



**Definición.** Sea  $\mathbf{A} \in \mathbb{R}^{m \times m}$  de rango  $r$ , se define la inversa generalizada de  $\mathbf{A}$  denotada por  $\mathbf{A}^-$  como

$$\mathbf{A} \mathbf{A}^- \mathbf{A} = \mathbf{A}$$

**Definición.** Sea  $\mathbf{A} \in \mathbb{R}^{n \times p}$  se define la norma de Frobenius de  $\mathbf{A}$  como

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^n \sum_{j=1}^p a_{ij}^2} = \sqrt{\text{TR}(\mathbf{A}^T \mathbf{A})} = \sqrt{\sum_{i=1}^{\min\{n, p\}} \delta_i^2}$$

con  $\delta_i^2$  los autovalores no nulos de  $\mathbf{A}^T \mathbf{A}$ .

## Descomposicion de valores singulares

**Descomposición 1.** Dada  $\mathbf{A} \in \mathbb{R}^{n \times p}$ ,  $n \geq p$ , de rango  $r$  puede expresarse como

$$\mathbf{A} = \mathbf{U}\mathbf{\Delta}\mathbf{W}^T$$

- $\mathbf{U} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{W} \in \mathbb{R}^{p \times p}$  ortogonales, o sea,  
 $\mathbf{U}\mathbf{U}^T = \mathbf{I}_n$ ,  
 $\mathbf{W}\mathbf{W}^T = \mathbf{I}_p$ ,
- $\mathbf{\Delta} \in \mathbb{R}^{n \times p}$  una matriz formada por los *Valores Singulares* de  $\mathbf{A}$  (o sea, la raíz cuadrada de los autovalores de  $\mathbf{A}^T\mathbf{A}$ ) en su diagonal principal ordenados de mayor a menor.

Sean  $\lambda_1 \geq \dots \geq \lambda_r > 0$  los autovalores no nulos de  $\mathbf{A}^T \mathbf{A} \in \mathbb{R}^{p \times p}$ ,  $r \leq \min(p, n)$  y sean  $\delta_j = \sqrt{\lambda_j}$ . Luego, si  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_p)$ ,  $\mathbf{w}_j \in \mathbb{R}^p$ ,  $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_n)$ ,  $\mathbf{u}_j \in \mathbb{R}^n$ ,

$$\mathbf{A} = \sum_{i=1}^r \delta_i \mathbf{u}_i \mathbf{w}_i^T$$

Los vectores  $\{\mathbf{w}_1, \dots, \mathbf{w}_p\}$  son los autovectores de  $\mathbf{A}^T \mathbf{A}$ .

Como  $\{\mathbf{A}\mathbf{w}_1, \dots, \mathbf{A}\mathbf{w}_p\}$  son ortogonales y  $\|\mathbf{A}\mathbf{w}_i\| = \sqrt{\lambda_i} = \delta_i$ , podemos definir

$$\mathbf{u}_j = \frac{\mathbf{A}\mathbf{w}_j}{\delta_j} \quad 1 \leq j \leq r$$

y completarla a una base ortonormal de  $\mathbb{R}^n$ .

Es fácil ver que con esta elección  $\mathbf{A}\mathbf{W} = \mathbf{U}\mathbf{\Delta}$

## Descomposición 2

Dada  $\mathbf{A} \in \mathbb{R}^{n \times p}$ ,  $n \geq p$ , de rango  $r$  puede expresarse como

$$\mathbf{A} = \mathbf{L}\mathbf{\Delta}\mathbf{W}^T$$

con  $\mathbf{L} \in \mathbb{R}^{n \times p}$ ,  $\mathbf{\Delta} \in \mathbb{R}^{p \times p}$ ,  $\mathbf{W} \in \mathbb{R}^{p \times p}$ ,

- $\mathbf{L}^T \mathbf{L} = \mathbf{I}_p$ ,
- $\mathbf{W}^T \mathbf{W} = \mathbf{I}_p$ ,
- $\mathbf{\Delta} = \text{DIAG}(\delta_1, \dots, \delta_p)$ ,  $\delta_1 \geq \dots \geq \delta_r \geq \delta_{r+1} = \dots = \delta_p = 0$ .

Las  $p$  columnas de  $\mathbf{L} = (\ell_1, \dots, \ell_p)$  son ortonormales y son los autovectores de  $\mathbf{A}\mathbf{A}^T$  asociados a los  $p$  mayores autovalores.

