

○○  
○○○○○  
○

○○○○○  
○○

○○○  
○○○○○○○  
○○○

○○  
○○○○○  
○○○○○

○○  
○○  
○○○○

# Repaso de Multivariado I

**Graciela Boente**



Dados vectores  $\mathbf{x}_1, \dots, \mathbf{x}_n$  indicaremos por

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_n^T \end{pmatrix} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(p)})$$

donde  $\mathbf{x}^{(j)} = (x_{1,j}, \dots, x_{n,j})^T$ . Por otra parte,

$$\text{COV}(\mathbf{x}, \mathbf{y}) = \mathbb{E}(\mathbf{xy}^T) - \mathbb{E}(\mathbf{x}) \mathbb{E}(\mathbf{y}^T)$$

$$\text{VAR}(\mathbf{x}) = \text{COV}(\mathbf{x}) = \text{COV}(\mathbf{x}, \mathbf{x})$$



**Propiedades.** Dados  $\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{x}_i \in \mathbb{R}^p$

- i)  $\mathbb{E}(\mathbf{A}\mathbf{X}\mathbf{B} + \mathbf{C}) = \mathbf{A}\mathbb{E}\mathbf{X}\mathbf{B} + \mathbf{C}$
- ii) Si  $\mathbf{x} \in \mathbb{R}^p, \mathbf{y} \in \mathbb{R}^q$  son independientes  $\Rightarrow \text{COV}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ .
- iii) Dados vectores aleatorios  $\mathbf{x} \in \mathbb{R}^p, \mathbf{y} \in \mathbb{R}^q \Rightarrow$   
 $\text{Cov}(\mathbf{A}\mathbf{x}, \mathbf{B}\mathbf{y}) = \mathbf{A}\text{Cov}(\mathbf{x}, \mathbf{y})\mathbf{B}^T$ .
- iv) Dado un vector aleatorio  $\mathbf{x} \in \mathbb{R}^p, \text{VAR}(\mathbf{A}\mathbf{x}) = \mathbf{A}\text{VAR}(\mathbf{x})\mathbf{A}^T$

**Lema.** Si  $\mathbf{x}_1, \dots, \mathbf{x}_n$  son independientes,  $\mathbf{x}_i \in \mathbb{R}^p, \mathbb{E}\mathbf{x}_i = \boldsymbol{\mu}_i,$   
 $\text{COV}(\mathbf{x}_i) = \boldsymbol{\Sigma}_i$  entonces dada  $\mathbf{A} = (a_{ij}) \in \mathbb{R}^{p \times p}$

- i)  $\mathbb{E}(\mathbf{X}^T \mathbf{A} \mathbf{X}) = \sum_{i=1}^n a_{ii} \boldsymbol{\Sigma}_i + (\mathbb{E}\mathbf{X})^T \mathbf{A} \mathbb{E}\mathbf{X}$
- ii) Si  $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$  entonces  $\mathbb{E}(\mathbf{X}^T \mathbf{A} \mathbf{X}) = \text{TR}(\mathbf{A})\boldsymbol{\Sigma} + (\mathbb{E}\mathbf{X})^T \mathbf{A} \mathbb{E}\mathbf{X}$

$$\mathbf{X}^T \mathbf{A} \mathbf{X} = \sum_{1 \leq i, s \leq n} a_{is} \mathbf{x}_i \mathbf{x}_s^T.$$



La matriz de covarianza de  $\mathbf{X}$  se define como la matriz de covarianza de

$$\mathbf{y} = \text{VEC}(\mathbf{X}) = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}$$

**Propiedad.** Si  $\mathbf{x}_1, \dots, \mathbf{x}_n$  son independientes,  $\mathbf{x}_i \in \mathbb{R}^p$ ,  $\mathbb{E}\mathbf{x}_i = \boldsymbol{\mu}$ ,  $\text{Cov}(\mathbf{x}_i) = \boldsymbol{\Sigma}$  entonces

i)  $\mathbb{E}(\mathbf{X}) = \mathbf{1}_n \boldsymbol{\mu}^T$

ii)  $\text{Cov}(\mathbf{X}) = \text{Cov}(\mathbf{y}) = \mathbf{I} \otimes \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \dots & \mathbf{0} \\ \vdots & \vdots & & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma} \end{pmatrix}$



Dados  $\mathbf{x}_1, \dots, \mathbf{x}_n$  independientes,  $\mathbf{x}_i \in \mathbb{R}^p$ ,  $\mathbb{E}\mathbf{x}_i = \boldsymbol{\mu}$ ,  $\text{COV}(\mathbf{x}_i) = \boldsymbol{\Sigma}$

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i = \frac{1}{n} \mathbf{X}^T \mathbf{1}_n$$

$$\begin{aligned} \mathbf{S} &= \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \\ &= \frac{1}{n-1} \mathbf{Q} = \frac{1}{n-1} \mathbf{X}^T \left( \mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^T \right) \mathbf{X} \end{aligned}$$

Se tiene que

$$\mathbb{E}(\bar{\mathbf{x}}) = \boldsymbol{\mu}$$

$$\mathbb{E}(\mathbf{S}) = \boldsymbol{\Sigma}$$



## Definición 1

- Sea  $\boldsymbol{\mu} \in \mathbb{R}^p$  y  $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$  simétrica y definida positiva  
Se dice que  $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  si su densidad está dada por

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}$$

- Si  $\mathbf{x} \sim N(\mathbf{0}, \text{diag}(\lambda_1, \dots, \lambda_p))$

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}}} \frac{1}{\prod_{j=1}^p \lambda_j^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \sum_{j=1}^p \frac{x_j^2}{\lambda_j} \right\}$$

Por lo tanto,  $x_1, \dots, x_p$  son independientes  $x_j \sim N(0, \lambda_j)$ .

- En particular, si  $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_p)$ ,  $x_1, \dots, x_p$  son i.i.d.  $N(0, 1)$ .
- Si  $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  y  $\mathbf{A} \in \mathbb{R}^{p \times p}$  es no singular  $\implies$   
 $\mathbf{Ax} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$



## Definición 2

- Si  $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_p) \implies \varphi_{\mathbf{x}}(\mathbf{t}) = \exp\{-\frac{1}{2}\|\mathbf{t}\|^2\}$
- Si  $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \implies \varphi_{\mathbf{x}}(\mathbf{t}) = \exp\{i\mathbf{t}^T \boldsymbol{\mu}\} \exp\{-\frac{1}{2}\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}\}$

## Definición 2

Se dice que  $\mathbf{x}$  es normal multivariada si y sólo si  $\forall \mathbf{t} \in \mathbb{R}^p$  se tiene que  $\mathbf{t}^T \mathbf{x}$  es normal univariada.



## Teorema

Sea  $\mathbf{x}_1, \dots, \mathbf{x}_n$  i.i.d.  $\mathbf{x}_i \sim N(\mathbf{0}, \Sigma)$  con  $\Sigma > 0$ . Definamos  $\mathbf{X}^T = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ , o sea,  $\mathbf{X} = (\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)})$ .

Se tiene,

- $\mathbf{x}^{(j)} \sim N(\mathbf{0}, \sigma_{jj} \mathbf{I}_n)$
- Dado  $\mathbf{a} \in \mathbb{R}^n$ ,  $\mathbf{X}^T \mathbf{a} = \sum_{i=1}^n a_i \mathbf{x}_i \sim N(\mathbf{0}, \|\mathbf{a}\|^2 \Sigma)$
- Dados  $\mathbf{a}_i \in \mathbb{R}^n$ ,  $1 \leq i \leq r$  con  $r \leq n$  ortogonales, entonces  $\mathbf{X}^T \mathbf{a}_i$  son independientes.
- Dado  $\mathbf{b} \in \mathbb{R}^p$ ,  $\mathbf{X} \mathbf{b} = \sum_{j=1}^p b_j \mathbf{x}^{(j)} \sim N(\mathbf{0}, \sigma_{\mathbf{b}}^2 \mathbf{I}_n)$  con  $\sigma_{\mathbf{b}}^2 = \mathbf{b}^T \Sigma \mathbf{b}$ .





