u_1 = 0$$

Así se procede con los demás u_k, $k = 1, \dots, p$.

Solución de (1):

Sea $L(u_1, \lambda) = u_1^t D_p^{-1} S_r u_1 - \lambda (u_1^t D_p^{-1} u_1 - 1)$

$$\frac{\partial L}{\partial u_1} = \cancel{2} D_p^{-1} S_r u_1 - \cancel{2} \lambda D_p^{-1} u_1 = 0_{(p)}$$

$$\Rightarrow D_p^{-1} S_r u_1 = D_p^{-1} \lambda u_1 \Rightarrow \boxed{S_r u_1 = \lambda u_1}$$

(λ, u_1) autovalor/vec de S_r

Multiplicando por $u_1^t D_p^{-1}$:

$$u_1^t D_p^{-1} S_r u_1 = u_1^t D_p^{-1} \lambda u_1 = \lambda \underbrace{u_1^t D_p^{-1} u_1}_{=1}$$

$$\Rightarrow u_1^t D_p^{-1} S_r u_1 = \lambda$$

Como estamos maximizando, $\lambda \equiv$ máx. autovalor de S_r

$\Rightarrow u_1 \equiv$ autovector asociado al máximo autovalor de S_r.

• PREGUNTA 7

Concepto de vértice en el problema del AC por filas. Interpretación.

Los vértices de las filas son una distribución de probabilidad degenerada que concentra toda la información en la i-ésima fila o categoría.

(B_1) $v_{f_1}^t = (1, 0, \dots, 0)_{p\times 1}$ $\longrightarrow$ concentra toda la probabilidad en la 1ª categoría

$\vdots$

(B_i) $v_{f_i}^t = (0, \dots, \overset{i}{1}, 0, \dots, 0)_{p\times 1}$

"Cada vértice v_{f_i} es una distrib. degenerada que concentra toda su información en B_i"

$\vdots$

(B_p) $v_{f_p}^t = (0, \dots, 0, 1)_{p\times 1}$

(Los vértices son distribuciones extremas).

Interpretación:

Como $d_{\chi^2}(r_i, v_{f_i}) = d_e^2(Pr_i, Pv_{f_i})$ podemos interpretar la distancia del perfil fila r_i a la característica B_j.

- PREGUNTA 8

Determina las coordenadas de m_r' en la nueva base.

Veamos que $(0, m_r)$ es un autovalor /autovector de S_r.
Para ello, tenemos que comprobar que se verifica

$$S_r \cdot m_r = 0_p$$

Entonces,

$$S_r \cdot m_r = \sum_{i=1}^{n} f_{i.} (r_i - m_r)(r_i - m_r)^t \underbrace{D_p^{-1} m_r}_{\begin{pmatrix} \frac{1}{f_{\cdot 1}} & & \\ & \ddots & \\ & & \frac{1}{f_{\cdot p}} \end{pmatrix} \begin{pmatrix} f_{\cdot 1} \\ \vdots \\ f_{\cdot p} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_p = \mathbb{1}_p} = \ldots$$

Además,

$$\left.\begin{matrix} r_i^t \cdot \mathbb{1}_p = 1 \\ m_r^t \cdot \mathbb{1}_p = 1 \end{matrix}\right\} \underset{\text{Por tanto,}}{\Longrightarrow} (r_i - m_r)^t D_p^{-1} m_r = 0 \quad (\text{escalar})$$

Luego,

$$\ldots = 0_p .$$

$\Rightarrow$ como $(0, m_r)$ es autovalor / autovector de S_r, m_r es un eje de la nueva base $\Rightarrow$ $m_r' D_p^{-1} u_k = 0$ (para $u_k \neq m_r$)

$\Rightarrow$ $(0, \ldots, 0, 1)$ son las coordenadas de m_r^t.

Las coordenadas de r_i sobre m_r:

$$\underset{\text{(coordenadas de } m_r)}{\begin{bmatrix} r_1^t D_p^{-1} m_r \\ \vdots \\ r_n^t D_p^{-1} m_r \end{bmatrix}} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Cuando proyectamos todos tienen coordenada 1.
Es constante ya que la varianza de dicha columna vale 0.