Basta tomar  $\mathbf{L} = (\mathbf{u}_1, \dots, \mathbf{u}_p) = \mathbf{U}_p$  en la definición 1

Las  $p$  columnas de  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_p)$  son los autovectores de  $\mathbf{A}^T \mathbf{A}$  asociados a los  $p$  mayores autovalores.

## Descomposición 2

Sean

$$\mathbf{L}_r = (\ell_1, \dots, \ell_r)$$

$$\mathbf{W}_r = (\mathbf{w}_1, \dots, \mathbf{w}_r)$$

$$\mathbf{\Delta}_r = \text{DIAG}(\delta_1, \dots, \delta_r)$$

entonces

$$\mathbf{A}^T \mathbf{A} = \mathbf{W} \mathbf{\Delta}^2 \mathbf{W}^T \quad \mathbf{A}^T \mathbf{A} \mathbf{w}_j = \delta_j^2 \mathbf{w}_j$$

$$\mathbf{A} \mathbf{A}^T = \mathbf{L} \mathbf{\Delta}^2 \mathbf{L}^T \quad \mathbf{A} \mathbf{A}^T \ell_j = \delta_j^2 \ell_j$$

$$\mathbf{A} = \mathbf{L}_r \mathbf{\Delta}_r \mathbf{W}_r^T = \sum_{j=1}^r \delta_j \ell_j \mathbf{w}_j^T$$

## Aproximación en norma Frobenius

Sea  $\mathbf{A} \in \mathbb{R}^{n \times p}$ ,  $n \geq p$ , de rango  $r$  con descomposición de valores singulares

$$\mathbf{A} = \sum_{j=1}^r \delta_j \ell_j \mathbf{w}_j^T$$

Sea  $\mathcal{C} = \{\mathbf{C} \in \mathbb{R}^{n \times p}, \text{ de rango } s < r\}$ , luego

$$\min_{\mathbf{C} \in \mathcal{C}} \|\mathbf{A} - \mathbf{C}\|_F^2 = \min_{\mathbf{C} \in \mathcal{C}} \sum_{i=1}^n \sum_{j=1}^r (a_{ij} - c_{ij})^2 = \sum_{j=s+1}^r \delta_j^2 = \|\mathbf{A} - \mathbf{A}_{(s)}\|_F^2$$

donde

$$\mathbf{A}_{(s)} = \sum_{j=1}^s \delta_j \ell_j \mathbf{w}_j^T$$

## Aproximación en norma Frobenius

La bondad del ajuste dado por  $\mathbf{A}_{(s)}$  se mide por

$$\rho_{(s)}^2 = 1 - \frac{\|\mathbf{A} - \mathbf{A}_{(s)}\|_F^2}{\|\mathbf{A}\|_F^2} = \frac{\sum_{j=1}^s \delta_j^2}{\sum_{j=1}^r \delta_j^2}$$

**Lema.** Si las filas de  $\mathbf{A}$  suman 0 ( $\mathbf{A}^T \mathbf{1}_n = \mathbf{0}$ ), entonces,  $\mathbf{A}_{(s)}$  es la matriz de rango  $s$  tal que la diferencia entre sus filas mejor aproximan las diferencias entre las filas de  $\mathbf{A}$ , es decir, si  $\mathcal{C}_0 = \{\mathbf{C} \in \mathbb{R}^{n \times p}, \mathbf{C}^T \mathbf{1}_n = \mathbf{0}\}$

$$\operatorname{argmin}_{\mathbf{C} \in \mathcal{C}_0} \sum_{j=1}^n \sum_{k=1}^n \|\mathbf{a}_j - \mathbf{a}_k - (\mathbf{c}_j - \mathbf{c}_k)\|^2 = \mathbf{A}_{(s)}$$

donde  $\mathbf{a}_1, \dots, \mathbf{a}_n$  son las filas de  $\mathbf{A}$  y  $\mathbf{c}_1, \dots, \mathbf{c}_n$  las de  $\mathbf{C}$ .