## Propiedades

- Si  $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_p) \implies \varphi_{\mathbf{x}}(\mathbf{t}) = \exp\{-\frac{1}{2}\|\mathbf{t}\|^2\}$
- Si  $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \implies \varphi_{\mathbf{x}}(\mathbf{t}) = \exp\{i\mathbf{t}^T \boldsymbol{\mu}\} \exp\{-\frac{1}{2}\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}\}$
- $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \iff \mathbf{a}^T \mathbf{x} \sim N(\mathbf{a}^T \boldsymbol{\mu}, \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a}), \forall \mathbf{a} \in \mathbb{R}^p$
- $\mathbf{x} = (x_1, \dots, x_p)^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \implies \mathbb{E}(\mathbf{x}) = \boldsymbol{\mu}$  y  $\text{COV}(\mathbf{x}) = \boldsymbol{\Sigma}$
- Sea  $\mathbf{x} = (x_1, \dots, x_p)^T \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , entonces,  
 $x_1, \dots, x_p$  son independientes  $\iff \boldsymbol{\Sigma}$  es diagonal.



## Propiedades

Sea  $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  con  $\boldsymbol{\Sigma} > 0$ . Definamos  $\mathbf{x} = \begin{pmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{pmatrix}$ ,

$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}$  y  $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$  con  $\mathbf{x}^{(i)}, \boldsymbol{\mu}^{(i)} \in \mathbb{R}^{p_i}$ ,

$\boldsymbol{\Sigma}_{ij} \in \mathbb{R}^{p_i \times p_j}$ ,  $p_1 + p_2 = p$ .

Entonces,

- $\mathbf{x}^{(1)} \sim N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11})$  y  $\mathbf{x}^{(2)} \sim N(\boldsymbol{\mu}^{(2)}, \boldsymbol{\Sigma}_{22})$ .
- Más aún,  $\mathbf{x}^{(1)}$  y  $\mathbf{x}^{(2)}$  son independientes  $\iff \boldsymbol{\Sigma}_{21} = 0$ .
- Dada  $\mathbf{A} \in \mathbb{R}^{q \times p}$ ,  $\text{rg}(\mathbf{A}) = q \implies \mathbf{Ax} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$

En particular, si  $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_p)$  y  $\mathbf{H} = (\mathbf{h}_1, \dots, \mathbf{h}_q) \in \mathbb{R}^{p \times q}$ , es ortogonal incompleta, o sea,  $\mathbf{H}^T \mathbf{H} = \mathbf{I}_q$ , entonces

$\mathbf{y} = \mathbf{H}^T \mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_q)$ .



## Propiedades

- d) Sea  $\mathbf{\Sigma} = \mathbf{H}\mathbf{\Lambda}\mathbf{H}^T$ , con  $\mathbf{H}$  ortogonal y  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$ ,  
 $\lambda_1 \geq \dots \geq \lambda_p$ .  
 Si  $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}) \implies \mathbf{H}^T(\mathbf{x} - \boldsymbol{\mu}) \sim N(\mathbf{0}, \mathbf{\Lambda})$ .
- e) Si  $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}) \iff \mathbf{x} = \mathbf{A}\mathbf{z} + \boldsymbol{\mu}$  con  $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_p)$  y  
 $\mathbf{A}\mathbf{A}^T = \mathbf{\Sigma}$ .
- f) Si  $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}) \implies (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \chi_p^2$
- g) Si  $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}) \implies \mathbf{x}^T \mathbf{\Sigma}^{-1} \mathbf{x} \sim \chi_p^2(\delta^2)$  con  $\delta^2 = \boldsymbol{\mu}^T \mathbf{\Sigma}^{-1} \boldsymbol{\mu}$



## Propiedades

Sea  $\mathbf{x} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_p)$

y sea  $\mathbf{H}_1 = (\mathbf{h}_1, \dots, \mathbf{h}_q) \in \mathbb{R}^{p \times q}$  ortogonal incompleta, o sea,  $\mathbf{H}_1^T \mathbf{H}_1 = \mathbf{I}_q$ .

Sea  $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$  ortogonal, o sea,  $\mathbf{H}^T \mathbf{H} = \mathbf{H} \mathbf{H}^T = \mathbf{I}_p$

Entonces

a)  $\mathbf{z} = \mathbf{H}_1^T \mathbf{x} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_q)$

b)  $\mathbf{z}$  es independiente de  $\mathbf{x}^T \mathbf{x} - \mathbf{z}^T \mathbf{z}$

c)  $\frac{\mathbf{x}^T \mathbf{x} - \mathbf{z}^T \mathbf{z}}{\sigma^2} \sim \chi_{p-q}^2$



## Teorema: Distribución condicional

Sea  $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  con  $\boldsymbol{\Sigma} > 0$ . Definamos  $\mathbf{x} = \begin{pmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \end{pmatrix}$ ,

$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix}$  y  $\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}$  con  $\mathbf{x}^{(i)}, \boldsymbol{\mu}^{(i)} \in \mathbb{R}^{p_i}$ ,

$\boldsymbol{\Sigma}_{ij} \in \mathbb{R}^{p_i \times p_j}$ ,  $p_1 + p_2 = p$ .

Entonces,

$$\mathbf{x}^{(1)} | \mathbf{x}^{(2)} = \mathbf{x}_0 \sim N\left(\boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}^{(2)}), \boldsymbol{\Sigma}_{11.2}\right)$$

donde  $\boldsymbol{\Sigma}_{11.2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{21}$ .



## Estimación

Sean  $\mathbf{x}_1, \dots, \mathbf{x}_n$  i.i.d.,  $\mathbf{x}_i \in \mathbb{R}^p$ ,  $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,  $\boldsymbol{\Sigma} > 0$

- La familia  $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  es una familia exponencial.
- $\mathbf{Q} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$  y  $\bar{\mathbf{x}}$  son estadísticos suficientes y completos.
- Por lo tanto, cualquier estimador insesgado basado en  $\mathbf{Q}$  y  $\bar{\mathbf{x}}$  resulta IMVU.



## Propiedad

- Los estimadores de máxima verosimilitud de  $\boldsymbol{\mu}$  y  $\boldsymbol{\Sigma}$  son  $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$  y  $\hat{\boldsymbol{\Sigma}} = \mathbf{Q}/n$ . Además si

$$L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{j=1}^n f_{\mathbf{x}}(\mathbf{x}_j, \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

tenemos que

$$L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = (2\pi)^{-\frac{np}{2}} \left( \det(\hat{\boldsymbol{\Sigma}}) \right)^{-\frac{n}{2}} e^{-\frac{np}{2}}$$

- $\mathbb{E}(\bar{\mathbf{x}}) = \boldsymbol{\mu}$

$$\mathbb{E}(\hat{\boldsymbol{\Sigma}}) = \frac{n-1}{n} \boldsymbol{\Sigma}$$

luego el estimador insesgado de  $\boldsymbol{\Sigma}$  es

$$\mathbf{S} = \frac{\mathbf{Q}}{n-1}$$



## Distribución de los estimadores

Sean  $\mathbf{x}_1, \dots, \mathbf{x}_n$  i.i.d.,  $\mathbf{x}_i \in \mathbb{R}^p$ ,  $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ,  $\boldsymbol{\Sigma} > 0$  y  $n \geq p + 1$

a)  $\bar{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, (1/n)\boldsymbol{\Sigma})$ ,

$$\mathbf{Q} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \sim \mathcal{W}(\boldsymbol{\Sigma}, p, n - 1) .$$

b)  $\bar{\mathbf{x}}$  y  $\mathbf{Q}$  son independientes.

