Repaso de Multivariado I

Graciela Boente

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

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

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

$$\begin{aligned} \mathrm{Cov}\left(\mathbf{x}, \mathbf{y}\right) &= \mathbb{E}\left(\mathbf{x}\mathbf{y}^{\mathrm{T}}\right) - \mathbb{E}\left(\mathbf{x}\right) \mathbb{E}\left(\mathbf{y}^{\mathrm{T}}\right) \\ \mathrm{Var}\left(\mathbf{x}\right) &= \mathrm{Cov}\left(\mathbf{x}\right) = \mathrm{Cov}\left(\mathbf{x}, \mathbf{x}\right) \end{aligned}$$

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

i) $\mathbb{E}(\mathbf{AXB} + \mathbf{C}) = \mathbf{A}\,\mathbb{E}\mathbf{X}\,\mathbf{B} + \mathbf{C}$

ii) Si $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^q$ son independientes $\Rightarrow \mathrm{Cov}(\mathbf{x}, \mathbf{y}) = 0$.

iii) Dados vectores aleatorios $\mathbf{x} \in \mathbb{R}^p$, $\mathbf{y} \in \mathbb{R}^q \Rightarrow$
$\mathrm{Cov}(\mathbf{Ax}, \mathbf{By}) = \mathbf{A}\mathrm{Cov}(\mathbf{x}, \mathbf{y})\,\mathbf{B}^{\mathrm{T}}$.

iv) Dado un vector aleatorio $\mathbf{x} \in \mathbb{R}^p$, $\mathrm{Var}(\mathbf{Ax}) = \mathbf{A}\mathrm{Var}(\mathbf{x})\,\mathbf{A}^{\mathrm{T}}$

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

i) $\mathbb{E}(\mathbf{X}^{\mathrm{T}}\mathbf{AX}) = \sum_{i=1}^n a_{ii}\boldsymbol{\Sigma}_i + (\mathbb{E}\mathbf{X})^{\mathrm{T}}\,\mathbf{A}\,\mathbb{E}\mathbf{X}$

ii) Si $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ entonces $\mathbb{E}(\mathbf{X}^{\mathrm{T}}\mathbf{AX}) = \mathrm{TR}(\mathbf{A})\boldsymbol{\Sigma} + (\mathbb{E}\mathbf{X})^{\mathrm{T}}\,\mathbf{A}\,\mathbb{E}\mathbf{X}$

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

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

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

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

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

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

Dados $\mathbf{x}_1, \ldots, \mathbf{x}_n$ son independientes, $\mathbf{x}_i \in \mathbb{R}^p$, $\mathbb{E}\mathbf{x}_i = \boldsymbol{\mu}$, $\mathrm{Cov}(\mathbf{x}_i) = \boldsymbol{\Sigma}$

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

Se tiene que

$$\begin{aligned}
\mathbb{E}(\overline{\mathbf{x}}) &= \boldsymbol{\mu} \\
\mathbb{E}(\mathbf{S}) &= \boldsymbol{\Sigma}
\end{aligned}$$

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

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

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

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

  Por lo tanto, $x_1, \ldots, x_p$ son independientes $x_j \sim N(0, \lambda_j)$.
- En particular, si $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_p)$, $x_1, \ldots, x_p$ 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 $\Longrightarrow$
  $\mathbf{A}\mathbf{x} + \mathbf{b} \sim N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{T}})$

Definición 2

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

Definición 2

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

Teorema

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

Se tiene,

a) $\mathbf{x}^{(j)} \sim N(\mathbf{0}, \sigma_{jj}\, \mathbf{I}_n)$

b) Dado $\mathbf{a} \in \mathbb{R}^n$, $\mathbf{X}^{\mathrm{T}}\mathbf{a} = \sum_{i=1}^{n} a_i \mathbf{x}_i \sim N(\mathbf{0}, \|\mathbf{a}\|^2 \mathbf{\Sigma})$

c) Dados $\mathbf{a}_i \in \mathbb{R}^n$, $1 \leq i \leq r$ con $r \leq n$ ortogonales, entonces $\mathbf{X}^{\mathrm{T}}\mathbf{a}_i$ son independientes.

d) Dado $\mathbf{b} \in \mathbb{R}^p$, $\mathbf{Xb} = \sum_{j=1}^{p} b_j \mathbf{x}^{(j)} \sim N(\mathbf{0}, \sigma_{\mathbf{b}}^2 \mathbf{I_n})$

Propiedades

- Si $\mathbf{x} \sim N(\mathbf{0}, \mathbf{I}_p) \Longrightarrow \varphi_{\mathbf{x}}(\mathbf{t}) = \exp\{-\frac{1}{2}\|\mathbf{t}\|^2\}$
- Si $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longrightarrow \varphi_{\mathbf{x}}(\mathbf{t}) = \exp\{i\mathbf{t}^{\mathrm{T}}\boldsymbol{\mu}\}\exp\{-\frac{1}{2}\mathbf{t}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{t}\}$
- $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longleftrightarrow \mathbf{a}^{\mathrm{T}}\mathbf{x} \sim N(\mathbf{a}^{\mathrm{T}}\boldsymbol{\mu}, \mathbf{a}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{a})$, $\forall \mathbf{a} \in \mathbb{R}^p$
- $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \Longrightarrow \mathbb{E}(\mathbf{x}) = \boldsymbol{\mu}$ y $\mathrm{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$
- Sea $\mathbf{x} = (x_1, \ldots, x_p)^{\mathrm{T}} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, entonces, $x_1, \ldots, x_p$ son independientes $\Longleftrightarrow \boldsymbol{\Sigma}$ 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 $\mathbf{x}^{(i)}, \boldsymbol{\mu}^{(i)} \in \mathbb{R}^{p_i}$,

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

Entonces,

a) $\mathbf{x}^{(1)} \sim N(\boldsymbol{\mu}^{(1)}, \boldsymbol{\Sigma}_{11})$ y $\mathbf{x}_2 \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$.

b) Más aún, $\mathbf{x}^{(1)}$ y $\mathbf{x}^{(2)}$ son independientes $\iff \boldsymbol{\Sigma}_{21} = 0$.

c) Dada $\mathbf{A} \in \mathbb{R}^{q \times p}$, $rg(\mathbf{A}) = q \implies \mathbf{A}\mathbf{x} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{T}})$