Para la solución de (2):

Sea $L(u_2, \beta_1, \beta_2) = u_2^t D_p^{-1} S_r u_2 - \beta_1 (u_2^t D_p^{-1} u_2 - 1) - \beta_2 (u_2^t D_p^{-1} u_1)$

Entonces,

$$\frac{\partial L}{\partial u_2} = 2 D_p^{-1} S_r u_2 - 2\beta_1 D_p^{-1} u_2 - \beta_2 D_p^{-1} u_1 = 0_{(p)}$$

$$\Rightarrow 2 u_1^t D_p^{-1} S_r u_2 - 2\beta_1 \underbrace{u_1^t D_p^{-1} u_2}_{\substack{\shortparallel \\ 0}} - \beta_2 \underbrace{u_1^t D_p^{-1} u_1}_{\substack{\shortparallel \\ 1}} = 0_{(1)}$$

Como $u_2^t D_p^{-1} S_r u_1 = \lambda u_2^t D_p^{-1} u_1 = 0 \Rightarrow \underline{\underline{\beta_2 = 0}}$

Por tanto,

$$D_p^{-1} S_r u_2 = \beta_1 D_p^{-1} u_2 \Longrightarrow \boxed{S_r u_2 = \beta_1 u_2}$$

y se tiene que (β_1, u_2) son autovalor / autovector de S_r.

• PREGUNTA 6

Demuestra que el cuadrado de la distancia euclídea entre las proyecciones de dos filas P_{r_i} y $P_{r_{i'}}$ coincide con la distancia chi-cuadrado entre las filas r_i y $r_{i'}$.

$$d_e^2 (P_{r_i}, P_{r_{i'}}) = (P_{r_i} - P_{r_{i'}})^t (P_{r_i} - P_{r_{i'}}) =$$

$$= \sum_{\alpha=1}^{p} (r_i^t D_p^{-1} u_\alpha - r_{i'}^t D_p^{-1} u_\alpha)(r_i^t D_p^{-1} u_\alpha - r_{i'}^t D_p^{-1} u_\alpha)$$

$$= (r_i - r_{i'})^t D_p^{-1} \underbrace{\sum_{\alpha=1}^{p} u_\alpha u_\alpha^t D_p^{-1}}_{I_p} (r_i - r_{i'})$$

$$= (r_i - r_{i'})^t D_p^{-1} (r_i - r_{i'}) = d_{\chi^2}(r_i, r_{i'})$$

* NOTA :

Los nuevos ejes $u_1, \ldots, u_p$ verifican

$$\begin{pmatrix} u_1^t \\ \vdots \\ u_p^t \end{pmatrix} D_p^{-1} (u_1 \ldots u_p) = I_p$$

$$(u_1 \ldots u_p) \begin{pmatrix} u_1^t \\ \vdots \\ u_p^t \end{pmatrix} D_p^{-1} = \sum_{\alpha=1}^{p} u_\alpha u_\alpha^t D_p^{-1} = I_p$$

Las coordenadas de r_i sobre los nuevos ejes son:

$$P_{r_i} : (r_i^t D_p^{-1} u_1, \ldots, r_i^t D_p^{-1} u_p)$$

con

$$r_i = \sum_{\alpha=1}^{p} (r_i^t D_p^{-1} u_\alpha) u_\alpha = \sum_{\alpha=1}^{p} u_\alpha (u_\alpha^t D_p^{-1} r_i) =$$

$$= \underbrace{\sum_{\alpha=1}^{p} u_\alpha u_\alpha^t D_p^{-1}}_{I_p} r_i$$

- PREGUNTA 9

Demuestra que $(0, m_r)$ es un autovalor-vector de la matriz

$$S_r = \sum_{i=1}^{n} f_{i\cdot} (r_i - m_r)(r_i - m_r)^t D_p^{-1}$$

Para ello, tenemos que comprobar que se verifica

$$S_r \cdot m_r = 0_p .$$

Luego,

$$S_r \cdot m_r = \sum_{i=1}^{n} f_{i\cdot} (r_i - m_r)(r_i - m_r)^t \underbrace{D_p^{-1} (m_r)}_{} = \ldots$$