c)

$$T^2 = n(n-1) (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \mathbf{Q}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) = n (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \sim T_{p, n-1}^2$$

o sea,

$$\frac{n-p}{p} \frac{T^2}{n-1} \sim \mathcal{F}_{p, n-p}$$





## Motivación

Sea  $\mathbf{x} \in \mathbb{R}^p$  tal que

$$\mathbb{E}(\mathbf{x}) = \mathbf{0} \quad \text{VAR}(\mathbf{x}) = \boldsymbol{\Sigma}$$

El método de componentes principales busca elegir  $q$  combinaciones lineales

$$z_1 = \gamma_1^T \mathbf{x}, \quad z_2 = \gamma_2^T \mathbf{x}, \quad \dots \quad z_q = \gamma_q^T \mathbf{x}$$

de modo tal que si  $\mathbf{z} = (z_1, \dots, z_q)$  entonces,  $\mathbf{z}$  explica una porción razonable de la dispersión total medida a través de  $\text{TR}(\boldsymbol{\Sigma})$ .

Como ejemplo, tomemos las dimensiones del caparazón de las tortugas.



## Ejemplo Tortugas

En este ejemplo se miden la dimensiones del caparazón de las tortugas siendo

- $x_1 = 10 \log(\text{longitud del caparazón})$ ,
- $x_2 = 10 \log(\text{ancho del caparazón})$

Se estudiaron 24 machos y 24 hembras.

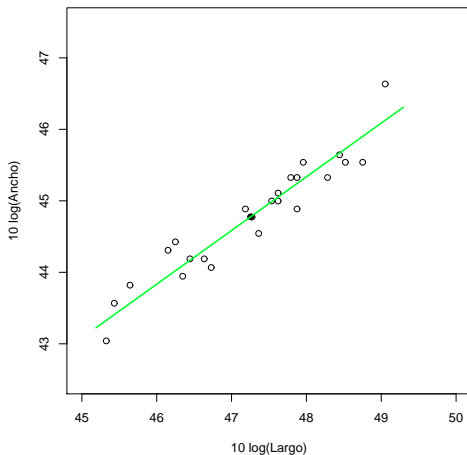
$$\bar{\mathbf{x}}_M = \begin{pmatrix} 47.254 \\ 44.776 \end{pmatrix} \quad \bar{\mathbf{x}}_H = \begin{pmatrix} 49.004 \\ 46.229 \end{pmatrix}$$

y

$$\hat{\boldsymbol{\gamma}}_{1,M} = (0.7996, 0.6005)^T \quad \hat{\boldsymbol{\gamma}}_{1,H} = (0.7892, 0.6141)^T$$

○○  
○○○○○  
○○○●○○○  
○○○○○  
○○○○○○○○  
○○○○○○  
○○○○○○  
○○○○○○○  
○○  
○○○○

# Machos



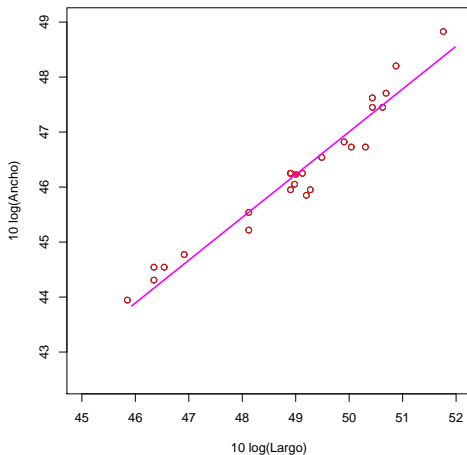


## Ejemplo Tortugas: Machos

- Observemos que los 24 puntos se distribuyen en forma bastante pareja a ambos lados de la recta.
- Esto está relacionado con un concepto introducido por Flury (1990) llamado *auto-consistencia* que tienen las componentes principales en el caso de datos normales.
- Si la forma del gráfico muestra curvatura, entonces veremos que en algunos segmentos de la recta hay demasiados puntos de un lado de la recta y muy pocos del otro, lo que hace dudar de que un ajuste lineal sea adecuado.



# Hembras



○○  
○○○○○  
○

○○○○○●  
○○

○○○  
○○○○○○○○  
○○○○

○○  
○○○○○○  
○○○○○

○○  
○○  
○○○○

## Ejemplo Tortugas

- En este ejemplo, ninguna de las dos variables  $x_1$  o  $x_2$  puede ser declarada como **independiente** o **dependiente**.
- Esto constituye la diferencia esencial con el análisis de regresión.
- La recta que obtuvimos **no es la recta de regresión** y se obtuvo minimizando la distancia de los puntos a la recta pero **midiendo la distancia** no verticalmente como en regresión sino **en forma ortogonal a la recta**.
- Es el principio de mínimos cuadrados ortogonales de Pearson (1901).



## Definición

Sea  $\mathbf{x} \in \mathbb{R}^p$  tal que

$$\mathbb{E}(\mathbf{x}) = \boldsymbol{\mu}$$

$$\text{VAR}(\mathbf{x}) = \boldsymbol{\Sigma}$$

Sean

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

$$\boldsymbol{\Gamma}^T \boldsymbol{\Sigma} \boldsymbol{\Gamma} = \boldsymbol{\Lambda}, \quad \boldsymbol{\Sigma} = \sum_{j=1}^p \lambda_j \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T$$

Luego, podemos escribir a  $\mathbf{x}$  como

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



## Definición

Sea el vector  $\mathbf{v} = \mathbf{\Gamma}^T(\mathbf{x} - \boldsymbol{\mu})$ . Las coordenadas  $v_1, \dots, v_p$  de  $\mathbf{v}$  se llaman **las componentes principales de  $\mathbf{x}$** .

La  $j$ -ésima componente principal es, por lo tanto,

$$v_j = \boldsymbol{\gamma}_j^T(\mathbf{x} - \boldsymbol{\mu}),$$

corresponde a la proyección ortogonal de  $(\mathbf{x} - \boldsymbol{\mu})$  en la dirección  $\boldsymbol{\gamma}_j$ .

Se llama  **$j$ -ésima componente principal estandarizada** a la variable

$$z_j = \lambda_j^{-\frac{1}{2}} v_j = \lambda_j^{-\frac{1}{2}} \boldsymbol{\gamma}_j^T(\mathbf{x} - \boldsymbol{\mu})$$

**Propiedad 1.** Las componentes principales  $v_1, \dots, v_p$  son no correlacionadas y  $\text{VAR}(v_j) = \lambda_j$ , o sea,

$$\text{VAR}(\mathbf{v}) = \mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$$





## Lemas previos

**Lema 1.** Sea  $\Sigma \in \mathbb{R}^{p \times p}$  una matriz simétrica definida no-negativa. Sean  $\lambda_1 \geq \dots \geq \lambda_p$  los autovalores de  $\Sigma$  y  $\gamma_1, \dots, \gamma_p$  los autovectores de  $\Sigma$  asociados a  $\lambda_1 \geq \dots \geq \lambda_p$ . Entonces

a)  $\sup_{\mathbf{u} \neq \mathbf{0}} \frac{\mathbf{u}^T \Sigma \mathbf{u}}{\mathbf{u}^T \mathbf{u}} = \lambda_1$  y el supremo se alcanza en  $\gamma_1$ .

b)  $\inf_{\mathbf{u} \neq \mathbf{0}} \frac{\mathbf{u}^T \Sigma \mathbf{u}}{\mathbf{u}^T \mathbf{u}} = \lambda_p$  y el infimo se alcanza en  $\gamma_p$ .

c)  $\sup_{\substack{\mathbf{u} \neq \mathbf{0} \\ \mathbf{u}^T \gamma_i = 0 \quad 1 \leq i \leq k}} \frac{\mathbf{u}^T \Sigma \mathbf{u}}{\mathbf{u}^T \mathbf{u}} = \lambda_{k+1}$  y el supremo se alcanza en  $\gamma_{k+1}$ .