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

Propiedades

d) Sea $\mathbf{\Sigma} = \mathbf{H\Lambda H}^{\mathrm{T}}$, con $\mathbf{H}$ ortogonal y $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, $\lambda_1 \geq \ldots, \lambda_p$.
Si $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}) \Longrightarrow \mathbf{H}^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu}) \sim N(\mathbf{0}, \mathbf{\Lambda})$.

e) Si $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}) \Longleftrightarrow \mathbf{x} = \mathbf{Az} + \boldsymbol{\mu}$ con $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_p)$ y $\mathbf{AA}^{\mathrm{T}} = \mathbf{\Sigma}$.

f) Si $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}) \Longrightarrow (\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu) \sim \chi^2_p$

g) Si $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{\Sigma}) \Longrightarrow \mathbf{x}^{\mathrm{T}} \mathbf{\Sigma}^{-1} \mathbf{x} \sim \chi^2_p(\delta^2)$ con $\delta^2 = \boldsymbol{\mu}^{\mathrm{T}} \mathbf{\Sigma}^{-1} \boldsymbol{\mu}$

Propiedades

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

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

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

Entonces

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

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

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

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 $\mathbf{x}^{(i)}, \boldsymbol{\mu}^{(i)} \in \mathbb{R}^{p_i}$, $\boldsymbol{\Sigma}_{ii} \in \mathbb{R}^{p_i \times p_i}$, $p_1 + p_2 = p$.

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

Definición 1

- Sea $\mathbf{\Sigma} \in \mathbb{R}^{p\times p}$ simétrica y definida positiva y $n \geq p$.
- Sea $\mathbf{W} = (w_{ij}) \in \mathbb{R}^{p\times p}$ simétrica y definida positiva con probabilidad 1.
- Se dice que $\mathbf{W} \sim \mathcal{W}(\mathbf{\Sigma}, p, n)$ si la densidad conjunta de los $p(p+1)/2$ elementos distintos de $\mathbf{W}$ está dada por

$$f(\mathbf{w}) = c^{-1}|\mathbf{W}|^{\frac{n-p-1}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{W}\right)\right\}$$

con $\mathbf{w} = (w_{11}, w_{12}, \ldots, w_{pp})^{\mathrm{T}}$ y

$$c = 2^{\frac{np}{2}}\,|\mathbf{\Sigma}|^{\frac{n}{2}}\,\pi^{\frac{p(p-1)}{4}}\prod_{j=1}^{p}\Gamma\left(\frac{1}{2}(n+1-j)\right)$$

Es fácil ver que si $p = 1$ $\mathcal{W}(\sigma^2, 1, n) = \sigma^2\chi^2_n$.

Veremos que $\forall\,\mathbf{u} \neq 0$, $\mathbf{u} \in \mathbb{R}^p$, $\mathbf{u}^{\mathrm{T}}\mathbf{W}\mathbf{u} \sim (\mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{u})\chi^2_n$.

Definición 2

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

  $$\mathbf{X} = \begin{pmatrix} \mathbf{x}_i^{\mathrm{T}} \\ \vdots \\ \mathbf{x}_n^{\mathrm{T}} \end{pmatrix}, \text{ o sea, } \mathbf{X}^{\mathrm{T}} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

  - La distribución $\mathcal{W}(\mathbf{\Sigma}, p, n)$ es la distribución de

    $$\mathbf{W} = \mathbf{X}^{\mathrm{T}}\mathbf{X} = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}}$$

- Sean $\mathbf{x}_1, \ldots, \mathbf{x}_n$ independientes, $\mathbf{x}_i \in \mathbb{R}^p$, $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}_i, \mathbf{\Sigma})$

  - La distribución Wishart no central con parámetro de no centralidad $\mathbf{\Delta} = \sum_{i=1}^{n} \boldsymbol{\mu}_i \boldsymbol{\mu}_i^{\mathrm{T}}$, $\mathcal{W}(\mathbf{\Sigma}, p, n)(\mathbf{\Delta})$, es la distribución de

    $$\mathbf{W} = \mathbf{X}^{\mathrm{T}}\mathbf{X} = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}}$$

- Sean $\mathbf{x}_1, \ldots, \mathbf{x}_n$ independientes, $\mathbf{x}_i \in \mathbb{R}^p$, $\mathbf{x}_i \sim N_p(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ y $\mathbf{W} = \mathbf{X}^{\mathrm{T}}\mathbf{X} = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}}$. O sea, $\mathbf{W} \sim \mathcal{W}(\boldsymbol{\Sigma}, p, n)(\boldsymbol{\Delta})$ entonces

$$\mathbb{E}(\mathbf{W}) = n\boldsymbol{\Sigma} + \mathbf{M}^{\mathrm{T}}\mathbf{M}$$

  con $\mathbf{M} = \mathbb{E}(\mathbf{X})$.

- En particular, $\mathbf{W} \sim \mathcal{W}(\boldsymbol{\Sigma}, p, n)$

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

Propiedades

- Si $n < p$ entonces, $\text{rg}(\mathbf{W}) = \text{rg}(\mathbf{X}) \leq \min(n, p) = n \Longrightarrow \mathbf{W}$ es singular

- Sean $\mathbf{x}_1, \ldots, \mathbf{x}_n$ i.i.d., $\mathbf{x}_i \in \mathbb{R}^p$, $\mathbf{x}_i \sim N_p(\mathbf{0}, \mathbf{\Sigma})$.
  Si $n \geq p$ y $\mathbf{\Sigma} > 0$, se puede ver que $\mathbf{W} = \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}}$ tiene densidad y la densidad de $\mathbf{W}$ es la dada en la definición 1 (Kshirsagar, A.M. (1972) Multivariate Analysis, pág. 51-58, 77-78).

- O sea, ambas definiciones son equivalentes si $n \geq p$ y $\mathbf{\Sigma} > 0$.

Propiedades

Teorema (Okamoto, 1973). Sean $\mathbf{X}^{\mathrm{T}} = (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^{p \times n}$ y $\mathbf{A} \in \mathbb{R}^{n \times n}$ simétrica de rango r.