$$\begin{pmatrix} \frac{1}{f_{\cdot 1}} & & \\ & \ddots & \\ & & \frac{1}{f_{\cdot p}} \end{pmatrix} \begin{pmatrix} f_{\cdot 1} \\ \vdots \\ f_{\cdot p} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_p = \mathbb{1}_p$$

Además,

$$\left. \begin{array}{l} r_i^t \cdot \mathbb{1}_p = 1 \\ m_r^t \cdot \mathbb{1}_p = 1 \end{array} \right\} \Rightarrow (r_i - m_r)^t D_p^{-1} m_r = 0$$

Por tanto, $S_r \cdot m_r = 0_p$

(con ello queda demostrado que $(0, m_r)$ es un autovalor-vector de S_r).

• PREGUNTA 10

Determina las coordenadas de las filas del autovector m_r. ¿Qué evidencian?:

$$\begin{bmatrix} r_1^t D_p^{-1} m_r \\ \vdots \\ r_n^t D_p^{-1} m_r \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

, donde se evidencia que la varianza de la columna es 0 (al ser la columna cte.)

• PREGUNTA 11

Demuestra que la varianza ponderada de las coordenadas correspondientes al autovector u_k se puede expresar como:

$$\sum_{i=1}^{n} f_{i\cdot} (r_i^t D_p^{-1} u_k)^2 \quad :$$

$$V_k = u_k^t \cdot D_p^{-1} \cdot S_r \cdot u_k = u_k^t D_p^{-1} \underbrace{\sum_{i=1}^{n} f_{i\cdot} (r_i - m_r)(r_i - m_r)^t D_p^{-1} u_k}_{S_r \text{ (no depende de la base)}} =$$

$$= \sum_{i=1}^{n} f_{i\cdot} (r_i^t D_p^{-1} u_k - \underbrace{m_r^t D_p^{-1} u_k}_{\overset{\shortparallel}{0} \text{ si } u_k \neq m_r})^2 = \sum_{i=1}^{n} f_{i\cdot} (r_i^t D_p^{-1} u_k)^2$$

(Media ponderada $m_r^t D_p^{-1} u_k = 0$ si $u_k \neq m_r$)

- ## PREGUNTA 12

Define la inercia total y demuestra que es igual a la suma de los autovalores de S_r :

La inercia total se define como : $InT = \frac{1}{N} \chi^2$

Veamos que $InT = \frac{1}{N} \chi^2 = tr(S_r) = \lambda_1 + \cdots + \lambda_p$

Tenemos: $\chi^2 = N \cdot \sum_{i=1}^{n} f_{i.} \; d_{\chi^2}(r_i, m_r) = N \cdot \sum_{i=1}^{n} f_{i.} (r_i - m_r)^t D_p^{-1} (r_i - m_r)$

$$S_r = \sum_{i=1}^{n} f_{i.} (r_i - m_r)(r_i - m_r)^t D_p^{-1}$$

$$tr(S_r) = \lambda_1 + \cdots + \lambda_p = \sum_{i=1}^{n} f_{i.} \; tr\left((r_i - m_r)(r_i - m_r)^t D_p^{-1}\right) =$$

$$= \sum_{i=1}^{n} f_{i.} \underbrace{\left((r_i - m_r)^t D_p^{-1} (r_i - m_r)\right)}_{1 \times 1 \text{ (escalar)}} = \sum_{i=1}^{n} f_{i.} \; d_{\chi^2}(r_i, m_r)$$

Entonces, $InT = \frac{1}{N} \chi^2 = \frac{1}{N} \cdot N \cdot \sum_{i=1}^{n} f_{i.} \, d_{\chi^2}(r_i, m_r) = tr(S_r) = \lambda_1 + \cdots + \lambda_p$

4. ANÁLISIS DISCRIMINANTE

• PREGUNTA 1.

