$$\mathcal{X}_s = \begin{pmatrix} \frac{(x_{11}-\bar{x}_{\cdot 1})}{S_{x_1x_1}^{1/2}} & \frac{(x_{12}-\bar{x}_{\cdot 2})}{S_{x_2x_2}^{1/2}} & \cdots & \frac{x_{1p}-\bar{x}_{\cdot p}}{S_{x_px_p}^{1/2}} \\ \vdots & \vdots & & \vdots \\ \frac{(x_{n1}-\bar{x}_{\cdot 1})}{S_{x_1x_1}^{1/2}} & \frac{(x_{n2}-\bar{x}_{\cdot 2})}{S_{x_2x_2}^{1/2}} & \cdots & \frac{x_{np}-\bar{x}_{\cdot p}}{S_{x_px_p}^{1/2}} \end{pmatrix}$$

↑

"matriz de datos estandarizados"

$$\bar{\mathbf{x}}^s_{(n)} \equiv \frac{1}{n}\mathcal{X}_s' \mathbb{1}_n = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^n (x_{i1}-\bar{x}_{\cdot 1}) \\ \vdots \\ \sum_{i=1}^n (x_{ip}-\bar{x}_{\cdot p}) \end{pmatrix}$$

$$= \begin{pmatrix} \bar{x}_{\cdot 1}-\bar{x}_{\cdot 1} \\ \vdots \\ \bar{x}_{\cdot p}-\bar{x}_{\cdot p} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}}_{\leftarrow p\times 1} \quad \cdots\cdots\cdots \text{(sm)}$$

$$\therefore \quad \hat{\Sigma}^s = \frac{1}{n}\mathcal{X}_s'\mathcal{X}_s - \bar{\mathbf{x}}^s_{(n)}\bar{\mathbf{x}}^{s\,\prime}_{(n)} = \frac{1}{n}\mathcal{X}_s'\mathcal{X}_s$$

es la matriz con entrada k,l.

$$\hat{\Sigma}^s_{kl} = \frac{1}{n}\sum_{i=1}^n \frac{(x_{ik}-\bar{x}_{\cdot k})(x_{il}-\bar{x}_{\cdot l})}{\sqrt{S_{x_kx_k}}\sqrt{S_{x_lx_l}}} \quad \left(= \hat{\rho}_{x_kx_l} \right)$$

$$\hat{\Sigma}^s = \hat{\mathcal{R}} \leftarrow \text{matriz de correlaciones de } \mathcal{X}$$

$$\hat{\mathcal{R}} = \mathcal{G}_{\hat{\mathcal{R}}}\mathcal{L}_{\hat{\mathcal{R}}}\mathcal{G}_{\hat{\mathcal{R}}}' \quad \text{descomposición de Jordan de } \hat{\mathcal{R}}$$

$\mathbb{L}_{\hat{\mathcal{R}}} = \mathrm{DIAG}(l_1^{\hat{\mathcal{R}}}, \ldots, l_p^{\hat{\mathcal{R}}})$, $l_1^{\hat{\mathcal{R}}} \geq \cdots \geq l_p^{\hat{\mathcal{R}}}$ son

los valores propios de $\hat{\mathcal{R}}$, con correspondientes

vectores propios $g_1^{\hat{\mathcal{R}}}, \ldots, g_p^{\hat{\mathcal{R}}}$. Como

$tr(\hat{\mathcal{R}}) = \sum_{j=1}^{p} l_j^{\hat{\mathcal{R}}}$ y la matriz $\hat{\mathcal{R}}$ tiene

diagonal dada por $\mathbb{1}_p = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, entonces

$$\sum_{j=1}^{p} l_j^{\hat{\mathcal{R}}} = p .$$

Las componentes principales normalizadas

son entonces (dado que $\overline{x}_{(n)}^{S} = 0$)

$$\mathbb{Z} = \mathcal{X}_S \, \mathbb{G}_{\hat{\mathcal{R}}} \qquad \ldots \text{(CPN)}$$

(compárese con la ecuación (CP), página 35).