Si los elementos de $\mathbf{X}$ tienen densidad conjunta, entonces

$$\mathbb{P}\left(\mathrm{rg}\left(\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X}\right) = \min(p, r)\right) = 1$$

$$\mathbb{P}\left(\text{los autovalores no nulos de } \mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X} \text{ sean distintos}\right) = 1\,.$$

Corolario. Sea $n \geq p$, $\mathbf{\Sigma} > 0$ y $\mathbf{W} \sim \mathcal{W}(\mathbf{\Sigma}, p, n)$, entonces

$$\mathbb{P}\left(\mathrm{rg}\left(\mathbf{W}\right) = p\right) = 1$$

o sea, $\mathbb{P}\left(\mathbf{W} > 0\right) = 1$ y los autovalores de $\mathbf{W}$ son distintos con probabilidad 1.

Propiedades

Sea $\mathbf{W} \sim \mathcal{W}(\mathbf{\Sigma}, p, n)$

a) Sea $\mathbf{C} \in \mathbb{R}^{q \times p}$, $\text{rg}(\mathbf{C}) = q \leq p \Longrightarrow \mathbf{C}\mathbf{W}\mathbf{C}^{\mathrm{T}} \sim \mathcal{W}(\mathbf{C}\mathbf{\Sigma}\mathbf{C}^{\mathrm{T}}, q, n)$.

En particular, si $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}^{\mathrm{T}}$ con $\mathbf{C}$ triangular inferior $\Longrightarrow$ $\mathbf{C}^{-1}\mathbf{W}(\mathbf{C}^{-1})^{\mathrm{T}} \sim \mathcal{W}(\mathbf{I}_p, p, n)$

b) Si $\mathbf{u} \neq 0$, $\mathbf{u} \in \mathbb{R}^p \Longrightarrow \mathbf{u}^{\mathrm{T}}\mathbf{W}\mathbf{u}/(\mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{u}) \sim \chi^2_n$.

c) En particular,

$$\frac{w_{jj}}{\sigma_{jj}} \sim \chi^2_n \, .$$

d) Si $\mathbf{y} \neq 0$, $\mathbf{y} \in \mathbb{R}^p$, es un vector aleatorio independiente de $\mathbf{W}$ tal que $\mathbb{P}(\mathbf{y} \neq \mathbf{0}) = 1 \Longrightarrow$

$$\frac{\mathbf{y}^{\mathrm{T}}\mathbf{W}\mathbf{y}}{\mathbf{y}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{y}} \sim \chi^2_n$$

Propiedades

e) Sea $\mathbf{W} = \begin{pmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{pmatrix} \sim \mathcal{W}(\mathbf{\Sigma}, p, n)$, con $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$, $\mathbf{W}_{11} \in \mathbb{R}^{k \times k}$ y $\mathbf{\Sigma}_{11} \in \mathbb{R}^{k \times k}$, entonces

- $\mathbf{W}_{11} \sim \mathcal{W}(\mathbf{\Sigma}_{11}, k, n)$
- $\mathbf{W}_{22} \sim \mathcal{W}(\mathbf{\Sigma}_{22}, p-k, n)$
- Si $\mathbf{\Sigma}_{12} = \mathbf{0} \Longrightarrow \mathbf{W}_{11}$ y $\mathbf{W}_{22}$ son independientes

g) Si $\mathbf{W}_1, \ldots, \mathbf{W}_k$ son independientes $\mathbf{W}_i \sim \mathcal{W}(\mathbf{\Sigma}, p, n_i)$ entonces

$$\sum_{j=1}^{k} \mathbf{W}_j \sim \mathcal{W}(\mathbf{\Sigma}, p, \sum_{j=1}^{k} n_j)$$

Teorema de descomposición de Bartlett

Sea $\mathbf{D} \sim \mathcal{W}(\mathbf{I}_p, p, n)$ con $n \geq p$ y $\mathbf{D} = \mathbf{B}\mathbf{B}^{\mathrm{T}}$ con $\mathbf{B}$ triangular inferior entonces

- Los elementos b_{ij} $(1 \leq j \leq i \leq p)$ son todos independientes,
- $b_{ii}^2 \sim \chi^2_{n-i+1}$ y
- $b_{ij} \sim N(0,1)$, $1 \leq j < i \leq p$
- o sea, la densidad de los elementos no nulos de $\mathbf{B}$ es

$$\prod_{i=1}^{p} \prod_{j=1}^{i-1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} b_{ij}^2} \prod_{\ell=1}^{p} f_{\chi^2_{n-k+1}}(b_{kk}^2)$$

Propiedades

Sea $\mathbf{\Sigma} > 0$, $\mathbf{x}_1, \ldots \mathbf{x}_n$ independientes $\mathbf{x}_i \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{X}^{\mathrm{T}} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ y $\mathbf{A} \in \mathbb{R}^{n \times n}$ simétrica.

a) Si $\mathbf{A} \in \mathbb{R}^{n \times n}$ es idempotente, con $\mathrm{rg}(\mathbf{A}) = r$ entonces

$$\mathbf{X}^{\mathrm{T}} \mathbf{A} \mathbf{X} \sim \mathcal{W}(\mathbf{\Sigma}, p, r)$$

b) Sea $\mathbf{Y} = \mathbf{A}\mathbf{X}$ y $\mathbf{Z} = \mathbf{C}\mathbf{X}$, si $\mathbf{A}\mathbf{C}^{\mathrm{T}} = \mathbf{0}$ entonces

$\mathbf{Y}$ y $\mathbf{Z}$ son independientes

Recordemos que

a) Sea $\mathbf{y} \sim N_p(\mathbf{0}, \sigma^2 \mathbf{I}_p)$ y $\mathbf{P} \in \mathbb{R}^{p \times p}$ es simétrica, entonces

$$\frac{\mathbf{y}^{\mathrm{T}} \mathbf{P} \mathbf{y}}{\sigma^2} \sim \chi^2_r \iff \mathbf{P}^2 = \mathbf{P} \quad \text{y} \quad \operatorname{rg}(\mathbf{P}) = r$$

b) Sea $\mathbf{y} \sim N_p(\mathbf{0}, \sigma^2 \mathbf{I}_p)$ y $\mathbf{P}_\ell \in \mathbb{R}^{p \times p}$ simétricas, $\ell = 1, 2$

Definamos

$$U_\ell = \frac{\mathbf{y}^{\mathrm{T}} \mathbf{P}_\ell \mathbf{y}}{\sigma^2}, \quad \ell = 1, 2\,.$$

Supongamos que $U_\ell \sim \chi^2_{r_\ell}$ entonces

$$U_1 \text{ y } U_2 \text{ son independientes} \iff \mathbf{P}_1 \mathbf{P}_2 = \mathbf{0}$$

Teorema

Sea $\mathbf{\Sigma} > 0$, $\mathbf{x}_1, \ldots \mathbf{x}_n$ independientes $\mathbf{x}_i \sim N_p(\mathbf{0}, \mathbf{\Sigma})$,
$\mathbf{X}^{\mathrm{T}} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$.