## Lemas previos

**Teorema de Courant–Fisher.** Sea  $\Sigma \in \mathbb{R}^{p \times p}$  una matriz simétrica definida no-negativa. Sean

- $\lambda_1 \geq \dots \geq \lambda_p$  los autovalores de  $\Sigma$  y
- $\gamma_1, \dots, \gamma_p$  los autovectores de  $\Sigma$  asociados a  $\lambda_1 \geq \dots \geq \lambda_p$ .

Entonces

$$\inf_{\mathbf{B} \in \mathbb{R}^{p \times k}} \sup_{\mathbf{B}^T \mathbf{u} = 0} \frac{\mathbf{u}^T \Sigma \mathbf{u}}{\mathbf{u}^T \mathbf{u}} = \lambda_{k+1}$$

y se alcanza en  $\mathbf{B}_0 = (\gamma_1, \dots, \gamma_k)$ .



## Lemas previos

**Teorema de separación de Poincaré.** Sea  $\Sigma \in \mathbb{R}^{p \times p}$  una matriz simétrica definida no-negativa. Sean

- $\lambda_1 \geq \dots \geq \lambda_p$  los autovalores de  $\Sigma$  y
- $\gamma_1, \dots, \gamma_p$  los autovectores de  $\Sigma$  asociados a  $\lambda_1 \geq \dots \geq \lambda_p$ .

Entonces, si  $\mathbf{B} \in \mathbb{R}^{p \times k}$  es tal que  $\mathbf{B}^T \mathbf{B} = \mathbf{I}_k$ , se tiene que

$$\begin{aligned} \lambda_j(\mathbf{B}^T \Sigma \mathbf{B}) &\leq \lambda_j = \lambda_j(\Sigma) & 1 \leq j \leq k \\ \lambda_{k-j}(\mathbf{B}^T \Sigma \mathbf{B}) &\geq \lambda_{p-j} = \lambda_{p-j}(\Sigma) & 0 \leq j \leq k-1 \\ \lambda_s(\mathbf{B}^T \Sigma \mathbf{B}) &\geq \lambda_{p-k+s} = \lambda_{p-k+s}(\Sigma) & 1 \leq j \leq k \end{aligned}$$

donde  $\lambda_j(\mathbf{A})$  indica el  $j$ -ésimo autovalor de  $\mathbf{A}$ .



## Propiedades de optimalidad

**Propiedad 1.** (*Pearson, 1901*) Sea  $\mathbf{x} \in \mathbb{R}^p$  tal que

$$\mathbb{E}(\mathbf{x}) = \mathbf{0} \quad \text{VAR}(\mathbf{x}) = \mathbf{\Sigma}$$

- $\lambda_1 \geq \dots \geq \lambda_p$  los autovalores de  $\mathbf{\Sigma}$  y
- $\gamma_1, \dots, \gamma_p$  los autovectores de  $\mathbf{\Sigma}$  asociados a  $\lambda_1 \geq \dots \geq \lambda_p$ .
- $\mathcal{H}_0$  el subespacio generado por  $\gamma_1, \dots, \gamma_q$  donde  $\lambda_q > \lambda_{q+1}$ .

Indiquemos por  $\pi(\mathbf{x}, \mathcal{H})$  a la proyección ortogonal de  $\mathbf{x}$  sobre el subespacio  $\mathcal{H}$ . Entonces, se tiene que para todo subespacio  $\mathcal{H}$  de dimensión  $q$

$$\mathbb{E}\|\mathbf{x} - \pi(\mathbf{x}, \mathcal{H}_0)\|^2 \leq \mathbb{E}\|\mathbf{x} - \pi(\mathbf{x}, \mathcal{H})\|^2$$

o sea, las componentes principales dan el mejor ajuste lineal de dimensión  $q$ .



## Propiedades de optimalidad

**Propiedad 2.** Sea  $\mathbf{x} \in \mathbb{R}^p$  tal que

$$\mathbb{E}(\mathbf{x}) = \mathbf{0} \quad \text{VAR}(\mathbf{x}) = \mathbf{\Sigma}$$

- $\lambda_1 \geq \dots \geq \lambda_p$  los autovalores de  $\mathbf{\Sigma}$  y
- $\gamma_1, \dots, \gamma_p$  los autovectores de  $\mathbf{\Sigma}$  asociados a  $\lambda_1 \geq \dots \geq \lambda_p$ .

Sea  $\mathbf{y} \in \mathbb{R}^q$ ,  $q < p$ ,  $\mathbf{A} \in \mathbb{R}^{p \times q}$ ,  $\text{rango}(\mathbf{A}) = q$ ,  $\mathbf{b} \in \mathbb{R}^p$  y definamos

$$\mathbf{M} = \mathbb{E}(\mathbf{x} - \mathbf{A}\mathbf{y} - \mathbf{b})(\mathbf{x} - \mathbf{A}\mathbf{y} - \mathbf{b})^T$$

El error de la aproximación (o predicción lineal de  $\mathbf{x}$  basada en  $\mathbf{y}$ ) puede medirse mediante

$$\text{TR}(\mathbf{M}) = \mathbb{E}\|\mathbf{x} - \mathbf{A}\mathbf{y} - \mathbf{b}\|^2$$

El mínimo de  $\text{TR}(\mathbf{M})$  se alcanza en

$$\mathbf{b}_0 = \mathbf{0}, \quad \mathbf{A}_0 = (\gamma_1, \dots, \gamma_q), \quad \mathbf{y}_0 = (v_1, \dots, v_q)^T$$



## Propiedades de optimalidad

**Propiedad 3.** Sea  $\mathbf{x} \in \mathbb{R}^p$  tal que

$$\mathbb{E}(\mathbf{x}) = \boldsymbol{\mu} \quad \text{VAR}(\mathbf{x}) = \boldsymbol{\Sigma}$$

Sean  $\lambda_1 \geq \dots \geq \lambda_p$  los autovalores de  $\boldsymbol{\Sigma}$  y  $\gamma_1, \dots, \gamma_p$  los autovectores de  $\boldsymbol{\Sigma}$  asociados a  $\lambda_1 \geq \dots \geq \lambda_p$ . Entonces,

a)  $\max_{\|\mathbf{a}\|=1} \text{VAR}(\mathbf{a}^T \mathbf{x}) = \text{VAR}(v_1)$ , o sea, el máximo se alcanza en  $\gamma_1$ .

b)  $\max_{\substack{\|\mathbf{a}\|=1 \\ \text{Cov}(\mathbf{a}^T \mathbf{x}, v_j) = 0 \quad 1 \leq j \leq k}} \text{VAR}(\mathbf{a}^T \mathbf{x}) = \text{VAR}(v_{k+1})$ ,

es decir, el máximo se alcanza en  $\gamma_{k+1}$ .

La condición  $\text{COV}(\mathbf{a}^T \mathbf{x}, v_j) = 0$  asegura que no se repite información.

c)  $\sum_{j=1}^p \text{VAR}(v_j) = \text{TR}(\boldsymbol{\Sigma})$ .



## Propiedades de optimalidad

**Propiedad 4.** Sea  $\mathbf{x} \in \mathbb{R}^p$  tal que

$$\mathbb{E}(\mathbf{x}) = \boldsymbol{\mu} \quad \text{VAR}(\mathbf{x}) = \boldsymbol{\Sigma}$$

- $\lambda_1 \geq \dots \geq \lambda_p$  los autovalores de  $\boldsymbol{\Sigma}$  y
- $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p$  los autovectores de  $\boldsymbol{\Sigma}$  asociados a  $\lambda_1 \geq \dots \geq \lambda_p$ .