## Formas bilineales

Sea  $\mathbf{A} \in \mathbb{R}^{n \times p}$ ,  $n \geq p$ , de rango  $r$  con descomposición de valores singulares

$$\mathbf{A} = \sum_{j=1}^r \delta_j \ell_j \mathbf{w}_j^T = \sum_{j=1}^r \delta_j \mathbf{u}_j \mathbf{w}_j^T$$

Tenemos que

$$(\mathbf{u}_1, \mathbf{w}_1) = \underset{\|\mathbf{u}\|=1, \|\mathbf{w}\|=1}{\operatorname{argmax}} \mathbf{u}^T \mathbf{A} \mathbf{w} \quad , \quad \mathbf{u}_1^T \mathbf{A} \mathbf{w}_1 = \delta_1$$

$$(\mathbf{u}_2, \mathbf{w}_2) = \underset{\substack{\|\mathbf{u}\|=1, \|\mathbf{w}\|=1 \\ \mathbf{u}^T \mathbf{u}_1=0, \mathbf{w}^T \mathbf{w}_1=0}}{\operatorname{argmax}} \mathbf{u}^T \mathbf{A} \mathbf{w} \quad , \quad \mathbf{u}_2^T \mathbf{A} \mathbf{w}_2 = \delta_2$$

$$\vdots \quad \quad \quad \vdots$$

$$(\mathbf{u}_j, \mathbf{w}_j) = \underset{\substack{\|\mathbf{u}\|=1, \|\mathbf{w}\|=1 \\ \mathbf{u}^T \mathbf{u}_\ell=0, \mathbf{w}^T \mathbf{w}_\ell=0, 1 \leq \ell \leq j}}{\operatorname{argmax}} \mathbf{u}^T \mathbf{A} \mathbf{w} \quad , \quad \mathbf{u}_j^T \mathbf{A} \mathbf{w}_j = \delta_j$$

## Factorización de Matrices

Sea  $\mathbf{A} \in \mathbb{R}^{n \times p}$ ,  $n \geq p$ , de rango  $r$  entonces,  $\mathbf{A}$  puede factorizarse

$$\mathbf{A} = \mathbf{G}\mathbf{H}^T$$

- $\mathbf{G} \in \mathbb{R}^{n \times r}$ ,  $\text{rango}(\mathbf{G}) = r$ ,

$$\mathbf{G} = \begin{pmatrix} \mathbf{g}_1^T \\ \vdots \\ \mathbf{g}_n^T \end{pmatrix} \quad \mathbf{g}_i \in \mathbb{R}^r$$

- $\mathbf{H} \in \mathbb{R}^{p \times r}$ ,  $\text{rango}(\mathbf{H}) = r$

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}_1^T \\ \vdots \\ \mathbf{h}_p^T \end{pmatrix} \quad \mathbf{h}_j \in \mathbb{R}^r$$

## Factorización de Matrices

$$a_{ij} = \mathbf{g}_i^T \mathbf{h}_j$$

Esta factorización asigna

- cada  $\mathbf{g}_i$  a una fila de  $\mathbf{A}$  y
- cada  $\mathbf{h}_j$  a una columna de  $\mathbf{A}$

Por lo tanto, si  $r = 2$  podríamos plotear los  $n + p$  vectores  $\{\mathbf{g}_i\}_{1 \leq i \leq n} \cup \{\mathbf{h}_j\}_{1 \leq j \leq p}$  de dimensión 2 en lugar de las  $n$  filas de dimensión  $p$  de la matriz  $\mathbf{A}$ .

La descomposición anterior no es única.

## Existencia de una Factorización

Dada  $\mathbf{A} \in \mathbb{R}^{n \times p}$ ,  $n \geq p$ , de rango  $r$  por la descomposición de valores singulares

$$\mathbf{A} = \mathbf{L}_r \mathbf{\Delta}_r \mathbf{W}_r^T = \sum_{j=1}^r \delta_j \ell_j \mathbf{w}_j^T$$

donde

$$\mathbf{A}^T \mathbf{A} \mathbf{w}_j = \delta_j^2 \mathbf{w}_j \quad \mathbf{A}^T \mathbf{A} \mathbf{W}_r = \mathbf{W}_r \mathbf{\Delta}_r^2$$

- $\mathbf{L}_r^T \mathbf{L}_r = \mathbf{I}_r$ ,  $\mathbf{L}_r = (\ell_1, \dots, \ell_r)$  son los autovectores de  $\mathbf{A} \mathbf{A}^T$  asociados a los  $r$  autovalores no nulos ( $\ell_j = \mathbf{A} \mathbf{w}_j / \delta_j$ )
- $\mathbf{W}_r^T \mathbf{W}_r = \mathbf{I}_r$ ,  $\mathbf{W}_r = (\mathbf{w}_1, \dots, \mathbf{w}_r)$  son los autovectores de  $\mathbf{A}^T \mathbf{A}$  asociados a los  $r$  autovalores no nulos
- $\mathbf{\Delta}_r = \text{DIAG}(\delta_1, \dots, \delta_r)$ ,  $\delta_1 \geq \dots \geq \delta_r > 0$ .

