# Tema 6

## Análisis de correspondencia por filas:

Objetivo: Determinar una base $u_1, \dots, u_p$ de $\mathbb{R}^p$ tal que:

a) $d_{\chi^2}(r_i, r_j) = d_e^2(P_{r_i}, P_{r_j})$

b) $V_1 \geq V_2 \geq \dots \geq V_p$

$$\begin{bmatrix} r_1' \\ \vdots \\ r_n' \end{bmatrix} \longrightarrow \Big[$$

Determinación de la base: Supongamos que tengo una base $u_1, \dots, u_p$ de $\mathbb{R}^p$ con a) $u_k^t D_p^{-1} u_k = 1 \ \forall k$ b) $u_k^t D_p^{-1} u_\ell = 0 \quad k \neq \ell$

Se deduce:

1. $\begin{bmatrix} u_1^t \\ \vdots \\ u_n^t \end{bmatrix} D_p^{-1} [u_1 \dots u_p] = I_p \qquad [u_1 \dots u_p] \begin{bmatrix} u_1^t \\ \vdots \\ u_p^t \end{bmatrix} D_p^{-1} = I_p$

   $\sum_{k=1}^{p} u_k u_k^t D_p^{-1} = I_p$

2. Coordenadas de $r_i$ respecto de $u_1 \dots u_p$. $P_{r_i}^t = (r_i^t D_p^{-1} u_1, \dots, r_i^t D_p^{-1} u_p)$

   Verifica que:

   $$\sum_{k=1}^{p} \underbrace{(r_i^t D_p^{-1} u_k)}_{\text{Escalar}} u_k = \sum_{k=1}^{p} u_k (r_i^t D_p^{-1} u_k) \overset{\substack{\text{escalar =} \\ \text{a su trasp}}}{=} \sum_{k=1}^{p} \underbrace{u_k \cdot u_k^t D_p^{-1}}_{I_p} r_i = r_i$$

   Podemos expresar $P_{r_i}^t = r_i^t D_p^{-1} (u_1 - u_p)$

3. $\delta_e^2(P_{r_i}, P_{r_j}) = (P_{r_i} - P_{r_j})^t (P_{r_i} - P_{r_j}) \overset{\downarrow}{=} (r_i^t - r_j^t) D_p^{-1} \underbrace{(u_1 \dots u_p) \begin{pmatrix} u_1^t \\ \vdots \\ u_n^t \end{pmatrix} D_p^{-1}}_{u \ I_p} (r_i - r_j) =$

   $= (r_i - r_j)^t D_p^{-1} (r_i - r_j) = \delta_{\chi^2}(r_i, r_j)$

   $\overset{\shortparallel}{\delta_{\chi^2}(r_i, r_j)} \overset{\shortparallel}{\ } (r_i - r_j)^t D_p^{-1} (r_i - r_j)$

4. $$\begin{bmatrix} P_{r_1}' \\ \vdots \\ P_{r_n}' \end{bmatrix} = \begin{bmatrix} \cdots & \overset{\text{coord asociadas a } u_k}{r_1^t D_p^{-1} u_k} & \cdots \\ & \vdots & \\ & \underbrace{r_n^t D_p^{-1} u_k}_{\text{columna } k} & \end{bmatrix} \underbrace{\begin{pmatrix} f_{1\cdot} \\ \vdots \\ f_{n\cdot} \end{pmatrix}}_{\text{Ponderac. (pesos)}}$$

Media ponderada de la columna $k$: $\sum_{i=1} r_k^t D_p^{-1} u_k f_{i.} = \sum_{i=1} f_{i.} r_k^t D_p^{-1} u_k = m_r^t D_p^{-1} u_k$

5) Varianza ponderada de la columna $k$: $V_k = \sum_{i=1}^{n} (r_k^t D_p^{-1} u_k - m_r^t D_p^{-1} u_k)^2 f_{i.} =$

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

6) $$\sum_{k=1}^{p} V_k = \sum_{k=1}^{p} \underbrace{u_k^t D_p^{-1} S_r u_k}_{\text{Escalar}} = \sum tr(u_k^t D_p^{-1} S_r u_k) = tr(\underbrace{\sum u_k u_k^t D_p^{-1}}_{Id} S_r) =$$

