

Analyse des données

TP 3

Objectif

Dans ce présent TP, il s'agit de réaliser une analyse en composantes principales (ACP) normée, de visualiser et d'interpréter les résultats obtenus.

Exercice

Considérons la matrice des données ci-dessous résumant les résultats d'une enquête effectuée sur douze familles.

L'étude consiste à étudier les frais mensuels des chefs de familles, de différents types de métiers, de sept différents produits (denrées) alimentaires.

Les individus représentent les **différents types de métiers** des chefs de familles marqués par le **nombre de personnes** qui composent chaque famille :

- (Ouv, i) = Ouvrier avec une famille de i personnes, i = 4, 5, 6, 7.
- (Salr, i) = Salarié avec une famille de i personnes, i = 4, 5, 6, 7.
- (Cad. S, i) = Cadre supérieur avec une famille de i personnes, i = 4, 5, 6, 7.

Les variables sont les sept produits consommés :

- Produits laitiers, Pommes de terre, Fruits, Viande blanche, Viande, Pain et Légumes secs.

Par conséquent, les données sont les frais mensuels de ces produits alimentaires.

<i>Produits Métiers</i>	<i>Produits laitiers</i>	<i>Pommes de terre</i>	<i>Fruits</i>	<i>Viande Blanche</i>	<i>Viande</i>	<i>Pain</i>	<i>Légumes secs</i>
<i>I1 (Ouv, 7)</i>	495	776	423	759	1848	655	486
<i>I2 (Salr, 7)</i>	518	995	548	893	2056	584	319
<i>I3 (Cad.S, 7)</i>	561	1097	887	1167	2630	515	284
<i>I4 (Ouv, 6)</i>	414	660	367	638	1620	534	407
<i>I5 (Salr, 6)</i>	400	699	484	762	1856	460	416
<i>I6 (Cad.S, 6)</i>	304	789	621	1149	2366	385	282
<i>I7 (Ouv, 5)</i>	324	563	341	544	1507	406	407
<i>I8 (Salr, 5)</i>	319	608	396	568	1501	386	363
<i>I9 (Cad.S, 5)</i>	243	843	689	1148	2345	438	341
<i>I10 (Ouv, 4)</i>	247	428	354	526	1437	332	427
<i>I11 (Salr, 4)</i>	239	559	388	567	1527	293	258
<i>I12 (Cad.S, 4)</i>	235	767	562	927	1948	372	433

Pour atteindre le but de cette étude, nous réalisons une ACP normée sur les données de cette matrice.

Pour ce faire :

Analyse des données

- 1) Afficher la matrice des corrélations.
- 2) Afficher un tableau résumant les informations suivantes :
 - Les valeurs propres.
 - Le **Taux** d'inertie (par rapport à l'inertie totale) porté par chaque **axe principal**.
 - Les **Taux cumulés** de l'information retenue.
- 3) Déterminer les **coordonnées de tous les individus** sur tous les axes principaux dans un tableau.
- 4) Visualiser graphiquement tous les individus dans le plan principal. Nommer vos axes ainsi que les individus dans votre graphe. Préciser (afficher) le taux d'inertie porté par chaque axe au-dessous de l'axe correspondant.
- 5) Déterminer les **coordonnées des variables** sur tous les axes principaux.
- 6) Afficher un tableau résumant les **corrélations** entre les **variables et les axes principaux**.
- 7) Visualiser graphiquement toutes les variables dans le plan principal (**cercle de corrélations**). Nommer vos axes ainsi que les variables dans votre graphe. Préciser (afficher) le taux d'inertie porté par chaque axe au-dessous de l'axe correspondant.
- 8) Interpréter les graphes obtenus. Analyser bien le nuage des variables.
- 9) Déterminer la **qualité de représentation** de chaque individu dans la construction du plan factoriel. (Afficher les résultats dans un tableau tout en précisant la qualité de représentation de chaque individu sur chaque axe).
- 10) Calculer la **contribution** de chaque individu à l'inertie de chaque axe principal. Afficher les contributions dans un tableau.
- 11) Marquer graphiquement tous les individus ayant une très bonne qualité de représentation.
- 12) Déterminer les individus qui sont mal présentes sur le plan principal. Par quoi, pouvez-vous interpréter ça ?