## Existencia de una Factorización $\mathbf{A} = \mathbf{GH}^T$

Tomando

$$\mathbf{G} = \mathbf{L}_r \quad \mathbf{H} = \mathbf{W}_r \mathbf{\Delta}_r$$

tenemos que  $\mathbf{A} = \mathbf{GH}^T$  donde  $\mathbf{G}$  y  $\mathbf{H}$  cumplen

a)  $\mathbf{HH}^T = \mathbf{A}^T \mathbf{A}$

b)  $\mathbf{GG}^T = \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T$

La factorización  $\mathbf{A} = \mathbf{GH}^T$  donde  $\mathbf{G}$  y  $\mathbf{H}$  cumplen a) y b) es única salvo transformaciones ortogonales.

Está claro que también podríamos haber elegido

$$\mathbf{G} = \mathbf{L}_r \mathbf{\Delta}_r^{1-\alpha} \quad \mathbf{H} = \mathbf{W}_r \mathbf{\Delta}_r^\alpha$$

La elección anterior corresponde a  $\alpha = 1$  y es la que usaremos para el biplot.

## Biplot

Introducido por Gabriel (1971) un biplot es un gráfico, generalmente bidimensional ( $k = 2$ ), capaz de representar con un cierto grado de aproximacion tanto a las observaciones  $\mathbf{x}_i \in \mathbb{R}^p$ ,  $p > 2$ , como a las  $p$  variables. Las observaciones se representan mediante puntos en el gráfico, mientras que las variables se representan mediante flechas.

Sea  $\mathbf{x} \in \mathbb{R}^p$  tal que  $\mathbb{E}(\mathbf{x}) = \mathbf{0}$ ,  $\text{VAR}(\mathbf{x}) = \mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T$

Sean

- $\lambda_1 \geq \dots \geq \lambda_p$  los autovalores de  $\mathbf{\Sigma}$
- $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p$  los autovectores de  $\mathbf{\Sigma}$  asociados a  $\lambda_1 \geq \dots \geq \lambda_p$
- $\mathbf{\Gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)$ ,  $\mathbf{\Gamma}\mathbf{\Gamma}^T = \mathbf{I}_p$
- $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$

$$\mathbf{x} = \boldsymbol{\mu} + \sum_{j=1}^p \boldsymbol{\gamma}_j^T (\mathbf{x} - \boldsymbol{\mu}) \boldsymbol{\gamma}_j$$

## Representación de las $n$ observaciones

- $v_j = \gamma_j^T(\mathbf{x} - \boldsymbol{\mu})$ ,
- $\mathbf{v}^{(r)} = (v_1, \dots, v_r)^T = \boldsymbol{\Gamma}_r^T(\mathbf{x} - \boldsymbol{\mu})$  son las primeras  $r$  componentes principales

Si  $\mathcal{H}_r$  el subespacio generado por  $\gamma_1, \dots, \gamma_r$

$$d_p(\mathbf{x}_1, \mathbf{x}_2) \approx d_p(\pi(\mathbf{x}_1 - \boldsymbol{\mu}, \mathcal{H}_r), \pi(\mathbf{x}_2 - \boldsymbol{\mu}, \mathcal{H}_r)) = d_r(\mathbf{v}_1^{(r)}, \mathbf{v}_2^{(r)})$$

## Representación de las $n$ observaciones

Dada una muestra  $\mathbf{x}_1, \dots, \mathbf{x}_n$  y  $\mathbf{S} = \mathbf{Q}/(n-1) = \widehat{\mathbf{\Gamma}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Gamma}}^T$  donde

- $\widehat{\lambda}_1 \geq \dots \geq \widehat{\lambda}_p$  son los autovalores de  $\mathbf{S}$
- $\widehat{\gamma}_1, \dots, \widehat{\gamma}_p$  los autovectores de  $\mathbf{S}$  asociados a  $\widehat{\lambda}_1 \geq \dots \geq \widehat{\lambda}_p$
- $\widehat{\mathbf{\Gamma}} = (\widehat{\gamma}_1, \dots, \widehat{\gamma}_p)$
- $\widehat{\mathbf{\Lambda}} = \text{diag}(\widehat{\lambda}_1, \dots, \widehat{\lambda}_p)$

$$\widehat{\mathbf{v}}_i^{(r)} = \widehat{\mathbf{\Gamma}}_r^T (\mathbf{x}_i - \bar{\mathbf{x}}) = (\widehat{v}_{i,1}, \dots, \widehat{v}_{i,r})^T$$

$$d_p(\mathbf{x}_i, \mathbf{x}_j) \approx d_p(\pi(\mathbf{x}_i - \bar{\mathbf{x}}, \widehat{\mathcal{H}}_r), \pi(\mathbf{x}_j - \bar{\mathbf{x}}, \widehat{\mathcal{H}}_r)) = d_q(\widehat{\mathbf{v}}_i^{(r)}, \widehat{\mathbf{v}}_j^{(r)})$$

