Hasta ahora se han estudiado las componentes principales para **un vector** aleatorio $\mathbb{X}$ cuando sus primeros dos momentos existen y asumiendo $\Sigma > 0$. ¿Cómo se pueden construir estimadores de $\mathbb{Y}_1, \ldots, \mathbb{Y}_p$ usando los datos $\mathbb{x}_1, \ldots, \mathbb{x}_n$?

En la práctica se sugiere reemplazar las características poblacionales $\mu\ (= \mathbb{E}(\mathbb{X}))$ y $\Sigma\ (= VAR(\mathbb{X}))$ por sus estimadores muestrales

$$\hat{\mu} \equiv \frac{1}{n}\sum_{i=1}^{n} \mathbb{x}_i = \overline{\mathbb{x}}_{(n)} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1} \\ \frac{1}{n}\sum_{i=1}^{n} x_{i2} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{ip} \end{pmatrix}$$

$\overline{x}_{(n)}$ has j-th row given by

$$\overline{x}_{\cdot j} \equiv \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

el promedio (sobre todos los individuos) de la j-ésima variable.

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \underbrace{(\mathbb{x}_i - \overline{\mathbb{x}}_{(n)})}_{p\times 1}\underbrace{(\mathbb{x}_i - \overline{x}_{(n)})'}_{1\times p}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} x_{i1} - \overline{x}_{\cdot 1} \\ x_{i2} - \overline{x}_{\cdot 2} \\ \vdots \\ x_{ip} - \overline{x}_{\cdot p} \end{pmatrix} \left(x_{i1} - \overline{x}_{\cdot 1}, x_{i2} - \overline{x}_{\cdot 2}, \ldots, x_{ip} - \overline{x}_{\cdot p}\right)$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} (x_{i1}-\bar{x}_{\cdot 1})^2 & (x_{i1}-\bar{x}_{\cdot 1})(x_{i2}-\bar{x}_{\cdot 2}) & \cdots & (x_{i1}-\bar{x}_{\cdot 1})(x_{ip}-\bar{x}_{\cdot p}) \\ (x_{i2}-\bar{x}_{\cdot 2})(x_{i1}-\bar{x}_{\cdot 1}) & (x_{i2}-\bar{x}_{\cdot 2})^2 & \cdots & (x_{i2}-\bar{x}_{\cdot 2})(x_{ip}-\bar{x}_{\cdot p}) \\ \vdots & & \ddots & \vdots \\ (x_{ip}-\bar{x}_{\cdot p})(x_{i1}-\bar{x}_{\cdot 1}) & \cdots & & (x_{ip}-\bar{x}_{\cdot p})^2 \end{pmatrix}$$

$\hat{\Sigma}$ tiene entrada j,k dada por $\longrightarrow$

$$\hat{\Sigma}_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij}-\bar{x}_{\cdot j})(x_{ik}-\bar{x}_{\cdot k}) \qquad \begin{array}{l} j=1,2,\ldots,p, \\ k=1,2,\ldots,p. \end{array}$$

<u>Nota</u>: En algunos referencias se considera la "corrección de Bessel", definiendo $\hat{\Sigma}$ como la matriz de dimensiones $p\times p$ con entrada j,k dada por:

$$\hat{\Sigma}^{(c)}_{jk} = \frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_{\cdot j})(x_{ik}-\bar{x}_{\cdot k}) \cdots (A)$$

Esta corrección tiene el fin de hacer de de (A) un estimador insesgado para $\operatorname{cov}(X_j, X_k)$. $\left(\hat{\Sigma}^{(c)} = \frac{n}{n-1}\hat{\Sigma} \right.$ es estimador insesgado para $\left. \Sigma \right)$

Sea $\mathcal{X}$ la matriz de datos dada por:

$$\mathcal{X} \equiv \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \quad \text{de dimensiones } n \times p.$$

$\mathcal{X}$ contiene renglones dados por los individuos de la muestra (el i-ésimo renglón de $\mathcal{X}$ es x_i').