$= tr(S_r)$

↑ No depende de la base

7) Problema a resolver: $u_1$ tal que $\max_{u_1} u_1^t D_p^{-1} S_r u_1$ s.a $u_1^t D_p^{-1} u_1 = 1$

y $u_2$ tal que $\max_{u_2} u_2^t D_p^{-1} S_r u_2$ s.a $u_2^t D_p^{-1} u_2 = 1$

$u_2^t D_p^{-1} u_1 = 0$

<u>Teorema</u>: Sea $A$ una matriz simétrica, sea $M$ una matriz simétrica def. pos. $\Rightarrow$

(a) $\max_{u_1} u_1^t A u_1$ s.a $u_1^t M u_1 = 0$ $\Rightarrow$ da solución $(\lambda_{MAX}, u_{MAX})$ de $M^{-1}A$ (Autovect, autovalor)

(b) $\max_{u_2} u_2^t A u_2$ s.a $u_2^t M u_2 = 1$, $u_2^t M u_1 = 0$ $\Rightarrow$ da solución $(\lambda_{2^o}, u_{2^o})$ de $M^{-1}A$

<u>Dem</u>:

(a) $L(u_1, \lambda) = u_1^t A u_1 - \lambda(u_1^t M u_1 - 1)$

$$\frac{\delta L}{\delta u_1} = 2Au_1 - 2\lambda M u_1 = \underset{\uparrow \text{ vector dim } p}{O_{(p)}} \Rightarrow \underbrace{Au_1 = \lambda M u_1}_{} \Rightarrow M^{-1} A u_1 = \lambda u_1$$

$\Rightarrow (\lambda_1, u_1)$ aut de $M^{-1}A$

Multiplicamos por $u_1^t$: $\boxed{u_1^t A u_1} = \lambda \underbrace{u_1^t M u_1}_{1}$

↑ Como es lo que queremos maximizar entonces la soluc. viene dada por $(\lambda_{max}, u_{max})$ de $M^{-1}A$

Primera y seg rest.

(b) $L(u_2, \beta_1, \beta_2) = u_2^t A u_2 - \beta_1 (u_2^t M u_2 - 1) - \beta_2 (u_2^t M u_1)$

$$\frac{\delta L}{\delta u_2} = 2Au_2 - 2\beta_2 M u_2 - \beta_2 M u_1 = 0_{(p)}^{(*)}$$ Mult $u_1^t$

$$2\underset{\overset{\shortparallel}{0}}{u_1^t A u_2} - 2\beta_1 \underbrace{u_1^t M u_2}_{C \text{ por hipot.}} - \underbrace{\beta_2 u_1^t M u_1}_{1 \text{ por (a)}} = 0_{(1)}$$

De ... ; $u_2^t A u_1 = \lambda u_2^t M u_1 = 0$ , por tanto de sacamos

mult $u_2^t$ $\qquad \overset{\shortparallel}{u_1^t M u_2} \qquad \overset{\shortparallel}{0}$

la condición de que $\beta = 0$

(*) Tenemos la solución $Au_2 = \beta_1 M u_2 \Rightarrow M^{-1} A u_2 = \beta_1 u_2$

$u_2^t A u_2 = \beta_1$ $\qquad (\beta_1, u_2)$ autv/actov de $M^{-1}A$

(Es lo que queremos Max)

Solución de (7) : $u_1, \dots, u_p$ son autovect de $S_r$

$\lambda_1 \geq \dots \geq \lambda_p$ son autovalores de $S_r$

y $V_k = \lambda_k \quad k = 1, \dots, p$

(8) El nº de autovalores no nulos de la matriz $S_r$ es :

$\min(n-1, p-1)$

Veamos primero que $S_r \cdot m_r = 0 \cdot m_r = 0_p$

$$\sum_{i=1}^{n} f_{i\cdot} (r_i - m_r) \underbrace{(r_i - m_r)^t \underbrace{D_p^{-1} m_r}}_{\overset{\shortparallel}{C}} = 0_p$$

$r_i^t 1_p = 1$

$m_r^t 1_p = 1$

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

(9) Coordenadas de $r_i$ sobre $m_r$

$$\begin{bmatrix} r_1^t D_p^{-1} m_r \\ \vdots \\ r_n^t D_p^{-1} m_r \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$ ← Tiene varianza 0 (columna cte)

(10) Coordenadas de $r_i$ sobre $u_k$ ($\neq m_r$)

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

Media ponderada : $m_r^t D_p^{-1} u_k = 0$ por $u_k^t D_p^{-1} u_k = 0 \ (k \neq l)$

