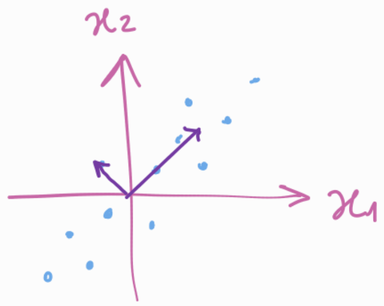
$$X = \begin{array}{c|ccc} & x_1 & & x_p \\ \hline x_1 & x_{11} & & x_{1p} \\ \vdots & & & \\ x_n & x_{n1} & & x_{np} \end{array}$$

Filas → individuos (n)

Columnas → variables (p)

Ej: $p = 2$



ANÁLISIS DE COMPONENTES PRINCIPALES

1) Estandarizar columnas de X.

$$\mu_1 = \frac{1}{n}\sum_{i=1}^{n} x_{i1}$$

$$\sigma_1^2 = \frac{1}{n-1}\sum (x_{i1} - \mu_1)^2 \quad (*)$$

$$Y = \begin{pmatrix} \frac{x_{11}-\mu_1}{\sqrt{\sigma_1^2}} & \cdots & \frac{x_{1p}-\mu_p}{\sqrt{\sigma_p^2}} \\ \vdots & & \vdots \\ \frac{x_{n1}-\mu_1}{\sqrt{\sigma_1^2}} & & \frac{x_{np}-\mu_p}{\sqrt{\sigma_p^2}} \end{pmatrix}$$

Supongamos que $X = Y$ (columnas ya están estandarizadas)

(*) En Python por defecto es $\frac{1}{n}\sum (x_{i1} - \mu_1)^2$ ⇝ resultados ligeramente distintos.

2) Calcular matriz de covarianza: $\Sigma = \frac{1}{n-1} Y^T Y$

$\Sigma \in \mathbb{R}^{p \times p}$

3) Hallamos valores y vectores propios de Σ

$$\underbrace{\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p}_{\text{ordenados}} \qquad \underbrace{a_1, \dots, a_p}_{a_i \text{ vector propio asociado a } \lambda_i}$$

4) $\overset{\text{PC1}}{Z_1} = a_{11} x_1 + a_{21} x_2 + \dots + a_{p1} x_p$

$$a_1 = (a_{11}, a_{21}, \dots, a_{p1})$$

$\vdots$

$\underset{\text{PCi}}{Z_i} = a_{1i} x_1 + a_{2i} x_2 + \dots + a_{pi} x_p$.

Notar que $Z_1 = X a_1 \rightsquigarrow \begin{array}{c} x_1 \\ \vdots \\ x_n \end{array} \begin{pmatrix} x_{11} & \dots & x_{1p} \\ & & \\ & & \end{pmatrix} \begin{pmatrix} a_{11} \\ \vdots \\ a_{p1} \end{pmatrix}$

$\begin{pmatrix} \bigcirc \\ \vdots \\ \bigcirc \end{pmatrix}$

$a_{11} x_{11} + \dots + a_{p1} x_{1p} = \langle x_1, a_1 \rangle$
$= C_{11}$

$a_{11} x_{n1} + \dots + a_{p1} x_{np} = \langle x_n, a_1 \rangle$
$= C_{n1}$

En resumen:

$$Z = \begin{pmatrix} | & | & & | \\ z_1 & z_2 & \cdots & z_p \\ | & | & & | \end{pmatrix} = \underbrace{\begin{pmatrix} | & & | \\ x_1 & \cdots & x_p \\ | & & | \end{pmatrix}}_{X} \cdot \underbrace{\begin{pmatrix} | & & | \\ a_1 & \cdots & a_p \\ | & & | \end{pmatrix}}_{A}$$

$$= XA$$

y tenemos $\boxed{c_{ik} = \langle x_i, a_k \rangle}$

elemento ik de la matriz Z.

Contribución individuo x_i a la construcción del eje k:

$$ctr_k(i) = \frac{c_{ik}^2}{n \cdot \lambda_k}$$

CONTRIBUCIÓN VARIABLE x_j A LA CONSTRUCCIÓN DEL EJE K:

$$ctr_k(x_j) = a_{jk}^2$$

Recordar $a_j = (a_{1j}, \ldots, a_{pj})$

$$z_k = a_{1k}x_1 + \ldots + a_{pk}x_p$$

$$\downarrow$$

$$a_{jk}x_j$$

CALIDAD REPRESENTACIÓN INDIVIDUO x_i SOBRE EJE K

$$Cal_k(i) = \frac{c_{ik}^2}{n \cdot \|x_i\|^2}$$

CALIDAD REPRESENTACIÓN VARIABLE x_j SOBRE EJE K

$$Cal_k(x_j) = \frac{d_{jk}^2}{\|x_j\|^2}$$

donde $d_{jk} = Cor(z_k, x_j) = \sqrt{\lambda_k}\, a_{jk}$

REFERENCIAS:

- Diapositivas teórico (calidad y contribución en 30, 31, 38, 39)