Discriminación en dos poblaciones con distribuciones conocidas

(a) Define la probabilidad total de error de clasificación de una regla discriminante:

Sea P una población formada por dos grupos G_1 y G_2 con proporciones π_1 y π_2 respectivamente $(\pi_1 + \pi_2 = 1)$

Sea $\underline{X} = (X_1, \dots, X_p)'$ con función de densidad $f_i(\underline{x})$ en G_i, $i = 1,2$. Una regla discriminante viene determinada por una partición del espacio muestral Ω en dos subconjuntos Ω_1 y Ω_2, donde:

$$\Omega_1 \cup \Omega_2 = \Omega \quad \text{y} \quad \Omega_1 \cap \Omega_2 = \emptyset$$

, de forma que dado un individuo $\underline{x}_0 = (x_{01}, \dots, x_{0p})'$, lo asignamos al grupo G_i si $\underline{x}_0 \in \Omega_i$, $i = 1,2$.

Las probabilidades de los errores de clasificación son:

$$P(2|1) = P(\text{Asignar en } G_2 \text{ siendo de } G_1) = \int_{\Omega_2} f_1(\underline{x})\, dx$$

$$P(1|2) = P(\text{Asignar en } G_1 \text{ siendo de } G_2) = \int_{\Omega_1} f_2(\underline{x})\, dx$$

Se define la probabilidad total del error de clasificación de una regla discriminante como:

$$P(\Omega : f) = P(\text{Asignar en } G_2 \text{ siendo de } G_1 \text{ ó asignar en } G_1 \text{ siendo de } G_2) = P(2|1) \cdot \pi_1 + P(1|2) \cdot \pi_2$$

(b) Determina la regla discriminante que minimiza la probabilidad total de error de clasificación.

$$P(\Omega : f) = P(1|2)\cdot \pi_2 + P(2|1)\cdot \pi_1 = \pi_2 \cdot \int_{\Omega_1} f_2(\underline{x})dx + \pi_1 \int_{\Omega_2} f_1(x)\, dx =$$

$$= \pi_2 \cdot \int_{\Omega_1} f_2(\underline{x})\, dx + \pi_1 \left[1 - \int_{\Omega_1} f_1(\underline{x})\, dx \right] =$$

$$= \int_{\Omega_1} \left(\pi_2 f_2(x) - \pi_1 f_1(\underline{x}) \right) dx + \pi_1$$

Como $\pi_1 > 0$, $P(\Omega : f)$ se minimiza donde lo haga la integral, es decir, el π_1 final no influye en el mínimo.

Luego el recinto donde se minimiza la integral es:

$$\Omega_1^* = \{ \underline{x} \,/\, f_2(\underline{x})\pi_2 - f_1(\underline{x})\pi_1 \leq 0 \} = \{ \underline{x} \,/\, f_1(\underline{x})\pi_1 \geq f_2(\underline{x})\pi_2 \}$$

Así, clasificaremos un individuo en G_1 si: $\pi_1 f_1(\underline{x}_0) > f_2(\underline{x}_0)\pi_2$, y en G_2 si $\pi_2 f_2(\underline{x}_0) > \pi_1 f_1(\underline{x}_0)$.

(c) Plantea y resuelve el problema original de Fisher:

Sea P una población formada por dos grupos G_1 y G_2. Sea $\underline{X} = (\underline{X}_1, \dots, \underline{X}_p)'$ con distribución $(\underline{\mu}_1, \Sigma)$ en G_1 y $(\underline{\mu}_2, \Sigma)$ en G_2.

Se plantea el problema de determinar $b \in \mathbb{R}^p$ tal que:

$$\begin{aligned} &\max_{\underline{b}} \ (\underline{b}'\underline{\mu}_1 - \underline{b}'\underline{\mu}_2)^2 \\ &s.a.: \ \underline{b}'\hat{\Sigma}\underline{b} = 1 \end{aligned}$$

, donde $\underline{b}$ es el vector que más diferencia entre los grupos.