Sean

- $\mathbf{u} \neq \mathbf{0}$, $\mathbf{u} \in \mathbb{R}^p$, $\mathbf{y} = \mathbf{X}\mathbf{u}$
- $\sigma^2_{\mathbf{u}} = \mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{u}$
- $\mathbf{A}_1, \mathbf{A}_2 \in \mathbb{R}^{n \times n}$ matrices simétricas $\mathrm{rg}(\mathbf{A}_1) = r$, $\mathrm{rg}(\mathbf{A}_2) = s$
- $U_\ell = \dfrac{\mathbf{y}^{\mathrm{T}}\mathbf{A}_\ell\mathbf{y}}{\sigma^2_{\mathbf{u}}}$, $\ell = 1, 2$,
- $\mathbf{b} \neq \mathbf{0}$, $\mathbf{b} \in \mathbb{R}^n$

Entonces,

a) $\mathbf{X}^{\mathrm{T}}\mathbf{A}_1\mathbf{X} \sim \mathcal{W}(\mathbf{\Sigma}, p, r) \iff U_1 \sim \chi^2_r$ para cualquier $\mathbf{u} \neq \mathbf{0}$

b) $\mathbf{X}^{\mathrm{T}}\mathbf{A}_1\mathbf{X} \sim \mathcal{W}(\mathbf{\Sigma}, p, r)$ y $\mathbf{X}^{\mathrm{T}}\mathbf{A}_2\mathbf{X} \sim \mathcal{W}(\mathbf{\Sigma}, p, s)$ independientes entre sí
$\iff$
U_1 y U_2 son independientes tales que $U_1 \sim \chi^2_r$, $U_2 \sim \chi^2_s$, para cualquier $\mathbf{u} \neq \mathbf{0}$.

c) $\mathbf{X}^{\mathrm{T}}\mathbf{b}$ y $\mathbf{X}^{\mathrm{T}}\mathbf{A}_1\mathbf{X}$ son independientes $\iff$
U_1 e $\mathbf{y}^{\mathrm{T}}\mathbf{b}$ son independientes y tales que $\mathbf{y}^{\mathrm{T}}\mathbf{b} \sim N(0, \|\mathbf{b}\|^2\sigma^2_{\mathbf{u}})$, $U_1 \sim \chi^2_r$ para cualquier $\mathbf{u} \neq \mathbf{0}$.

Corolarios

Sea $\mathbf{\Sigma} > 0$, $\mathbf{x}_1, \ldots \mathbf{x}_n$ independientes $\mathbf{x}_i \sim N_p(\mathbf{0}, \mathbf{\Sigma})$,
$\mathbf{X}^{\mathrm{T}} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$.

Corolario 1 Sea $\mathbf{A} \in \mathbb{R}^{n \times n}$ matriz simétrica

$$\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X} \sim \mathcal{W}(\mathbf{\Sigma}, p, r) \iff \mathbf{A}^2 = \mathbf{A} \quad \text{y} \quad \mathrm{rg}(\mathbf{A}) = r$$

Corolario 2 Sea $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ matrices simétricas, $\mathbf{b} \neq \mathbf{0}$, $\mathbf{b} \in \mathbb{R}^n$

a) $\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X} \sim \mathcal{W}(\mathbf{\Sigma}, p, r)$ y $\mathbf{X}^{\mathrm{T}}\mathbf{B}\mathbf{X} \sim \mathcal{W}(\mathbf{\Sigma}, p, s)$ son independientes entre sí $\iff \mathbf{A}\mathbf{B} = 0$

b) $\mathbf{X}^{\mathrm{T}}\mathbf{A}\mathbf{X} \sim \mathcal{W}(\mathbf{\Sigma}, p, r)$ y $\mathbf{X}^{\mathrm{T}}\mathbf{b} \sim N_p(\mathbf{0}, \|\mathbf{b}\|^2\mathbf{\Sigma})$ independientes entre sí $\iff \mathbf{A}\mathbf{b} = 0$, $\mathbf{A}^2 = \mathbf{A}$ y $\mathrm{rg}(\mathbf{A}) = r$

Definición

Hotelling central

Sean $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{W} \sim \mathcal{W}(\boldsymbol{\Sigma}, p, m)$ independientes entre sí, al estadístico

$$T^2_{p,m} = m\,(\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{W}^{-1} (\mathbf{x} - \boldsymbol{\mu})\,,$$

se lo llama estadístico de Hotelling central.

Hotelling no central

Sean $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{W} \sim \mathcal{W}(\boldsymbol{\Sigma}, p, m)$ independientes entre sí, al estadístico

$$T^2_{p,m} = m\,\mathbf{x}^{\mathrm{T}} \mathbf{W}^{-1} \mathbf{x}\,,$$

se lo llama estadístico de Hotelling no central.

Teorema 1. Sean $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{W} \sim \mathcal{W}(\boldsymbol{\Sigma}, p, m)$ independientes entre sí.

a) El estadístico $T^2_{p,m} = m\,(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{W}^{-1}(\mathbf{x}-\boldsymbol{\mu})$, tiene distribución dada por

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

b) El estadístico $T^2_{p,m}(\lambda^2) = m\,\mathbf{x}^{\mathrm{T}}\mathbf{W}^{-1}\mathbf{x}$, tiene distribución dada por

$$\frac{m-p+1}{p}\frac{T^2_{p,m}}{m} \sim \mathcal{F}_{p,m-p+1}(\lambda^2) \quad \text{con} \quad \lambda^2 = \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$$

Estimación

Sean $\mathbf{x}_1, \ldots, \mathbf{x}_n$ 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.
- $\mathbf{Q} = \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}}$ y $\overline{\mathbf{x}}$ son estadísticos suficientes y completos.
- Por lo tanto, cualquier estimador insesgado basado en $\mathbf{Q}$ y $\overline{\mathbf{x}}$ resulta IMVU.

