

TP EX12 : Tableau de notes

Pour cet exercice on pourra se reporter, si nécessaire, au document « ACPrappels »

Les Données : Il s'agit d'un tableau de notes de 9 élèves dans 5 matières (données contenues dans le fichier notes.dat). Les élèves sont les individus, et les matières correspondent aux variables.

Noms	math	phys	fran	latin	dessin
Alain	6,0	6	5	5,5	8
Benoit	8,0	8	8	8	9
Cyril	6,0	7	11	9,5	11
Daisy	14,5	14,5	15,5	15	8
Emilie	14,0	14	12	12,5	10
Fanny	11,0	10	5,5	7	13
Gaétan	5,5	7	14	11,5	10
Hélène	13,0	12,5	8,5	9,5	12
Inès	9,0	9,5	12,5	12	18

Les noms des élèves ne sont pas contenus dans le fichier.

Il faudra donc associer les noms à chacune des lignes du tableau.

a) Représentation des Données

Reproduire le tableau ci-dessus, indiquer les moyennes par élève et par matière.

Dresser le tableau des corrélations.

Par une représentation graphique on cherchera à faire ressortir des informations sur les données.

ACP sur données centrées

On désire réaliser une ACP sur les données centrées par variable.

- Réaliser le centrage.
- Soit X la matrice de données centrées, il faudra déterminer les valeurs et les vecteurs propres de la matrice $X'X$.
- Il restera à trier les valeurs propres, avec les vecteurs propres associés, selon un ordre décroissant.

b) Inertie.

On demande de représenter, sur une même figure, les pourcentages d'inertie des valeurs propres, sous forme de barres, ainsi qu'une courbe correspondant aux inerties cumulées.

- Pour poursuivre l'exercice, calculer :
 - les nouvelles coordonnées des individus (Xu)
 - les nouvelles coordonnées des variables ($X'v$)

c) Qualités de représentation et contributions des individus.

Pour avoir des éléments supplémentaires d'appréciation des résultats en vue d'une interprétation, calculer et afficher les qualités de représentation (qlt) des individus et leurs contributions (ctr) à la fabrication des axes. On rappelle les formules suivantes :

- La qualité de représentation d'un individu (\mathbf{o}_i), de norme $\|\mathbf{o}_i\|$, par un axe k est donnée par :

$$qlt_k(\mathbf{o}_i) = c_{ik} / \|\mathbf{o}_i\|^2 \quad \text{avec } c_{ik} \text{ la coordonnée de l'individu } i \text{ sur l'axe } k.$$

L'individu i sera d'autant mieux représenté sur l'axe k , que cette valeur est proche de 1.

- La contribution d'un individu (\mathbf{o}_i) à la fabrication d'un axe k . C'est la part d'inertie de l'axe k pris en compte ou expliquée par l'individu i . Elle est donnée par :

$$ctr_k(\mathbf{o}_i) = q_i c_{ik}^2 / \lambda_k \quad \text{où } q_i \text{ est le poids de l'individu } i, \text{ et } \lambda_k \text{ l'inertie de l'axe } k.$$

Dans notre cas, où l'on recherche les valeurs et vecteurs propres de la matrice $\mathbf{X}'\mathbf{X}$ où \mathbf{X} est simplement centrée, c'est-à-dire dont les données n'ont pas été pondérées, les poids q_i valent 1 quelque soit i .

On vérifiera que la somme des contributions pour chaque axe vaut 100% ainsi que la somme des qualités de représentation par individu.

d) Etudes des représentations

On s'intéresse aux 3 axes qui cumulent le plus d'inertie pour comprendre les données. Pour chacun des plans formés en prenant ces axes deux à deux on demande de :

- présenter une figure pour le **nuage des individus** et une autre pour le **nuage des variables**.
- faire faire une **analyse** détaillée de l'ensemble des données, en se servant de tous les éléments que calculés.