Queremos reemplazar a  $\mathbf{x}$  por  $q < p$  combinaciones lineales elegidas de modo a perder lo menos posible.

Tomemos  $y_j = \mathbf{a}_j^T \mathbf{x}$ ,  $1 \leq j \leq q$  y supongamos que  $\|\mathbf{a}_j\| = 1$  y  $\mathbf{a}_j^T \mathbf{a}_\ell = 0$  si  $j \neq \ell$ .



## Propiedades de optimalidad

**Propiedad 4.** Luego,

$$\text{VAR}(\mathbf{a}_j^T \mathbf{x}) = \mathbf{a}_j^T \boldsymbol{\Sigma} \mathbf{a}_j$$

por lo que las  $q$  combinaciones lineales  $(y_1, \dots, y_q)$  aportan

$$\sum_{j=1}^q \mathbf{a}_j^T \boldsymbol{\Sigma} \mathbf{a}_j$$

de la variación total de  $\mathbf{x}$  medida a través de la  $\text{TR}(\boldsymbol{\Sigma})$ .

Entonces, se cumple que

$$\max_{\substack{\|\mathbf{a}_j\|=1 \\ \mathbf{a}_j^T \mathbf{a}_\ell = 0 \quad j \neq \ell}} \sum_{j=1}^q \mathbf{a}_j^T \boldsymbol{\Sigma} \mathbf{a}_j = \sum_{j=1}^q \boldsymbol{\gamma}_j^T \boldsymbol{\Sigma} \boldsymbol{\gamma}_j = \sum_{j=1}^q \lambda_j$$





## Propiedades de optimalidad

Si  $\lambda_{q+1} = \dots = \lambda_p = 0$ , entonces  $v_{q+1}, \dots, v_p$  tienen varianza 0, o sea,

$$\mathbb{P}(\gamma_j^T(\mathbf{x} - \boldsymbol{\mu}) = 0 \quad \text{para todo } q + 1 \leq j \leq p) = 1$$

es decir,  $\mathbf{x} - \boldsymbol{\mu}$  yace en un subespacio de dimensión  $q$ .

Si esto no ocurre, deberíamos elegir  $q$  tal que  $\sum_{j=1}^q \lambda_j$  sea un porcentaje alto de la variación total de  $\mathbf{x}$ , o sea, de modo que por ejemplo

$$\frac{\sum_{j=1}^q \lambda_j}{\sum_{j=1}^p \lambda_j} = 0.95$$

Han visto test para verificar esta hipótesis basados en una muestra

$\mathbf{x}_1, \dots, \mathbf{x}_n$ . [► Ir a](#)



## Correlaciones

Supongamos que  $\mathbb{E}\mathbf{x} = \boldsymbol{\mu}$  y  $\text{VAR}(\mathbf{x}) = \boldsymbol{\Sigma}$ . Sea

$$\boldsymbol{\gamma}_\ell = (\gamma_{\ell,1}, \dots, \gamma_{\ell,p})^T$$

La correlación entre  $x_j$ , la coordenada  $j$ -ésima de  $\mathbf{x}$ , y  $v_\ell$  está dada por

$$\text{Corr}(x_j, v_\ell) = \rho_{x_j, v_\ell} = \gamma_{\ell,j} \sqrt{\frac{\lambda_\ell}{\sigma_{jj}}} \quad (1)$$

Supongamos que predecimos a  $\mathbf{x}$  usando un predictor lineal basado en  $\mathbf{v}_q = (v_1, \dots, v_q)^T$ . El mejor predictor lineal de  $\mathbf{x}$  basado en  $\mathbf{v}_q$  es

$$\tilde{\mathbf{x}} = \boldsymbol{\mu} + \text{COV}(\mathbf{x}, \mathbf{v}_q) \{ \text{VAR}(\mathbf{v}_q) \}^{-1} \mathbf{v}_q = \boldsymbol{\mu} + \boldsymbol{\Gamma}_q \mathbf{v}_q$$

donde  $\boldsymbol{\Gamma}_q = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_q)$  y el residuo es  $\mathbf{u} = \mathbf{x} - \tilde{\mathbf{x}}$ .



## Correlaciones

Luego, si  $\mathbf{\Lambda}_q = \text{diag}(\lambda_1, \dots, \lambda_q)$

$$\text{VAR}(\mathbf{u}) = \mathbf{\Sigma} - \mathbf{\Gamma}_q \mathbf{\Lambda}_q \mathbf{\Gamma}_q^T$$

o, sea,

$$\text{VAR}(x_j - \tilde{x}_j) = \sigma_{jj} - \sum_{\ell=1}^q \lambda_{\ell} \gamma_{\ell,j}^2$$

El término  $\lambda_{\ell} \gamma_{\ell,j}^2$  es la parte de la varianza de  $x_j$  explicada por  $v_{\ell}$  y por (1) es igual a  $\sigma_{jj} \rho_{x_j, v_{\ell}}^2$  de donde

$$\text{VAR}(x_j - \tilde{x}_j) = \sigma_{jj} \left( 1 - \sum_{\ell=1}^q \rho_{x_j, v_{\ell}}^2 \right)$$



## Definición

Sean  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$ , se dice que  $\mathbf{y}$  es *auto-consistente para  $\mathbf{x}$*  si

$$\mathbb{E}(\mathbf{x}|\mathbf{y}) = \mathbf{y}$$

## Ejemplos

- Si  $\mathbf{y} = \mathbf{x}$  entonces  $\mathbf{y}$  es auto-consistente para  $\mathbf{x}$ .
- Sea  $\boldsymbol{\mu} = \mathbb{E}(\mathbf{x})$  y definamos  $\mathbf{y}$  tal que  $\mathbb{P}(\mathbf{y} = \boldsymbol{\mu}) = 1$  entonces  $\mathbf{y}$  es auto-consistente para  $\mathbf{x}$ .
- Sea  $X \sim N(0, 1)$  y definamos

$$Y = \begin{cases} -\sqrt{\frac{2}{\pi}} & \text{si } X < 0 \\ \sqrt{\frac{2}{\pi}} & \text{si } X \geq 0 \end{cases}$$

entonces,  $Y$  es auto-consistente para  $X$



## Ejemplos

### Ejemplos

- Sea  $\mathbf{x} = (X_1, X_2)^T$  tal que  $\mathbb{E}(X_1) = \mu_1$  y  $\mathbb{E}(X_2) = \mu_2$

Sea

$$\mathbf{y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ X_2 \end{pmatrix}$$

Cuando  $\mathbf{y}$  es auto-consistente para  $\mathbf{x}$ ?



## Propiedades

- Si  $\mathbf{y}$  es auto-consistente para  $\mathbf{x}$ , entonces  $\mathbb{E}(\mathbf{y}) = \mathbb{E}(\mathbf{x})$ .
- Si  $\mathbf{y}$  es auto-consistente para  $\mathbf{x}$  y existen  $\text{VAR}(\mathbf{x}) = \mathbf{\Sigma}_x$  y  $\text{VAR}(\mathbf{y}) = \mathbf{\Sigma}_y$  entonces

$$\mathbb{E}\|\mathbf{x} - \mathbf{y}\|^2 = \text{TR}(\mathbf{\Sigma}_x) - \text{TR}(\mathbf{\Sigma}_y) = \sum_{j=1}^p \text{VAR}(x_j) - \text{VAR}(y_j)$$



## Propiedades

- Sea  $\mathbf{x} \in \mathbb{R}^p$ ,  $\mathbb{E}(\mathbf{x}) = \boldsymbol{\mu}$ ,  $\text{VAR}(\mathbf{x}) = \boldsymbol{\Sigma}$ .