Propiedad

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

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

tenemos que

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

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

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

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

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

Distribución de los estimadores

Sean $\mathbf{x}_1, \ldots, \mathbf{x}_n$ 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) $\overline{\mathbf{x}} \sim N_p(\boldsymbol{\mu}, (1/n)\boldsymbol{\Sigma})$,
$\mathbf{Q} = \sum_{i=1}^{n} (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}} \sim \mathcal{W}(\boldsymbol{\Sigma}, p, n-1)$.

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

c)

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

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

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

El método de componentes principales busca elegir q combinaciones lineales

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

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

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

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

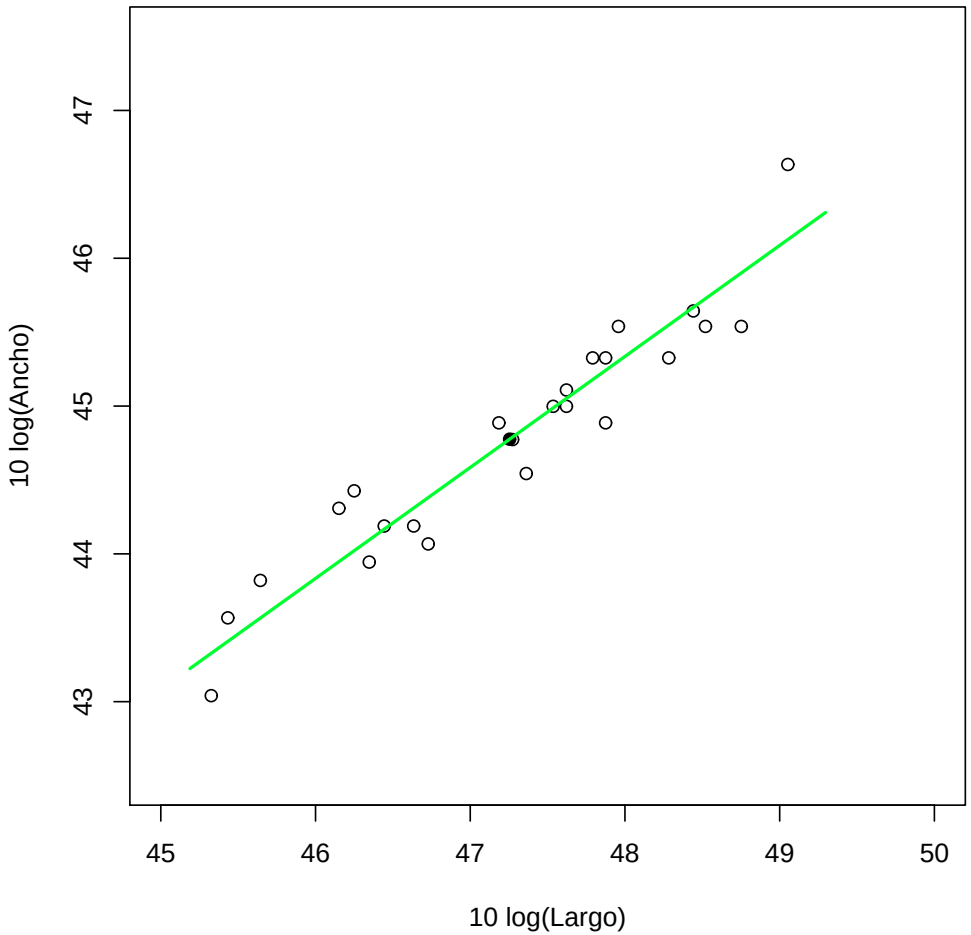
Se estudiaron 24 machos y 24 hembras.

$$\overline{\mathbf{x}}_{\mathrm{M}} = \begin{pmatrix} 47.254 \\ 44.776 \end{pmatrix} \qquad \overline{\mathbf{x}}_{\mathrm{H}} = \begin{pmatrix} 49.004 \\ 46.229 \end{pmatrix}$$

y

$$\widehat{\gamma}_{\mathrm{M}} = (0.7996, 0.6005)^{\mathrm{T}} \qquad \widehat{\gamma}_{\mathrm{H}} = (0.7892, 0.6141)^{\mathrm{T}}$$

