Chap III -

Analyse en composantes principales (ACP)

I- Aperçu de la méthode

p variables quantitatives $X_1, X_2, \dots X_p$
n individus
idée de l'ACP : extraire de la liste des p variables initiales (+ ou - redondants), une liste de nouvelles variables (ou facteurs) sans aucune redondance.

a. La transformation.

Redondance = . corrélation
. covariance.

Transformation des variables pour éliminer les covariances ou les corrélations.
⇒ on va diagonaliser Σ ou R.
⇒ . une nouvelle base (base des vecteurs propres) → matrice U
. une nouvelle matrice diagonale Λ (matrice diagonale des valeurs propres).

base initiale = variables initiales.
base de vecteurs propres : facteurs (nouvelles variables).

Remarque 1 : On choisit des vecteurs propres unitaires ou normés, c'est à dire dont la norme vaut 1.

Remarque 2 : Les vecteurs propres sont orthogonaux 2 à 2.

On obtient finalement une matrice F appelée matrice des scores factoriels, matrice des coordonnées des individus dans la nouvelle base :

$$F = X^s U \text{ ou } X^c U$$

On a donc :

- X^0 ou X^c : matrice d'origine (individus par rapport aux variables)
- R ou Σ : matrice de corrélation ou de variance - covariance
- U : matrice des vecteurs propres normés → facteurs
- $F = X^s U$ ou $X^c U$: matrice des individus par rapport aux facteurs.

Remarque : F est centrée mais pas forcément réduite.

Qualité globale d'explication (qge) : mesure de la qualité d'information portée par un facteur.

b. L'Interprétation.

Il faut tenter :

- d'interpréter les facteurs (axes)
- d'interpréter la position des individus par rapport aux facteurs.

On va étudier les corrélations existantes entre les facteurs et les variables.

→ matrice des corrélations entre les variables et les facteurs : <u>matrice des saturations</u> S.

~~S = RU~~

$$S = RU\Lambda^{-1/2} \quad \text{où} \quad \Lambda^{-1/2} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & & 0 \\ & \frac{1}{\sqrt{\lambda_1}} & \\ 0 & & \ddots \end{bmatrix}$$

λ_i = vp de R

↳ λ_i = variance du vecteur propre associé

$\sqrt{\lambda_i}$: écart type du vecteur propre associé.

or $R = \frac{1}{n} {}^tX^s X^s$

$$S = \frac{1}{n} {}^tX^s U \lambda^{-1/2} = \frac{1}{n} X^s F \lambda^{-1/2}$$

II- ACP d'un exemple.

matrice des données $A = \begin{bmatrix} 3 & 1{,}2 & 5 \\ 4 & 2 & 6 \\ 3 & 1{,}8 & 6 \end{bmatrix}$

(lait) (eau) (huile)

. <u>Première étape</u> : Déterminer A^c ou A^s.

analyse centrée et réduite.

. Calcul des moyennes :

$$\overline{X} = \frac{10}{3} \quad \overline{Y} = \frac{5}{3} \quad \overline{Z} = \frac{17}{3}$$

. Calcul des écart-types.

$$\sigma_x = \frac{\sqrt{2}}{3} \quad \sigma_y = \frac{2\sqrt{19}}{15\sqrt{3}} \quad \sigma_z = \frac{\sqrt{2}}{3}.$$

$$\Rightarrow \quad A^s = \begin{bmatrix} \frac{x_1 - \overline{X}}{\sigma_x} & \frac{y_1 - \overline{Y}}{\sigma_y} & \cdots \\ \frac{x_2 - \overline{X}}{\sigma_x} & \cdots & \cdot \\ \frac{x_3 - \overline{X}}{\sigma_x} & \cdots & \cdot \end{bmatrix}$$

$$A^s = \begin{bmatrix} -0,707 & -1,373 & -1,414 \\ 1,414 & 0,979 & 0,707 \\ -0,707 & 0,391 & 0,707 \end{bmatrix}$$

<u>Vérification</u> : A^s est centrée.

. Deuxième étape : Déterminer R.

$$R = \frac{1}{n} {}^tA^s . A^s = \begin{bmatrix} 1 & 0,693 & 0,5 \\ 0,693 & 1 & 0,925 \\ 0,5 & 0,925 & 1 \end{bmatrix}$$

<u>Vérification</u> : R est symétrique "des 1" sur la diagonale.

- Troisième étape : Diagonalisation de R.

a) Recherche des vp de R.

$$\det(R-\lambda I) = \begin{vmatrix} 1-\lambda & 0,693 & 0,5 \\ 0,693 & 1-\lambda & 0,925 \\ 0,5 & 0,925 & 1-\lambda \end{vmatrix} = \begin{vmatrix} 1-\lambda & 0,693 & 0,5 \\ 0 & -386-\lambda & -0,461+1,386\lambda \\ 0,5 & 0,925 & 1-\lambda \end{vmatrix}$$

$$L_1 - \frac{1-\lambda}{0,5} L_3 \to L_1$$

$$\det(R-\lambda I) = \begin{vmatrix} 0 & -1,157+1,85\lambda & -2\lambda^2+4\lambda-1,5 \\ 0 & -0,282-\lambda & -0,461+1,386\lambda \\ 0,5 & 0,925 & 1-\lambda \end{vmatrix}$$

$$= 0,5 \begin{vmatrix} -1,157+1,85\lambda & -2\lambda^2+4\lambda-1,5 \\ -0,282-\lambda & -0461+1,386\lambda \end{vmatrix}$$

$$= 0,5\,(-2\lambda^3 + 6\lambda^2 - 2,828\lambda + 0,11)$$
$$= -\lambda^3 + 3\lambda^2 - 1,414\lambda + 0,055.$$

Vérif : - polynome de degré 3
- terme dominant $(-1)^3\lambda^3 = -\lambda^3$.
- $\frac{-3}{-1} = \overline{x}$ des vp $= tr(R) = 3$

→ résolution de $\det(R-\lambda_j) = 0$.
résolution approchée
3 vp
vp : variance du facteur correspondant
$\lambda_1 = 2,43$
$\lambda_2 = 0,53$
$\lambda_3 = 0,004$