Representación de las variables e individuos con  $n > p$ 

$$\mathbf{A} = \tilde{\mathbf{X}} = \begin{pmatrix} \tilde{\mathbf{x}}_1^T \\ \vdots \\ \tilde{\mathbf{x}}_n^T \end{pmatrix} = \mathbf{G}_p \mathbf{H}_p^T$$

donde  $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \bar{\mathbf{x}}$ . Luego,  $r = \text{rango}(\mathbf{A}) = p$  y  $\mathbf{A}^T \mathbf{1}_n = \mathbf{0}$ .

$$\mathbf{A}^T \mathbf{A} = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = \mathbf{Q} \quad \mathbf{A} \mathbf{A}^T = \begin{pmatrix} \|\tilde{\mathbf{x}}_1\|^2 & \tilde{\mathbf{x}}_1^T \tilde{\mathbf{x}}_2 & \dots & \tilde{\mathbf{x}}_1^T \tilde{\mathbf{x}}_n \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\mathbf{x}}_n^T \tilde{\mathbf{x}}_1 & \tilde{\mathbf{x}}_n^T \tilde{\mathbf{x}}_2 & \dots & \|\tilde{\mathbf{x}}_n\|^2 \end{pmatrix}$$

- $\delta_j^2 = (n-1)\hat{\lambda}_j$  son los autovalores de  $\mathbf{A}^T \mathbf{A} = (n-1)\mathbf{S}$
- $\hat{\gamma}_j$  (autovectores de  $\mathbf{S}$ ) son los autovectores de  $\mathbf{A}^T \mathbf{A}$
- $\mathbf{H}_p = \hat{\Gamma} \mathbf{\Delta} = \sqrt{(n-1)} \hat{\Gamma} \hat{\Lambda}^{1/2}$ .

Representación de las variables e individuos con  $n > p \dots$ 

$$\mathbf{A} = \tilde{\mathbf{X}} = \mathbf{G}_p \mathbf{H}_p^T$$

$$\mathbf{A}^T \mathbf{A} = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = \mathbf{Q} \quad \mathbf{A} \mathbf{A}^T = \begin{pmatrix} \|\tilde{\mathbf{x}}_1\|^2 & \tilde{\mathbf{x}}_1^T \tilde{\mathbf{x}}_2 & \dots & \tilde{\mathbf{x}}_1^T \tilde{\mathbf{x}}_n \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{\mathbf{x}}_n^T \tilde{\mathbf{x}}_1 & \tilde{\mathbf{x}}_n^T \tilde{\mathbf{x}}_2 & \dots & \|\tilde{\mathbf{x}}_n\|^2 \end{pmatrix}$$

- Recordemos que  $\hat{v}_{i,j} = \hat{\gamma}_j^T (\mathbf{x}_i - \bar{\mathbf{x}})$ ,  $\mathbf{G}_p = \mathbf{L}_p = (\ell_1, \dots, \ell_p)$

$$\ell_j = \frac{\tilde{\mathbf{X}} \hat{\gamma}_j}{\delta_j} = \frac{(\hat{v}_{1,j}, \dots, \hat{v}_{n,j})^T}{\sqrt{(n-1) \hat{\lambda}_j}}$$

son los autovectores de  $\mathbf{A} \mathbf{A}^T$ , o sea, dan la relación entre las observaciones.

## Representación de las variables e individuos

Luego, podemos escribir a  $\tilde{\mathbf{X}}$  como

$$\tilde{\mathbf{X}} = \mathbf{G}_p \mathbf{H}_p^T = (n-1)^{\frac{1}{2}} \mathbf{G}_p \left\{ (n-1)^{-\frac{1}{2}} \mathbf{H}_p \right\}^T = \mathbf{G} \mathbf{H}^T$$

$$\mathbf{H} = \hat{\Gamma} \hat{\Lambda}^{1/2} = (\hat{\lambda}_1^{1/2} \hat{\gamma}_1, \dots, \hat{\lambda}_p^{1/2} \hat{\gamma}_p) = \begin{pmatrix} \mathbf{h}_1^T \\ \vdots \\ \mathbf{h}_p^T \end{pmatrix}$$