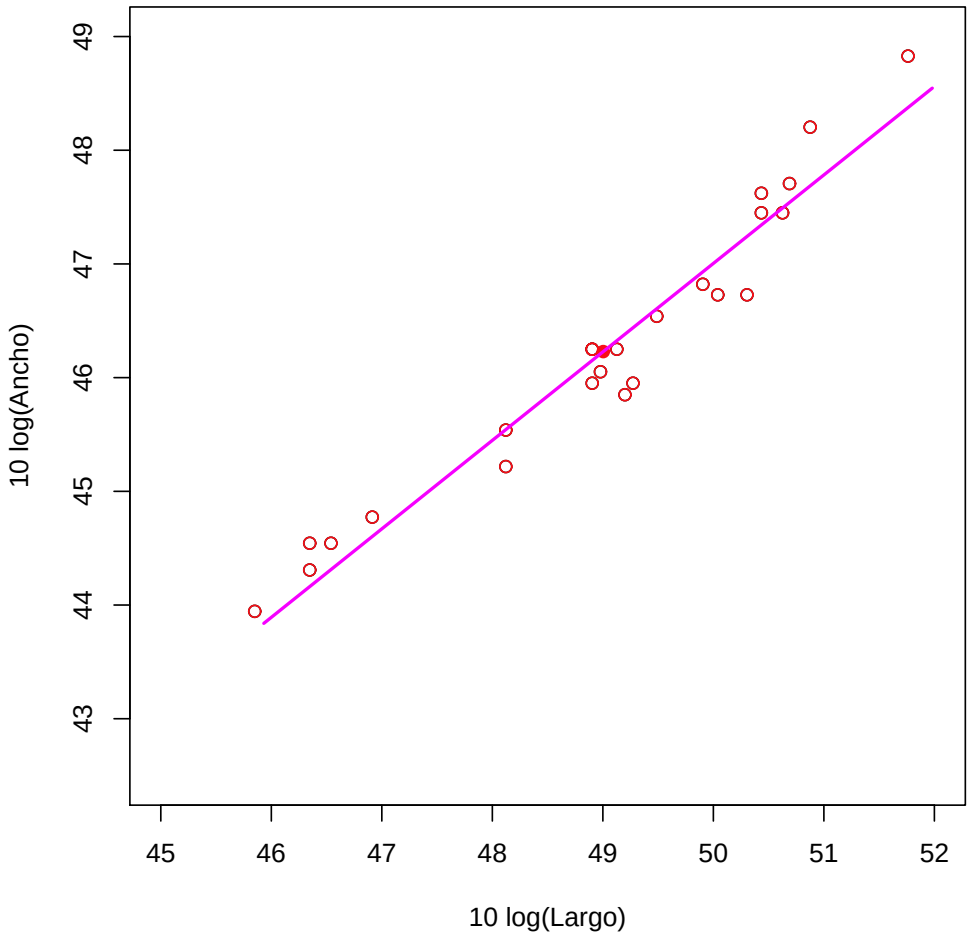
Machos



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


10 log(Ancho)
43
44
45
46
47
48
49
45
46
47
48
49
50
51
52
10 log(Largo)

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

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

Sean

- $\lambda_1 \geq \cdots \geq \lambda_p$ los autovalores de $\boldsymbol{\Sigma}$
- $\gamma_1, \ldots, \gamma_p$ los autovectores de $\boldsymbol{\Sigma}$ asociados a $\lambda_1 \geq \cdots \geq \lambda_p$
- $\boldsymbol{\Gamma} = (\gamma_1, \ldots, \gamma_p)$, $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}} = \mathbf{I}_p$
- $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_p)$

$$\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Gamma} = \boldsymbol{\Lambda}, \qquad \boldsymbol{\Sigma} = \sum_{j=1}^{p} \lambda_j \, \gamma_j \gamma_j^{\mathrm{T}}$$

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

$$\mathbf{x} = \boldsymbol{\mu} + \sum_{j=1}^{p} \gamma_j^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu}) \, \gamma_j$$

Definición

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

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

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

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

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

Propiedad 1. La componentes principales $v_1, \ldots, v_p$ son no correlacionadas y $\mathrm{VAR}(v_j) = \lambda_j$, o sea,

$$\mathrm{VAR}(\mathbf{v}) = \mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$$

Lemas previos

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

a) $\sup\limits_{\mathbf{u}\neq\mathbf{0}} \dfrac{\mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{u}}{\mathbf{u}^{\mathrm{T}}\mathbf{u}} = \lambda_1$ y el supremo se alcanza en γ_1.

b) $\inf\limits_{\mathbf{u}\neq\mathbf{0}} \dfrac{\mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{u}}{\mathbf{u}^{\mathrm{T}}\mathbf{u}} = \lambda_p$ y el infimo se alcanza en γ_p.

c) $\sup\limits_{\substack{\mathbf{u}\neq\mathbf{0} \\ \mathbf{u}^{\mathrm{T}}\gamma_i=0 \ 1\leq i\leq k}} \dfrac{\mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{u}}{\mathbf{u}^{\mathrm{T}}\mathbf{u}} = \lambda_{k+1}$ y el supremo se alcanza en γ_{k+1}.

Lemas previos

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

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

Entonces

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

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

Lemas previos

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

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

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

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

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

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

- $\lambda_1 \geq \cdots \geq \lambda_p$ los autovalores de $\mathbf{\Sigma}$ y
- $\gamma_1, \ldots, \gamma_p$ los autovectores de $\mathbf{\Sigma}$ asociados a $\lambda_1 \geq \cdots \geq \lambda_p$.
- $\mathcal{H}_0$ el subespacio generado por $\gamma_1, \ldots, \gamma_q$ 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

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

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

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

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

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

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

Propiedades de optimalidad

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

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

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

a) $\max_{\|\mathbf{a}\|=1} \text{VAR}(\mathbf{a}^{\mathrm{T}}\mathbf{x}) = \text{VAR}(v_1)$, o sea, el máximo se alcanza en γ_1.

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

es decir, el máximo se alcanza en γ_{k+1}.
La condición $\text{COV}(\mathbf{a}^{\mathrm{T}}\mathbf{x}, v_j) = 0$ asegura que no se repite información.

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

Propiedades de optimalidad

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

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

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

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

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

Propiedades de optimalidad

Propiedad 4. Luego,

$$\textsc{Var}(\mathbf{a}_j^{\mathrm{T}}\mathbf{x}) = \mathbf{a}_j^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{a}_j$$

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

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

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

Entonces, se cumple que

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

Propiedades de optimalidad

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

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

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 $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

Correlaciones

Supongamos que $\mathbb{E}\mathbf{x} = \boldsymbol{\mu}$ y $\textsc{Var}(\mathbf{x}) = \boldsymbol{\Sigma}$. Sea $\boldsymbol{\gamma}_\ell = (\gamma_{\ell,1}, \ldots, \gamma_{\ell,p})^{\mathrm{T}}$

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

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

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

$$\widetilde{\mathbf{x}} = \boldsymbol{\mu} + \textsc{Cov}(\mathbf{x}, \mathbf{v}_q)\{\textsc{Var}(\mathbf{v}_q)\}^{-1}\mathbf{v}_q = \boldsymbol{\mu} + \boldsymbol{\Gamma}_q \mathbf{v}_q$$

y el residuo es $\mathbf{u} = \mathbf{x} - \widetilde{\mathbf{x}}$.

Correlaciones

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

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

o, sea,

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

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

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

Componentes principales muestrales

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

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

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

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

y además,

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

Componentes principales muestrales: $\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Gamma}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Gamma}}^{\mathrm{T}}$

- $\widehat{\lambda}_1 > \cdots > \widehat{\lambda}_p$ los autovalores de $\widehat{\mathbf{\Sigma}}$ y
- $\widehat{\gamma}_1, \ldots, \widehat{\gamma}_p$ los autovectores de $\widehat{\mathbf{\Sigma}}$ asociados a $\widehat{\lambda}_1 > \cdots > \widehat{\lambda}_p$.