Reescribimos el problema como:

$$\begin{aligned} &\max_{\underline{b}} \ \underline{b}' \overbrace{(\underline{\mu}_1 - \underline{\mu}_2)}^{d} \overbrace{(\underline{\mu}_1 - \underline{\mu}_2)'}^{d'} \underline{b} \\ &s.a.: \ b'\hat{\Sigma} b = 1 \end{aligned}$$

, donde la solución a este tipo de problema ya sabemos que viene dada por el autovector asociado al máximo autovalor de $\hat{\Sigma}^{-1}\underline{d}\,\underline{d}^t$, que esta matriz (en particular) solo posee un autovalor no nulo, el cual es:

$\underline{b} = \hat{\Sigma}^{-1} \cdot d$, pero no está normalizado, luego no cumple que $\underline{b}'\hat{\Sigma}\underline{b} = 1$,

por tanto el autovector normalizado que necesitamos es, llamando

$a = d^t \hat{\Sigma}^{-1} d$, al autovalor asociado a $\underline{b}$, obtenemos:

$$\underline{b}^* = \frac{c}{a^{1/2}} = \frac{\hat{\Sigma}^{-1} \cdot d}{(d^t \hat{\Sigma}^{-1} d)^{1/2}}$$

(d) Demuestra que supuesto que las poblaciones se distribuyen según $N(\mu_i, \Sigma)$, $i=1,2$, la regla del discriminante que minimiza la probabilidad total de error coincide con la propuesta por Fisher.

- Regla propuesta por Fisher:

Asignar $\underline{x}_0$ a G_1 si $|\underline{b}'\underline{x}_0 - \underline{b}'\underline{\mu}_1| < |\underline{b}'\underline{x}_0 - \underline{b}'\underline{\mu}_2| \Leftrightarrow$

$$\Leftrightarrow (\underbrace{\underline{b}'\underline{x}_0}_{y_0} - \underline{b}'\underline{\mu}_1)^2 < (\underbrace{\underline{b}'\underline{x}_0}_{y_0} - \underline{b}'\underline{\mu}_2)^2 \Leftrightarrow$$

$$\Leftrightarrow \cancel{y_0^2} + (\underline{b}'\underline{\mu}_1)^2 - 2y_0\underline{b}'\underline{\mu}_1 < \cancel{y_0^2} + (\underline{b}'\underline{\mu}_2) - 2y_0\underline{b}'\underline{\mu}_2 \Leftrightarrow$$

$$\Leftrightarrow 2y_0 \underbrace{(\underline{b}'\underline{\mu}_2 - \underline{b}'\underline{\mu}_1)}_{<0} < \underbrace{(\underline{b}'\underline{\mu}_2)^2 - (\underline{b}'\underline{\mu}_1)^2}_{(\underline{b}'\underline{\mu}_2 + \underline{b}'\underline{\mu}_1)(\underline{b}'\underline{\mu}_2 - \underline{b}'\underline{\mu}_1)} \Leftrightarrow$$

$$\Leftrightarrow 2y_0 > \underline{b}'\underline{\mu}_1 + \underline{b}'\underline{\mu}_2 \Leftrightarrow \boxed{y_0 > \frac{\underline{b}'(\underline{\mu}_1 + \underline{\mu}_2)}{2}}$$

- Regla que minimiza la probab. del error:

Asignamos $\underline{x}_0$ a G_1 si $\pi_1 f_1(\underline{x}_0) > \pi_2 f_2(\underline{x}_0)$. Para el caso particular $f_i \sim N_p(\mu_i, \Sigma)$, $i=1,2$, si $\pi_1 = \pi_2$, asigno $\underline{x}_0$ en G_1 si $f_1(\underline{x}_0) > f_2(\underline{x}_0)$ (lo asignamos donde haya mas densidad).