$$\mathbf{G} = \begin{pmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{v}_{1,1} & \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{v}_{1,2} & \dots & \frac{1}{\sqrt{\hat{\lambda}_p}} \hat{v}_{1,p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{v}_{n,1} & \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{v}_{n,2} & \dots & \frac{1}{\sqrt{\hat{\lambda}_p}} \hat{v}_{n,p} \end{pmatrix} = \hat{\Lambda}^{-1/2} \begin{pmatrix} \hat{\mathbf{v}}_1^{(p)T} \\ \vdots \\ \hat{\mathbf{v}}_n^{(p)T} \end{pmatrix}$$

O sea, cada fila  $\mathbf{g}_i$  de  $\mathbf{G}$  corresponde a un individuo y cada fila  $\mathbf{h}_j$  de  $\mathbf{H}$  a una variable

## Representación de las $p$ variables

Recordemos que si  $\gamma_\ell = (\gamma_{\ell,1}, \dots, \gamma_{\ell,p})^T$ , la correlación entre  $x_j$ , la coordenada  $j$ -ésima de  $\mathbf{x}$ , y  $v_\ell$  está dada por

$$\text{CORR}(x_j, v_\ell) = \rho_{x_j, v_\ell} = \gamma_{\ell, j} \sqrt{\frac{\lambda_\ell}{\sigma_{jj}}}$$

Luego,

$$\mathbf{h}_j = \sqrt{s_{jj}} \begin{pmatrix} \hat{\rho}_{x_j, v_1} \\ \vdots \\ \hat{\rho}_{x_j, v_p}^T \end{pmatrix}$$

o sea,  $\mathbf{h}_j / \sqrt{s_{jj}}$  tiene por componentes las correlaciones estimadas de la variable  $x_j$  con los ejes  $\mathbf{v}^{(p)}$ .

# Representación de las $p$ variables

## Propiedad 1.

- El ángulo entre  $\mathbf{h}_j$  y  $\mathbf{h}_k$  es la correlación entre las variables  $j$  y  $k$
- $\|\mathbf{h}_j\|^2$  es la varianza muestral de la variable  $j$ ,  $s_{jj}$
- La covarianza muestral entre las variables  $j$  y  $k$  es  $s_{jk} = \mathbf{h}_j^T \mathbf{h}_k$
- $\|\mathbf{h}_j - \mathbf{h}_k\|^2$  es la varianza muestral de la diferencia entre las variables  $j$  y  $k$ .

Tenemos que  $\mathbf{G} = (n - 1)^{\frac{1}{2}} \mathbf{G}_p$  y  $\mathbf{H} = (n - 1)^{-\frac{1}{2}} \mathbf{H}_p$  donde

$$a) \mathbf{H}_p \mathbf{H}_p^T = \mathbf{A}^T \mathbf{A} = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} = (n - 1) \mathbf{S}$$

$$b) \mathbf{G}_p \mathbf{G}_p^T = \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \tilde{\mathbf{X}} \mathbf{S}^{-1} \tilde{\mathbf{X}}^T / (n - 1)$$

Por lo tanto,

$$\mathbf{H} \mathbf{H}^T = \mathbf{S} \quad \mathbf{G} \mathbf{G}^T = \tilde{\mathbf{X}} \mathbf{S}^{-1} \tilde{\mathbf{X}}$$

de donde

$$s_{jk} = \mathbf{h}_j^T \mathbf{h}_k$$

con lo cual se deduce la **Propiedad 1**. Por otra parte,

$$\begin{aligned} \mathbf{g}_i^T \mathbf{g}_j &= \tilde{\mathbf{x}}_i^T \mathbf{S}^{-1} \tilde{\mathbf{x}}_j \\ \|\mathbf{g}_i - \mathbf{g}_j\|^2 &= (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j)^T \mathbf{S}^{-1} (\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j) \\ &= (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{x}_j) \end{aligned}$$