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

$$\widehat{\mathbf{v}}_i = \widehat{\mathbf{\Gamma}}^{\mathrm{T}}(\mathbf{x}_i - \overline{\mathbf{x}})$$

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

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

$$\widehat{v}_j = \widehat{\gamma}_j^{\mathrm{T}}(\mathbf{x}_i - \overline{\mathbf{x}}),$$

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

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*

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

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

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

$$\widehat{\mathbf{\Lambda}} = \mathrm{diag}(1305.4337, 651.5147, 123.2253, 75.9081, 27.8237, 13.2150, 5.7182, 1.1248)$$

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

Ejemplo *Microtus multiplex*

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

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

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

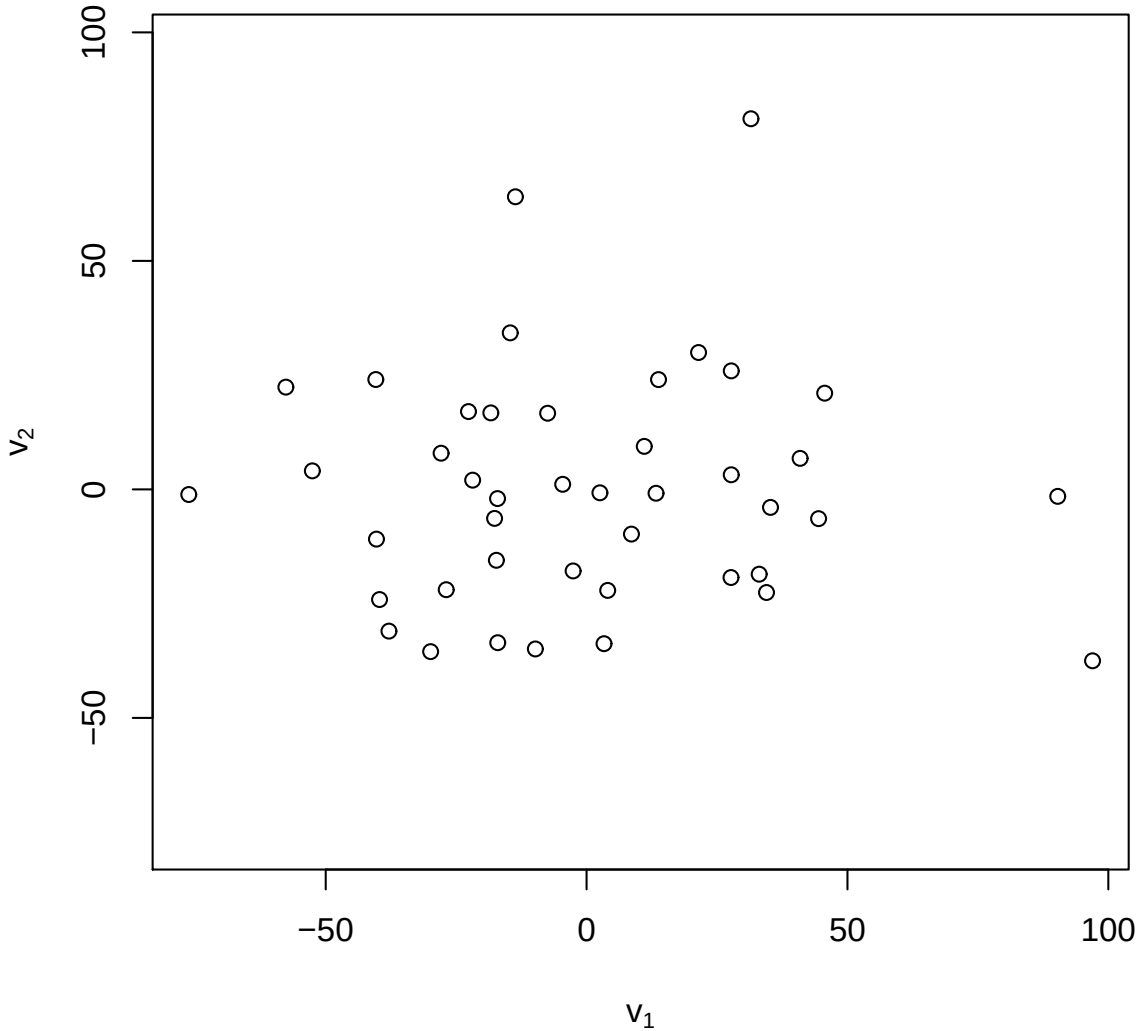
Luego, los desvíos estandar estimados de los autovalores $s_{\widehat{\lambda}_j}$ dan