Sea  $\mathbf{y} = \mathbf{x}_0 + \mathbf{a} \mathbf{a}^T (\mathbf{x} - \mathbf{x}_0)$  donde  $\mathbf{x}_0 \in \mathbb{R}^p$  es un punto fijo y  $\|\mathbf{a}\| = 1$ .

Si  $\mathbf{y}$  es auto-consistente para  $\mathbf{x}$  entonces

$$\mathbf{x}_0 = \boldsymbol{\mu} \quad \text{y} \quad \mathbf{a} = \boldsymbol{\gamma}_1$$

con  $\boldsymbol{\gamma}_1$  el autovector de  $\boldsymbol{\Sigma}$  asociado al mayor autovalor.

- Sea  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  y  $\boldsymbol{\Gamma}_q = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_q)$  con  $\boldsymbol{\Sigma} = \sum_{j=1}^p \lambda_j \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T$  y  $\lambda_1 \geq \dots \geq \lambda_p$ . Entonces,

$$\mathbf{y}_{(q)} = \boldsymbol{\mu} + \boldsymbol{\Gamma}_q \boldsymbol{\Gamma}_q^T (\mathbf{x} - \boldsymbol{\mu}) = \boldsymbol{\mu} + \boldsymbol{\Gamma}_q \mathbf{v}_q$$

es autoconsistente para  $\mathbf{x}$ .

Recordemos que  $\mathbf{y}_{(q)}$  era el mejor predictor lineal de  $\mathbf{x}$  basado en  $\mathbf{v}_q$ .



## Componentes principales muestrales

En la práctica,  $\boldsymbol{\mu}$  y  $\boldsymbol{\Sigma}$  son desconocidos y deben ser estimados a partir de una muestra aleatoria  $\mathbf{x}_1, \dots, \mathbf{x}_n$ .

$$\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}, \quad \mathbf{Q} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad \text{y} \quad \hat{\boldsymbol{\Sigma}} = \frac{\mathbf{Q}}{n}$$

Cuando  $\mathbf{x}$  tiene densidad, si  $n > p$ ,

$$\mathbb{P}(\mathbf{Q} > 0) = 1$$

y además,

$$\mathbb{P}(\lambda_1(\mathbf{Q}) > \lambda_2(\mathbf{Q}) > \dots > \lambda_p(\mathbf{Q})) = 1$$





## Componentes principales muestrales: $\hat{\Sigma} = \hat{\Gamma}\hat{\Lambda}\hat{\Gamma}^T$

- $\hat{\lambda}_1 > \dots > \hat{\lambda}_p$  los autovalores de  $\hat{\Sigma}$  y
- $\hat{\gamma}_1, \dots, \hat{\gamma}_p$  los autovectores de  $\hat{\Sigma}$  asociados a  $\hat{\lambda}_1 > \dots > \hat{\lambda}_p$ .

**Definición.** Para cada observación  $\mathbf{x}_i$  definimos el vector de componentes principales muestrales como

$$\hat{\mathbf{v}}_i = \hat{\Gamma}^T (\mathbf{x}_i - \bar{\mathbf{x}})$$

La coordenada  $j$ -ésima de  $\hat{\mathbf{v}}_i$ ,  $\hat{v}_{i,j}$ , se llama **la  $j$ -ésima componente principal de  $\mathbf{x}_i$** .

La  $j$ -ésima componente principal es, por lo tanto,

$$\hat{v}_j = \hat{\gamma}_j^T (\mathbf{x}_i - \bar{\mathbf{x}}),$$

corresponde a la proyección ortogonal de  $(\mathbf{x}_i - \bar{\mathbf{x}})$  en la dirección  $\hat{\gamma}_j$ .

Las propiedades que vimos anteriormente se cumplen en términos de la distribución empírica.



## 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

- Ancho del molar superior izquierdo # 1 (0.001mm)
- Ancho del molar superior izquierdo # 2 (0.001mm)
- Ancho del molar superior izquierdo # 3 (0.001mm)
- Largo de la fosa incisiva (0.001mm)
- Largo del hueso palatal (0.001mm)
- Largo del cráneo (0.01mm)
- Altura del cráneo sobre bullae (0.01mm)
- Ancho del cráneo a través del rostro (0.01mm)

obteniendose entonces vectores  $\mathbf{y}_i \in \mathbb{R}^8$ . Por conveniencia numérica, se presentan los resultados obtenidos con  $\mathbf{x}_i = \mathbf{y}_i/10$ .



## Ejemplo *Microtus multiplex*

$$\bar{\mathbf{x}} = (205.4535, 163.6465, 181.9930, 396.6488, 526.0209, 238.5977, 80.9442, 46.8698)^T$$

$$\mathbf{S} = \begin{pmatrix} 171.5130 & 97.4108 & 121.2151 & 158.7597 & 213.4108 & 88.4330 & 27.0469 & 23.2574 \\ 97.4108 & 102.3087 & 110.3706 & 161.7584 & 142.5469 & 73.9892 & 21.7843 & 17.8412 \\ 121.2151 & 110.3706 & 232.5688 & 250.9282 & 225.8311 & 110.3502 & 26.2622 & 24.0643 \\ 158.7597 & 161.7584 & 250.9282 & 737.7635 & 148.4182 & 187.5194 & 32.9356 & 42.2246 \\ 213.4108 & 142.5469 & 225.8311 & 148.4182 & 855.6855 & 159.8781 & 45.5893 & 36.5392 \\ 88.4330 & 73.9892 & 110.3502 & 187.5194 & 159.8781 & 87.0845 & 19.2189 & 19.3642 \\ 27.0469 & 21.7843 & 26.2622 & 32.9356 & 45.5893 & 19.2189 & 11.2949 & 5.2852 \\ 23.2574 & 17.8412 & 24.0643 & 42.2246 & 36.5392 & 19.3642 & 5.2852 & 5.7445 \end{pmatrix}$$

Los autovalores y autovectores de  $\mathbf{S}$  son  $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_p)$  y  $\hat{\Gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_p)$  donde

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

$$\hat{\Gamma} = \begin{pmatrix} 0.2719 & -0.0219 & -0.5571 & 0.6380 & 0.4369 & 0.1191 & -0.0428 & -0.0344 \\ 0.2179 & 0.0559 & -0.3577 & 0.1295 & -0.8556 & 0.2432 & -0.1161 & 0.0019 \\ 0.3409 & 0.0863 & -0.5097 & -0.7495 & 0.2152 & 0.0895 & 0.0174 & 0.0141 \\ 0.5404 & 0.7174 & 0.4063 & 0.0853 & 0.0603 & 0.1285 & 0.0277 & -0.0067 \\ 0.6404 & -0.6854 & 0.3389 & -0.0108 & -0.0046 & 0.0716 & -0.0002 & 0.0015 \\ 0.2328 & 0.0652 & -0.1129 & 0.0380 & -0.1206 & -0.9226 & -0.1878 & -0.1623 \\ 0.0563 & -0.0053 & -0.0875 & 0.0571 & -0.1100 & -0.1322 & 0.9686 & -0.1347 \\ 0.0534 & 0.0140 & -0.0402 & 0.0478 & -0.0209 & -0.1684 & 0.1010 & 0.9768 \end{pmatrix}$$



## Ejemplo *Microtus multiplex*

$$\frac{\hat{\lambda}_1}{\sum_{j=1}^p \hat{\lambda}_j} = 0.5923 \quad \frac{\hat{\lambda}_1 + \hat{\lambda}_2}{\sum_{j=1}^p \hat{\lambda}_j} = 0.8879 \quad \frac{\sum_{j=1}^3 \hat{\lambda}_j}{\sum_{j=1}^p \hat{\lambda}_j} = 0.9438$$