**Problema:**  $\mathbf{g}_i$  y  $\mathbf{h}_j$  son vectores de  $\mathbb{R}^p$

**Problema:**  $\mathbf{g}_j$  y  $\mathbf{h}_j$  son vectores de  $\mathbb{R}^p$

**Solución:** Aproximamos a

$$\tilde{\mathbf{X}} = (n-1)^{\frac{1}{2}} \mathbf{L} \hat{\mathbf{\Lambda}}^{1/2} \hat{\mathbf{\Gamma}}^T = \sum_{j=1}^p (n-1)^{\frac{1}{2}} \ell_j \hat{\lambda}_j^{\frac{1}{2}} \hat{\gamma}_j$$

por

$$\tilde{\mathbf{X}}_{(r)} = \sum_{j=1}^r (n-1)^{\frac{1}{2}} \ell_j \hat{\lambda}_j^{\frac{1}{2}} \hat{\gamma}_j = \mathbf{G}_r \mathbf{H}_r^T$$

$$\mathbf{H}_r = \hat{\mathbf{\Gamma}}_r \hat{\mathbf{\Lambda}}_r^{1/2} = \begin{pmatrix} \mathbf{h}_1^{(r)T} \\ \vdots \\ \mathbf{h}_p^{(r)T} \end{pmatrix} = (\hat{\lambda}_1^{1/2} \hat{\gamma}_1, \dots, \hat{\lambda}_r^{1/2} \hat{\gamma}_r)$$

$$\mathbf{G}_r = (n-1)^{\frac{1}{2}} \mathbf{L}_r = \begin{pmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{v}_{1,1} & \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{v}_{1,2} & \cdots & \frac{1}{\sqrt{\hat{\lambda}_r}} \hat{v}_{1,r} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{v}_{n,1} & \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{v}_{n,2} & \cdots & \frac{1}{\sqrt{\hat{\lambda}_r}} \hat{v}_{n,r} \end{pmatrix} = \hat{\mathbf{\Lambda}}^{-\frac{1}{2}} \begin{pmatrix} \hat{\mathbf{v}}_1^{(r)T} \\ \vdots \\ \hat{\mathbf{v}}_n^{(r)T} \end{pmatrix}$$

Luego, las filas  $\mathbf{h}_j^{(r)}$  de  $\mathbf{H}_r$  cumplen aproximadamente la **Propiedad 1**.

Basta observar que

$$s_{jk} = \sum_{m=1}^p \hat{\lambda}_m \hat{\gamma}_{jm} \hat{\gamma}_{km} \approx \sum_{m=1}^r \hat{\lambda}_m \hat{\gamma}_{jm} \hat{\gamma}_{km} = \left( \mathbf{h}_j^{(r)} \right)^T \mathbf{h}_k^{(r)}$$

Además, si  $\pi(\mathbf{x}_i, \hat{\mathcal{H}}_r)_\ell$  es la coordenada  $\ell$  de  $\pi(\mathbf{x}_i, \hat{\mathcal{H}}_r)$ ,

$$\pi(\mathbf{x}_i - \bar{\mathbf{x}}, \hat{\mathcal{H}}_r)_\ell = \mathbf{h}_\ell^{(r)T} \text{DIAG}(\hat{\lambda}_1, \dots, \hat{\lambda}_r)^{-\frac{1}{2}} \hat{\mathbf{v}}_i^{(r)} = \left( \mathbf{h}_\ell^{(r)} \right)^T \mathbf{g}_i^{(r)}$$

Biplot,  $h$ -Plot

En particular, si  $r = 2$ ,  $\tilde{\mathbf{X}} \approx \tilde{\mathbf{X}}_{(2)} = \mathbf{G}_2 \mathbf{H}_2^T$

$$\mathbf{H}_2 = (\hat{\lambda}_1^{1/2} \hat{\gamma}_1, \hat{\lambda}_2^{1/2} \hat{\gamma}_2) = \begin{pmatrix} \mathbf{h}_1^{(2)T} \\ \vdots \\ \mathbf{h}_p^{(2)T} \end{pmatrix}$$

$$\mathbf{G}_2 = \begin{pmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{v}_{1,1} & \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{v}_{1,2} \\ \vdots & \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{v}_{n,1} & \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{v}_{n,2} \end{pmatrix} = \hat{\Lambda}^{-1/2} \begin{pmatrix} \hat{\mathbf{v}}_1^{(2)T} \\ \vdots \\ \hat{\mathbf{v}}_n^{(2)T} \end{pmatrix}$$

## Biplot, $h$ -Plot

Se verifica que

- La representación de los individuos como puntos en un plano mediante las filas  $\mathbf{g}_i^{(2)}$  de  $\mathbf{G}_2$  equivale a proyectar las observaciones en el plano de las dos primeras componentes principales estandarizadas para que tengan varianza 1.
- Las distancias euclídeas entre puntos del plano equivalen a las distancias de Mahalanobis entre las observaciones originales.
- La representación de las variables mediante los vectores  $\mathbf{h}_j^{(2)} \in \mathbb{R}^2$  es tal que el ángulo entre vectores equivale aproximadamente a la correlación entre las variables y la distancia al origen de los vectores es aproximadamente el desvío estándar de la variable asociada.