Luego asigno $\underline{x}_0$ en G_1 si:

Clasifico $\underline{x}_0$ en G_2	Clasifico $\underline{x}_0$ en G_1
$\underline{b}'\underline{\mu}_2$	$\underline{b}'\underline{\mu}_1$

$$\frac{\underline{b}'(\mu_1+\mu_2)}{2}$$

, luego asigno $\underline{x}_0$ en G_1 si:

$$\boxed{\underline{b}'\underline{x}_0 > \frac{\underline{b}'(\underline{\mu}_1 + \underline{\mu}_2)}{2}}$$

- **PREGUNTA 2**

Coordenadas discriminantes canónicas.

(a) Planteamiento y solución:

Sea P una población formada por k grupos $G_1, \dots, G_k$ con proporciones $\pi_1, \dots, \pi_k$ respect. $\left(\sum_{i=1}^{k} \pi_i = 1\right)$. Sea $\underline{X} = (X_1, \dots, X_p)^t$ con fdd $f_i(\underline{x})$ en G_i, $i = 1, \dots, k$. Consideramos una muestra aleatoria de cada grupo $x_{i1}, \dots, x_{in_i}$ de G_i, $i = 1, \dots, k$.

El objetivo es determinar las comb. lineales de las vbles. $\underline{Y} = \underline{b}^t \underline{X}$, $\underline{b} \in \mathbb{R}^p$ que mas discrimina entre los grupos. Para ello consideramos el contraste:

$$H_0 : \underline{b}' \mu_1 = \dots = \underline{b}' \mu_k$$, bajo normalidad y homocedasticidad supuestas. El estadístico es:

$$F = \alpha \frac{\underline{b}' B \underline{b}}{\underline{b}' W \underline{b}}, \quad \text{donde} \quad B = \sum_{i=1}^{k} n_i (\bar{\underline{X}}_i - \bar{\bar{\underline{X}}})(\bar{\underline{X}}_i - \bar{\bar{\underline{X}}})^t, \ y \quad W = \sum_{i=1}^{k} S_i$$

y la región crítica $F \geq F_c$ (F crítica).

Así, debemos tomar $\underline{b}$ de forma que el estadístico sea máximo para que las medias sean lo mas distintas posibles. Planteamos los siguientes problemas de optimización:

(1) $$\sup_{b \neq 0} \frac{\underline{b}' B \underline{b}}{\underline{b}' W \underline{b}} \equiv \sup_{\text{s.a.: } \underline{b}' W \underline{b} = 1} \underline{b}' B \underline{b}$$, donde la solución

es el autovector $\underline{e}_1$ asociado al máximo autovalor de $W^{-1}B$, y definimos la primera coordenada discriminante como $\underline{Y}_1 = \underline{e}_1' \underline{X}$

(2) $\sup_{\underline{b} \neq 0} \dfrac{\underline{b}' B \underline{b}}{\underline{b}' W \underline{b}} \equiv \sup \underline{b}' B \underline{b}$ s.a.: $\underline{b}' W \underline{b} = 1$, $\underline{e}_1' W \underline{b} = 0$, cuya solución $\underline{e}_2$

es el autovector asociado al segundo mayor autovalor de $W^{-1}B$, luego definimos la segunda coord. discriminante

como: $\underline{Y}_2 = \underline{e}_2' \underline{X}$

$\vdots$

(Así con los r autovalores no nulos de $W^{-1}B$).

(b) Describe el criterio de clasificación basado en las coord. canónicas:

Determinados $\underline{e}_1, \ldots, \underline{e}_r \in \mathbb{R}^p$, usamos el criterio de clasificación para asignar un individuo a su grupo correspondiente. $\underline{x}_0$ será clasificado en G_i si:

$$d(C_r \cdot \underline{x}_0, C_r \cdot \bar{\underline{X}}_i) = \min_{s=1,\ldots,k} d(C_r \underline{x}_0, C_r \bar{\underline{X}}_s) \quad , \text{ donde}$$

$C_r = (\underline{e}_1, \ldots, \underline{e}_r)^t$