$$\underset{n \times p}{\mathbb{Z}} = (z_1, \ldots, z_p) \leftarrow z_j$$

la columna j tiene n observaciones de la j-ésima componente principal normalizada

compárese con $\mathbb{Y}$ en la ecuación (CP) página 35.

ejercicio: verifique que

(i) $\bar{\mathbb{Z}}'_{(n)} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}_{p\times 1}$

(ii) $\hat{\Sigma}_{\mathbb{Z}} = \mathcal{G}'_{\hat{\mathcal{R}}} \hat{\Sigma}^{s} \mathcal{G}_{\hat{\mathcal{R}}}$
$= \mathcal{G}'_{\hat{\mathcal{R}}} \hat{\mathcal{R}} \mathcal{G}_{\hat{\mathcal{R}}}$

A nivel poblacional las componentes principales normalizadas estarían dadas por

$$\mathbb{Z} = \Gamma'_{\mathcal{R}} \mathcal{D}^{-1/2}(\mathbb{X} - \mathbb{M}),$$ (← dims $p\times 1$)

en donde $\mathcal{R} = \Gamma_{\mathcal{R}} \Lambda_{\mathcal{R}} \Gamma'_{\mathcal{R}}$ es la matriz de correlaciones de $\mathbb{X}$ con entrada i,j $\rho_{ij} = corr(X_i, X_j)$; $i,j = 1,2,\dots,p$, $\mathcal{D}$ es una matriz diagonal y el elemento i en la diagonal es $VAR(X_i)$; $i = 1,2,\dots,p$.

$\mathbb{X}_s \equiv \mathcal{D}^{-1/2}(\mathbb{X} - \mathbb{M})$ es el vector aleatorio estandarizado con componente i dada por $\frac{X_i - \mu_i}{\sqrt{VAR(X_i)}}$; $i = 1,2,\dots,p$.

Para estudiar la correlación entre X_i y Z_j

$$\mathrm{cov}(\mathbb{X}_s, \mathbb{Z}_1) = \mathrm{cov}\left[\mathbb{D}^{-1/2}(\mathbb{X}-\mathbb{M}),\ \mathbb{\Gamma}_R' \mathbb{D}^{-1/2}(\mathbb{X}-\mathbb{M})\right]$$

$$= \mathbb{D}^{-1/2}\,\mathrm{cov}\left[(\mathbb{X}-\mathbb{M}),(\mathbb{X}-\mathbb{M})\right]\mathbb{D}^{-1/2}\,\mathbb{\Gamma}_R$$

$$= \mathbb{D}^{-1/2}\,\Sigma\,\mathbb{D}^{-1/2}\,\mathbb{\Gamma}_R$$

$$= \mathbb{R}\,\mathbb{\Gamma}_R = \mathbb{\Gamma}_R \Lambda_R \mathbb{\Gamma}_R' \mathbb{\Gamma}_R = \mathbb{\Gamma}_R \Lambda_R .$$

A nivel muestral el análogo estaría dado por

$$(\mathrm{A}) \cdots\cdots \widehat{\mathrm{cov}(\mathbb{X}_s, \mathbb{Z})} \overset{(a)}{=} \frac{1}{n}\mathcal{X}_s' \mathcal{Z} = \frac{1}{n}\mathcal{X}_s'\mathcal{X}_s \mathcal{G}_{\hat{R}}$$

$$= \hat{\mathbb{R}}\,\mathcal{G}_{\hat{R}} = \mathcal{G}_{\hat{R}} \mathcal{L}_{\hat{R}} \mathcal{G}_{\hat{R}}' \mathcal{G}_{\hat{R}}$$

$$= \mathcal{G}_{\hat{R}} \mathcal{L}_{\hat{R}} . \qquad (\text{dims } p\times p)$$

La ecuación (A) tiene una matriz cuya entrada i,j vale $\widehat{\mathrm{cov}(X_i^s, Z_j)}$, si de ahi quisieramos obtener $\widehat{\mathrm{corr}(X_i^s, Z_j)}$ tenemos que dividir cada entrada por

(a) ya que $\overline{\mathbb{Z}}_{(n)}' = \mathbb{O}$ (tarea i) y además de la ecuación (sm) $\overline{\mathbb{X}}_{(n)}^s = \mathbb{O}$ página 69