## $h$ -plot

El plot de  $\mathbf{h}_j^{(2)}$  se llama el  $h$ -plot de  $\mathbf{S}$  y una medida pérdida de ajuste posible es

$$\|\mathbf{S} - \mathbf{H}^{(2)}\mathbf{H}^{(2)\text{T}}\|_{\text{F}}$$

lo que lleva a la medida de bondad de ajuste

$$\rho_2 = \frac{\hat{\lambda}_1^2 + \hat{\lambda}_2^2}{\sum_{j=1}^p \hat{\lambda}_j^2}$$

donde  $\sum_{j=1}^p \hat{\lambda}_j^2 = \text{TR}(\mathbf{S}^2) = \sum_{j,k} s_{jk}^2$ .

**Otra opción** es tomar como medida de bondad del ajuste

$$\rho_{(2)}^2 = 1 - \frac{\|\tilde{\mathbf{X}} - \tilde{\mathbf{X}}_{(2)}\|_{\text{F}}^2}{\|\tilde{\mathbf{X}}\|_{\text{F}}^2} = \frac{\hat{\lambda}_1 + \hat{\lambda}_2}{\sum_{j=1}^p \hat{\lambda}_j}$$

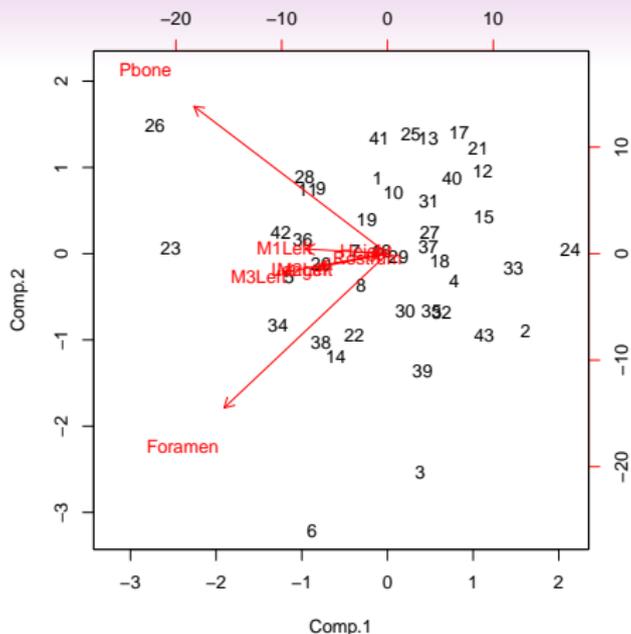
## Ejemplo *Microtus multiplex*

Los *Microtus multiplex* son una familia de roedores presentes en Europa. En este ejemplo se tomaron 43 especímenes y para cada uno se midieron 8 variables

- M1Left= Ancho del molar superior izquierdo # 1 (0.001mm)
- M2Left= Ancho del molar superior izquierdo # 2 (0.001mm)
- M3Left= Ancho del molar superior izquierdo # 3 (0.001mm)
- Foramen= Largo de la fosa incisiva (0.001mm)
- Pbone= Largo del hueso palatal (0.001mm)
- Length= Largo del cráneo (0.01mm)
- Height= Altura del cráneo sobre bullae (0.01mm)
- Rostrum= Ancho del cráneo a través del rostro (0.01mm)

obteniendo entonces vectores  $\mathbf{y}_i \in \mathbb{R}^8$ . Se presentan los resultados obtenidos con  $\mathbf{x}_i = \mathbf{y}_i/10$ .

$$\hat{\Lambda} = \text{diag}(1305.4337, 651.5147, 123.2253, 75.9081, 27.8237, 13.2150, 5.7182, 1.1248)$$



Da un plot de  $\mathbf{h}_j^{(2)}$ ,  $1 \leq j \leq 8$  en rojo y de  $\mathbf{g}_i^{(2)}$ ,  $1 \leq i \leq 43$  en negro. Los ángulos entre vectores dan aproximadamente las correlaciones y las normas el desvío estandar de la variable. Las distancias entre puntos representan la distancia de Mahalanobis entre las observaciones.

La bondad del ajuste medida por  $\rho_2 = 0.989803$  o  $\rho_{(2)}^2 = 0.8879224$  indica un buen ajuste.