$$\mathbb{1}_n \equiv \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}_{n \times 1}.$$

$$\overline{x}_{(n)} = \frac{1}{n} \mathcal{X}' \mathbb{1}_n.$$

(i)

$$\boxed{\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \overline{x}_{\cdot j})(x_{ik} - \overline{x}_{\cdot k}) \\ &= \frac{1}{n}\left(\sum_{i=1}^{n} x_{ij} x_{ik}\right) - \overline{x}_{\cdot j}\, \overline{x}_{\cdot k} \\ &= \text{entrada } j,k \text{ en } \hat{\Sigma} \end{aligned}}$$

Con estas notaciones, podemos escribir

$$(B) \ldots \quad \hat{\Sigma} \overset{(i)}{=} \frac{1}{n} \mathcal{X}'\mathcal{X} - \overline{x}_{(n)} \overline{x}_{(n)}' = \frac{1}{n}\left(\mathcal{X}'\mathcal{X} - \frac{1}{n} \mathcal{X}' \mathbb{1}_n \mathbb{1}_n' \mathcal{X}\right).$$

Usando la "matriz de centrado"

$$\mathcal{H} = \mathbb{I}_n - \frac{1}{n} \mathbb{1}_n \mathbb{1}_n', \qquad (\mathbb{I}_n \equiv \text{matriz identidad } n \times n)$$

también podemos escribir

$$(C) \ldots \quad \hat{\Sigma} = \frac{1}{n} \mathcal{X}' \mathcal{H} \mathcal{X}.$$

Claramente $\mathcal{H}$ es simétrica:

$$\mathcal{H}' = \left(\mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)' = \mathbb{I}_n' - \frac{1}{n}\left(\mathbb{1}_n\mathbb{1}_n'\right)'$$

$$= \mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' = \mathcal{H},$$

e idempotente:

$$\mathcal{H}\mathcal{H} = \left(\mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)\left(\mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)$$

$$= \mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' + \left(\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)\left(\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right).$$

Pero $\left(\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right)\left(\frac{1}{n}\mathbb{1}_n\mathbb{1}_n'\right) = \frac{1}{n^2}\left(\mathbb{1}_n\mathbb{1}_n'\right)\left(\mathbb{1}_n\mathbb{1}_n'\right)$

$$= \frac{1}{n^2}\underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}}_{n\times n}\underbrace{\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix}}_{n\times n} = \frac{1}{n^2}\begin{pmatrix} n & n & \cdots & n \\ n & n & \cdots & n \\ \vdots & \vdots & & \vdots \\ n & n & & n \end{pmatrix}$$

$$= \frac{1}{n}\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} = \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'$$

luego

$$\mathcal{H}\mathcal{H} = \mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' + \frac{1}{n}\mathbb{1}_n\mathbb{1}_n'$$

$$= \mathbb{I}_n - \frac{1}{n}\mathbb{1}_n\mathbb{1}_n' = \mathcal{H}$$

Usaremos la ecuación (C) y las propiedades de $\mathcal{H}$ para ver que $\hat{\Sigma}$ es positiva semi-definida:

Sea $a \in \mathbb{R}^p$, tal que $a \neq 0$, entonces por (C)

$$a'\hat{\Sigma}a = \frac{1}{n} a'\mathcal{X}'\mathcal{H}\mathcal{X}a = \frac{1}{n} a'\mathcal{X}'\mathcal{H}^2\mathcal{X}a$$

$$= \frac{1}{n} a'\mathcal{X}'\mathcal{H}'\mathcal{H}\mathcal{X}a = \frac{1}{n} b'b,$$

$\mathcal{H} = \mathcal{H}\mathcal{H}$
$= \mathcal{H}'\mathcal{H}$

donde $b \equiv \mathcal{H}\mathcal{X}a$ es un vector de dimensiones $n \times 1$.

Pero como $b'b = \sum_{i=1}^{n} b_i^2 \geq 0$, entonces

$$a'\hat{\Sigma}a \geq 0 \quad .$$

Usaremos $\hat{\mu}$ y $\hat{\Sigma}$ para calcular estimadores de $\gamma_1, \ldots, \gamma_p$. Para este fin, usaremos la descomposición de Jordan para $\hat{\Sigma}$

$$\hat{\Sigma} = \mathcal{G}\mathcal{L}\mathcal{G}' \quad \ldots\ldots (D)$$