Además un estimador del desvío estandar de  $\hat{\lambda}_j$  es

$$\sqrt{\frac{2}{n}} \lambda_j$$

Luego, los desvíos estandar estimados de los autovalores  $s_{\hat{\lambda}_j}$  dan

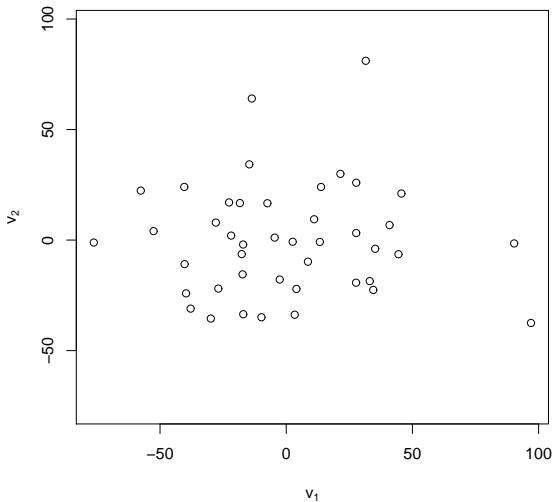
$j$	1	2	3	4	5	6	7	8
$\hat{\lambda}_j$	1305.434	651.515	123.225	75.908	27.824	13.215	5.718	1.125
$s_{\hat{\lambda}_j}$	281.537	140.509	26.575	16.371	6.001	2.850	1.233	0.243

Es decir,

- podemos pensar que la segunda componente está bien determinada, o sea, que  $\lambda_3 \neq \lambda_2$  y
- quizás dudemos sobre la tercera o sea, no podemos asegurar todavía que  $\lambda_3 \neq \lambda_4$ .

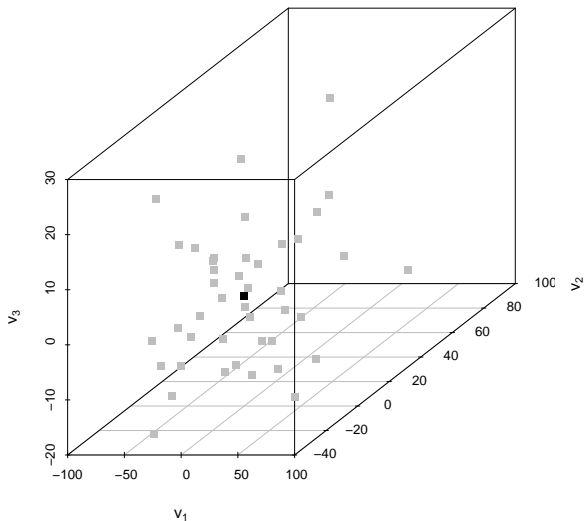


## Ejemplo *Microtus multiplex*: Dos Primeras CP





## Ejemplo *Microtus multiplex*: Tres Primeras CP





## Ejemplo *Microtus multiplex*

Correlaciones absolutas entre las variables y las 3 primeras componentes principales

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$
$\hat{\gamma}_1$	0.272	0.218	0.341	0.540	0.640	0.233	0.056	0.053
$ \hat{\rho}_{X_j, v_1} $	0.750	0.779	0.808	0.719	0.791	0.901	0.605	0.805
$\hat{\gamma}_2$	-0.022	0.056	0.086	0.717	-0.685	0.065	-0.005	0.014
$ \hat{\rho}_{X_j, v_2} $	0.043	0.141	0.144	0.674	0.598	0.178	0.04	0.149
$\hat{\gamma}_3$	-0.557	-0.358	-0.510	0.406	0.339	-0.113	-0.088	-0.040
$ \hat{\rho}_{X_j, v_3} $	0.472	0.393	0.371	0.166	0.129	0.134	0.289	0.186

Observemos que las coordenadas de  $\hat{\gamma}_1$  son todas positivas, esto ocurre porque  $\mathbf{S}$  tiene todos sus elementos positivos, o sea, todas las correlaciones son positivas.



## Ejemplo *Microtus multiplex*

Habíamos estudiado como testear  $H_{0,(1,2)} : \lambda_2 = \lambda_3$  y  $H_{0,(2,2)} : \lambda_3 = \lambda_4$  si son ciertas.

Recordemos que para testear  $H_{0,(r,h)} : \lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_{r+h}$  versus  $H_{1,(r,h)} : \lambda_{r+1} > \lambda_{r+2} > \dots > \lambda_{r+h}$  El test del cociente de máxima verosimilitud se basa en

$$M_{r,h} = \frac{\prod_{j=r+1}^{r+h} \hat{\lambda}_j}{\left( \frac{1}{h} \sum_{j=r+1}^{r+h} \hat{\lambda}_j \right)^h}$$

Rechazando para valores chicos de  $M_{r,h}$  y se tiene que

$$-n \log(M_{r,h}) \xrightarrow{D} \chi^2_{\frac{h(h+1)}{2} - 1}$$





En nuestro caso,  $H_{0,(1,2)} : \lambda_2 = \lambda_3$  y  $H_{0,(2,2)} : \lambda_3 = \lambda_4$

$$M_{1,2} = 0.5350 \quad -n \log(M_{1,2}) = 26.8942$$

$$M_{2,2} = 0.9435 \quad -n \log(M_{2,2}) = 2.4991$$

$$\chi_{2,0.05}^2 = 5.9915 \quad \chi_{2,0.01}^2 = 9.2103$$

Luego, rechazamos  $H_{0,(1,2)}$  pero no rechazamos  $H_{0,(2,2)}$ .  
Los  $p$ -valores son respectivamente,  $1.44 * 10^{-6}$  y 0.2866.

## Conclusión:

- No debemos dar ninguna interpretación relativa a  $v_3$  y  $v_4$  pues ese espacio no está bien determinado.
- Este resultado y el hecho que  $(\hat{\lambda}_1 + \hat{\lambda}_2) / \sum_{j=1}^p \hat{\lambda}_j = 0.8879$  sugeriría que la variabilidad en las mandíbulas de los roedores estudiados podría ser adecuadamente descrita por las dos primeras componentes principales.



## Ejemplo Microtus

$$\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{\sum_{j=1}^p \hat{\lambda}_j} = 0.8879$$

Estudiamos si vale  $H_{0,p_0}^*$

$$H_{0,p_0}^* : \frac{\lambda_1 + \lambda_2}{\sum_{j=1}^p \lambda_j} \leq p_0 \quad \text{versus} \quad H_{1,p_0}^* : \frac{\lambda_1 + \lambda_2}{\sum_{j=1}^p \lambda_j} > p_0$$

con  $p_0 = 0.80$  y  $p_0 = 0.85$ .

$p_0$	$\hat{\theta}_{p_0}$	$\hat{\sigma}_{p_0}$	$\sqrt{n} \hat{\theta}_{p_0} / \hat{\sigma}_{p_0}$	$p$ -valor
0.80	193.778	445.38	2.853	0.0022
0.85	83.579	357.04	1.535	0.0624



## Ejemplo Microtus

$$\frac{\hat{\lambda}_1 + \hat{\lambda}_2}{\sum_{j=1}^p \hat{\lambda}_j} = 0.8879$$

Estudiamos si vale  $H_{0,p_0}$

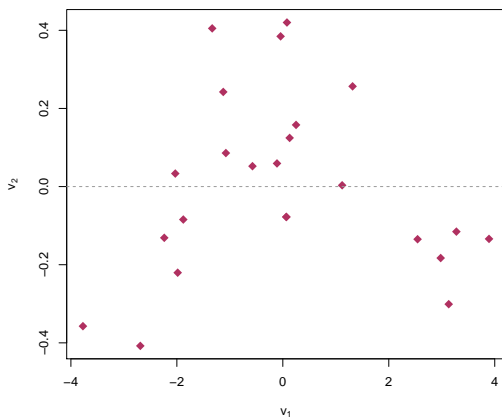
$$H_{0,p_0} : \frac{\lambda_1 + \lambda_2}{\sum_{j=1}^p \lambda_j} = p_0 \quad \text{versus} \quad H_{1,p_0} : \frac{\lambda_1 + \lambda_2}{\sum_{j=1}^p \lambda_j} \neq p_0$$

con  $p_0 = 0.90$ .