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

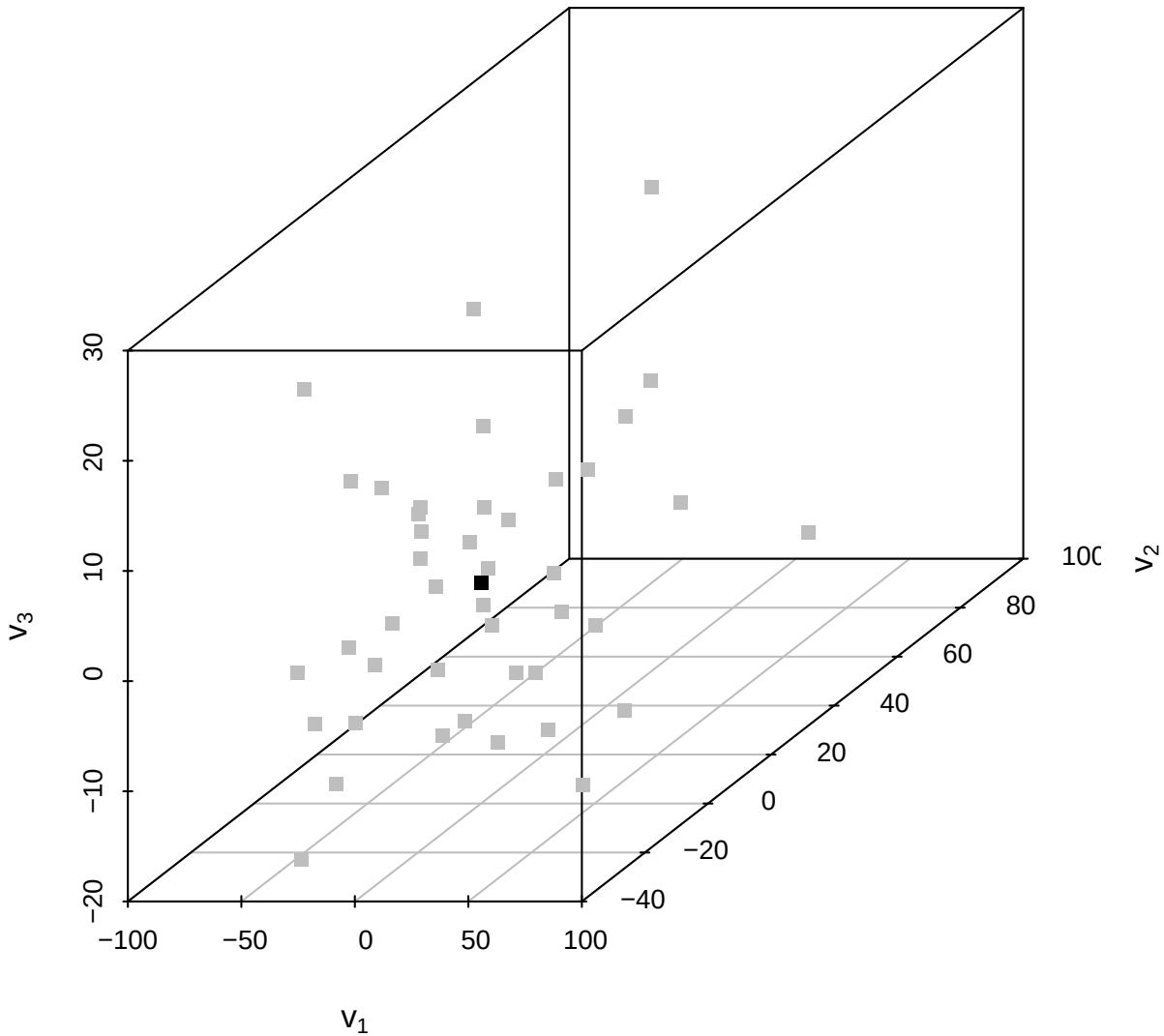
Es decir,

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

Observemos que las coordenadas de $\widehat{\gamma}_1$ 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 $H_{0,(r,h)} : \lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_{r+h}$ versus $H_{1,(r,h)} : \lambda_{r+1} > \lambda_{r+2} > \cdots > \lambda_{r+h}$ El test del cociente de máxima verosimilitud se basa en

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

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 $H_{0,(2,2)} : \lambda_3 = \lambda_4$

$$\begin{aligned} M_{1,2} &= 0.5350 & -n\log(M_{1,2}) &= 26.8942 \\ M_{2,2} &= 0.9435 & -n\log(M_{2,2}) &= 2.4991 \\ \chi^2_{2,0.05} &= 5.9915 & \chi^2_{2,0.01} &= 9.2103 \end{aligned}$$

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 $(\widehat{\lambda}_1 + \widehat{\lambda}_2)/\sum_{j=1}^{p} \widehat{\lambda}_j = 0.8879$ sugeriría que la variabilidad en las mandíbulas de los roedores estudiados podría ser adecuadamente descripta por las dos primeras componentes principales.

Inferencia en el caso normal

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

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

Además,

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

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

Normal Multivariada Singular

Definición. Sea $\mathbf{\Sigma}$ simétrica definida no–negativa. Se dice que $\mathbf{x} \sim N(\mathbf{0}, \mathbf{\Sigma})$ si su función característica es

$$\varphi(\mathbf{u}) = \exp\{-\frac{1}{2}\mathbf{u}^{\mathrm{T}}\mathbf{\Sigma}\mathbf{u}\}$$

Propiedad. Si $\mathbf{x} \sim N(\mathbf{0}, \mathbf{\Sigma})$ y rango$(\mathbf{\Sigma}) = q < p$ entonces

- $v_1, \ldots v_q$ son independientes
- $v_j \sim N(0, \lambda_j)$, $1 \leq j \leq q$
- para $j \geq q + 1$, $\mathbb{P}(v_j = 0) = 1$.

Normal Multivariada Singular: Ejemplo

Sea $\mathbf{x} \sim N(\mathbf{0}, \mathbf{\Sigma})$ con

$$\mathbf{\Sigma} = (1-\rho)\mathbf{I}_p + \rho \mathbf{1}_p \mathbf{1}_p^{\mathrm{T}} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}$$

Tomemos $\rho = -1/(p-1)$, de forma que $\mathbf{\Sigma}$ es singular.

- Los autovalores de $\mathbf{\Sigma}$ son $(1-\rho)$ con multiplicidad $p-1$ y $1+(p-1)\rho = 0$ con multiplicidad 1.
- El autovector asociado a $\lambda_p = 0$ es $\gamma_p = (1/\sqrt{p})\mathbf{1}_p$
- Cualquier conjunto de $p-1$ vectores ortogonales a $\mathbf{1}_p$ se pueden tomar como los autovectores asociados a $(1-\rho)$.

Normal Multivariada Singular: Ejemplo

Podemos tomar entonces como componentes principales

$$
\begin{aligned}
v_1 &= \frac{x_1 - x_2}{\sqrt{2}} \\
v_2 &= \frac{x_1 + x_2 - 2x_3}{\sqrt{2 \times 3}} \\
&\vdots \qquad\quad \vdots \\
v_{p-1} &= \frac{x_1 + \cdots + x_{p-1} - (p-1)x_p}{\sqrt{p(p-1)}} \\
v_p &= \frac{x_1 + \cdots + x_p}{\sqrt{p}}
\end{aligned}
$$

Normal Multivariada Singular: Ejemplo

- $v_1, \ldots, v_{p-1}$ son i.i.d. tales que $v_j \sim N\left(0, 1-\rho = \frac{p}{p-1}\right)$
- $\mathbb{P}(v_p = 0) = 1$

De esta forma, se obtiene por ejemplo, que

$$\begin{aligned}\mathbb{P}(3x_1 + x_2 - x_3 + x_4 + \cdots + x_p > 0) &= \mathbb{P}(\sqrt{p}v_p + \sqrt{6}v_2 + \sqrt{2}v_1 > 0) \\ &= \mathbb{P}(\sqrt{3}v_2 + v_1 > 0) = \frac{1}{2}\end{aligned}$$

pues

$$\sqrt{3}v_2 + v_1 \sim N\left(0, \frac{4(p-2)}{p-1}\right)$$