$\Rightarrow$ las media ponderada de todas las columnas es cero

## MEDIDAS PARA LA INTERPRETACIÓN

- $\chi^2 = N \sum_{i=1}^{n} f_{i.} \underbrace{\delta_{\chi^2}(r_i, m_r)}_{(r_i - m_r)^t D_p^{-1} (r_i - m_r)}$
- $tr(S_r) = \lambda_1 + \ldots + \lambda_p$
- $tr(S_r) = \sum_{i=1}^{n} f_{i.} \, tr\left( (r_i - m_r)(r_i - m_r)^t D_p^{-1} \right) = \sum_{i=1}^{n} f_{i.} \overbrace{\left( (r_i - m_r)^t D_p^{-1} (r_i - m_r) \right)}^{\substack{\delta_{\chi^2}(r_i, m_r) \\ \shortparallel \\ 1 \times 1}} =$

  $= \sum_{i=1}^{n} f_{i.} \, d_{\chi^2}(r_i, m_r)$

- Inercia total: $\frac{\chi^2}{N} = tr(S_r) = \lambda_1 + \ldots + \lambda_p$
- Inercia debida a $r_i$: $f_{i.} \, d_{\chi^2}(r_i, m_r)$ (%)
- Inercia debida a $u_k$ (eje $k$): $\lambda_k$ (%)

Contribución de $r_i$ a la inercia $u_k$: $\frac{f_{i.} (r_i^t D_p^{-1} u_k)^2}{\lambda_k}$

Contribución de $u_k$ a la inercia $r_i$:

tenemos que ver qué aporta $u_k$ a $f_{i.} \, d_{\chi^2}(r_i, m_r)$:

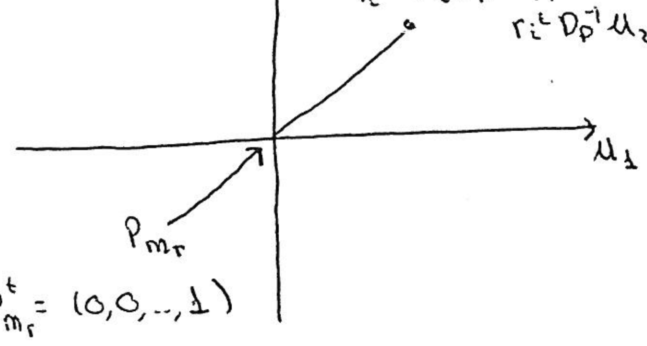
$d_{\chi^2}(r_i, m_r) = d_e^2 (P_{r_i}, P_{m_r}) = \sum_{\substack{u_k \\ u_k \neq m_r}} (r_i^t D_p^{-1} u_k)^2$

$(r_i^t D_p^{-1} u_1, \ldots, \underbrace{r_i^t D_p^{-1} u_p}_{=1})$

$(m_r^t D_p^{-1} u_1, \ldots, m_r^t D_p^{-1} u_p)$, con $m_r^t D_p^{-1} u_1 = 0$, $\ldots = 0$, $m_r^t D_p^{-1} u_p = 1$

porque $m_r$ es un elemento de la base (ortog a $u_i$)

Por tanto, la contribución será: $\frac{r_i^t D_p^{-1} u_k}{\delta_{\chi^2}(r_i, m_r) = \delta_e^2 (P_{r_i}, P_{m_r})}$ (1)




Si (1) es practicamente 1, / es una buena aproximación de la distancia $\chi^2$ o distancia euclidea entre $P_{r_i}$ y $P_{m_r}$. Si no es cercana a 0; solo las dos primeras coord no sirven.

nº autov no nulos =
= min (n-1, p-1)

## PUNTOS SUPLEMENTARIOS

$r_+^t = (r_{+1}, \dots, r_{+p})$ $\nearrow \geq 0$ $\searrow \Sigma = 1$

símbolo suma

Nuevas coordenadas: de $r_+^t \to (\dots r_+^t D_p^{-1} u_K \dots)$

de $m_r^t \to (\dots m_r^t D_p^{-1} u_K \dots) = (0, 0, \dots, 0, \pm 1)$

- Vértices de las filas: (B1) $V_{f_1}^t = (1, 0, \dots, 0)_{p \times 1}$

⋮

(Bp) $V_{f_p}^t = (0, 0, \dots 1)_{p \times 1}$ <- concentra toda la prob en la última categoría

$(\dots V_{f_i}^t D_p^{-1} u_K \dots)$

- Ejemplo: tabla de contingencia