$p_0$	$\hat{\theta}_{p_0}$	$\hat{\sigma}_{p_0}$	$\sqrt{n}  \hat{\theta}_{p_0}  / \hat{\sigma}_{p_0}$	$p$ -valor	$\mathcal{I}_{p_0}$
0.90	-26.619	279.46	0.6246	0.532	(-62.213, 8.976)



## Ejemplo Tortugas Hembras





## Inferencia en el caso normal

**Teorema.** Sean  $\mathbf{x}_1, \dots, \mathbf{x}_n$  i.i.d  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , con  $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$  entonces

- $\hat{\boldsymbol{\Lambda}} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_p)$  es el EMV de  $\boldsymbol{\Lambda}$
- $\hat{\boldsymbol{\Gamma}} = (\hat{\gamma}_1, \dots, \hat{\gamma}_p)$  es el EMV de  $\boldsymbol{\Gamma}$

Además,

$$H_j = \sqrt{n}(\hat{\lambda}_j - \lambda_j) \xrightarrow{D} N(0, 2\lambda_j^2)$$

asintóticamente independientes entre sí.

Si las observaciones no son normales se puede probar que  $H_j$  es asintóticamente normal con varianza  $c\lambda_j^2$  pero no son necesariamente independientes



## Inferencia en el caso normal

Por otra parte, si  $\mathbf{x}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , con

$$\boldsymbol{\Sigma} = \text{diag}(\lambda_1, \dots, \lambda_p)$$

donde  $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$  y  $\hat{\boldsymbol{\gamma}}_k = (\hat{\gamma}_{k,1}, \dots, \hat{\gamma}_{k,p})^T$  entonces,

- $H_j = \sqrt{n}(\hat{\lambda}_j - \lambda_j)$ ,  $F_{kj} = \sqrt{n}\hat{\gamma}_{k,j}$  y  $E_{kk} = \sqrt{n}(\hat{\gamma}_{k,k} - 1)$  son independientes
- $F_{kj} \xrightarrow{D} N\left(0, \frac{\lambda_k \lambda_j}{(\lambda_k - \lambda_j)^2}\right)$
- $E_{kk} \xrightarrow{P} 0$



$H_{0,(r,h)} : \lambda_{r+1} = \lambda_{r+2} = \dots = \lambda_{r+h}$  versus  $H_{1,(r,h)} : \lambda_{r+1} > \lambda_{r+2} > \dots > \lambda_{r+h}$

El test del cociente de máxima verosimilitud se basa en

$$M_{r,h} = \frac{\prod_{j=r+1}^{r+h} \hat{\lambda}_j}{\left( \frac{1}{h} \sum_{j=r+1}^{r+h} \hat{\lambda}_j \right)^h}$$

Rechazando para valores chicos de  $M_{r,h}$  y se tiene que

$$-n \log(M_{r,h}) \xrightarrow{D} \chi^2_{\frac{h(h+1)}{2}-1}$$



$$H_{0,k} = H_{0,(p-k,k)} : \lambda_{p-k+1} = \lambda_{p-k+2} = \cdots = \lambda_p \text{ versus}$$

$$H_{1,k} : \lambda_{p-k+1} > \lambda_{p-k+2} > \cdots > \lambda_p$$

El test del cociente de máxima verosimilitud se basa en

$$M_k = \frac{\prod_{j=p-k+1}^p \hat{\lambda}_j}{\left( \frac{1}{k} \sum_{j=p-k+1}^p \hat{\lambda}_j \right)^k}$$

Rechazando para valores chicos de  $M_k$  y se tiene que

$$-n \log(M_k) \xrightarrow{D} \chi_{\frac{k(k+1)}{2}-1}^2$$





## Test para porcentajes

Queremos testear

$$H_{0,p_0} : \frac{\sum_{j=1}^q \lambda_j}{\sum_{j=1}^p \lambda_j} = p_0 \quad \text{versus} \quad H_{1,p_0} : \frac{\sum_{j=1}^q \lambda_j}{\sum_{j=1}^p \lambda_j} \neq p_0$$

donde  $p_0$  es un porcentaje prefijado. También podría interesarnos

$$H_{0,p_0}^* : \frac{\sum_{j=1}^q \lambda_j}{\sum_{j=1}^p \lambda_j} \leq p_0 \quad \text{versus} \quad H_{1,p_0}^* : \frac{\sum_{j=1}^q \lambda_j}{\sum_{j=1}^p \lambda_j} > p_0$$

Para obtener un test para estas hipótesis nos basaremos en la distribución asintótica de los autovalores

$$\sqrt{n}(\hat{\lambda}_j - \lambda_j) \xrightarrow{D} N(0, 2\lambda_j^2)$$

y usaremos que son asintóticamente independientes entre sí.



## Test para porcentajes

Observemos que

$$\frac{\sum_{j=1}^q \lambda_j}{\sum_{j=1}^p \lambda_j} = p_0$$

es equivalente a  $\theta = 0$  donde

$$\theta_{p_0} = (1 - p_0) \sum_{j=1}^q \lambda_j - p_0 \sum_{j=q+1}^p \lambda_j.$$

Definamos el estimador de  $\theta$

$$\hat{\theta}_{p_0} = (1 - p_0) \sum_{j=1}^q \hat{\lambda}_j - p_0 \sum_{j=q+1}^p \hat{\lambda}_j.$$

Sea

$$\sigma_{p_0}^2 = 2(1 - p_0)^2 \sum_{j=1}^q \lambda_j^2 + 2p_0^2 \sum_{j=q+1}^p \lambda_j^2$$



## Test para porcentajes

Entonces,

$$\sqrt{n}(\hat{\theta}_{p_0} - \theta_{p_0}) \xrightarrow{D} N(0, \sigma_{p_0}^2)$$

Por lo tanto, definiendo

$$\hat{\sigma}_{p_0}^2 = 2(1 - p_0)^2 \sum_{j=1}^q \hat{\lambda}_j^2 + 2p_0^2 \sum_{j=q+1}^p \hat{\lambda}_j^2$$

resulta que

$$\sqrt{n}(\hat{\theta}_{p_0} - \theta_{p_0}) / \hat{\sigma}_{p_0} \xrightarrow{D} N(0, 1)$$



## Test para porcentajes

Un intervalo de confianza asintótico de nivel  $1 - \alpha$  para  $\theta$  es

$$\mathcal{I}_{p_0} = \left[ \hat{\theta}_{p_0} - \frac{\hat{\sigma}_{p_0}}{\sqrt{n}} z_{\frac{\alpha}{2}}, \hat{\theta}_{p_0} + \frac{\hat{\sigma}_{p_0}}{\sqrt{n}} z_{\frac{\alpha}{2}} \right]$$

Luego, rechazo  $H_{0,p_0}$  si  $0 \notin \mathcal{I}_{p_0}$  o equivalentemente si

$$\sqrt{n} \frac{|\hat{\theta}_{p_0}|}{\hat{\sigma}_{p_0}} \geq z_{\frac{\alpha}{2}}$$

Por otra parte, un test asintótico para  $H_{0,p_0}^*$  o sea, para  $H_{0,p_0}^* : \theta_{p_0} \leq 0$  versus  $\theta_{p_0} > 0$  rechaza si

$$\sqrt{n} \frac{\hat{\theta}_{p_0}}{\hat{\sigma}_{p_0}} \geq z_{\alpha}$$