Cours d'Analyse des Données

Excercices courigés ACP + AFC

2021/2022

département de mathématiques
université de chlef

Exercice 1 :

On considère la matrice de données X de type (2,3) suivante :

$$X = \begin{pmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}$$

1) Calculer le produit matriciel $X'X$ et s'assurer que c'est une matrice carrée et
symétrique.

2) Chercher les valeurs propres λ_i de $X'X$ et ses vecteurs propres associés u_i. 3) Vérifier que

$$tr(X'X) = tr(\Lambda) = \sum_i \lambda_i$$

Solution 1 :

$$X'X = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{pmatrix}$$

Ce qui est bien une matrice carrée d'ordre 3 et elle est symétrique.

$$2)\quad \det(X'X-\lambda I)=0 \Rightarrow \begin{vmatrix} 1-\lambda & 0 & -1 \\ 0 & 1-\lambda & -1 \\ -1 & -1 & 2-\lambda \end{vmatrix} = 0$$

$$\lambda(1-\lambda)(\lambda-3)=0$$

$\lambda_1=3$; $\lambda_2=1$; $\lambda_3=0$

Si $\lambda_1=3$, alors $X'Xu=\lambda_1 u$, c'est-à-dire $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 3x \\ 3y \\ 3z \end{pmatrix}$

$$u_1 = \left(-\frac{\sqrt{6}}{6};-\frac{\sqrt{6}}{6};\frac{\sqrt{6}}{3}\right) \qquad u_3 = \left(\frac{\sqrt{3}}{3};\frac{\sqrt{3}}{3};\frac{\sqrt{3}}{3}\right)$$

$$u_2 = \left(\frac{-\sqrt{2}}{2};\frac{\sqrt{2}}{2};0\right)$$

$$tr(X'X)=1+1+2=4=tr(\Lambda)=3+1+0$$

Exercice n° 1.

Soit le tableau de données

$$\mathbf{T} = \sqrt{10}\begin{bmatrix} 2 & 2 & 3 \\ 3 & 1 & 2 \\ 1 & 0 & 3 \\ 2 & 1 & 4 \\ 2 & 1 & 3 \end{bmatrix}$$

correspondant à des mesures effectuées sur 5 individus de poids statistiques égaux pour les trois variables T^1, T^2 et T^3. On va effectuer une ACP centrée-réduite sur ce tableau.

1. Calculer l'individu moyen, le vecteur $(\sigma_1, \sigma_2, \sigma_3)'$ des écarts types des variables et la matrice $\mathbf{X}$ des données centrées-réduites.

2. Calculer la matrice des corrélations $\mathbf{R}$

3. Effectuer la décomposition aux valeurs propres de $\mathbf{R}$

4. Les deux premiers vecteurs de $\mathbf{R}$ sont

$$\xi_1' = \frac{1}{2}\left(\sqrt{2}, 1, -1\right)' \text{ et } \xi_2' = \frac{1}{\sqrt{2}}(0, 1, 1)'.$$

   Ils sont associés aux valeurs propres

$$\lambda_1 = 1 + \frac{\sqrt{2}}{2} \text{ et } \lambda_2 = 1.$$

   Calculer les composantes principales $\mathbf{c}_1$ et $\mathbf{c}_2$ dont on vérifiera les propriétés statistiques

5. Représenter les individus dans le plan factoriel $(1, 2)$. Donner une interprétation de cette ACP

Correction exercice n°1

1. L'individu moyen est obtenu en faisant la moyenne des colonnes du tableau T, soit $\overline{\mathbf{x}} = \sqrt{10}\,(2,1,3)'$. Le vecteur des écarts types est obtenu en calculant les écarts types de chaque colonnes de T. Soit $\mathbf{T}_c$ la matrice des données centrées, $\mathbf{T}_c = \mathbf{T} - (\overline{\mathbf{x}}', \overline{\mathbf{x}}', \overline{\mathbf{x}}', \overline{\mathbf{x}}', \overline{\mathbf{x}}')'$

$$\mathbf{T}_c = \sqrt{10}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

le vecteur $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)'$ contient les termes en racine carrée des éléments diagonaux de la matrice $\mathbf{V} = \frac{1}{n}\mathbf{T}_c'\mathbf{T}_c$, $n = 5$, soit $\boldsymbol{\sigma}^2 = \frac{10}{5}\,(2,2,2)'$. Le calcul de la matrice $\mathbf{X}$ revient à diviser chaque colonne de $\mathbf{T}_c$ par l'écart-type de la variable correspondante :

$$\mathbf{X} = \frac{\sqrt{10}}{2}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & -1 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \frac{1}{2}\mathbf{T}_c$$

2. La matrice des corrélations

$$\mathbf{R} = \frac{1}{n}\mathbf{X}'\mathbf{X} = \frac{1}{5}\frac{10}{4}\left[\begin{array}{ccc} 2 & 1 & -1 \\ 1 & 2 & 0 \\ -1 & 0 & 2 \end{array}\right]$$

3. L'ACP centrée-réduite de $\mathbf{T}$ nécessite le calcul des vecteurs propres de $\mathbf{R}$. On résout le système $\det(\mathbf{R} - \lambda\mathbf{I}) = 0$ soit

$$(1-\lambda)\left(\lambda - 1 - \frac{1}{\sqrt{2}}\right)\left(\lambda - 1 + \frac{1}{\sqrt{2}}\right) = 0$$

et on obtient 3 valeurs propres $\lambda_1 = 1 + \frac{1}{\sqrt{2}}$, $\lambda_2 = 1$, $\lambda_3 = 1 - \frac{1}{\sqrt{2}}$.

4. Le calcul des 2 premières composantes principales est donné par

$$\mathbf{c}^i = \mathbf{X}\boldsymbol{\xi}_i,\ i = 1, 2$$

soit pour la première, associée à la valeur propre λ_1,

$$\mathbf{c}^1 = \frac{\sqrt{10}}{2} \times \frac{1}{2}\left[\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 0 & -1 \\ -1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{array}\right]\left(\begin{array}{c} \sqrt{2} \\ 1 \\ -1 \end{array}\right) = \frac{\sqrt{10}}{4}\left(\begin{array}{c} 1 \\ \sqrt{2}+1 \\ -\sqrt{2}-1 \\ -1 \\ 0 \end{array}\right)$$

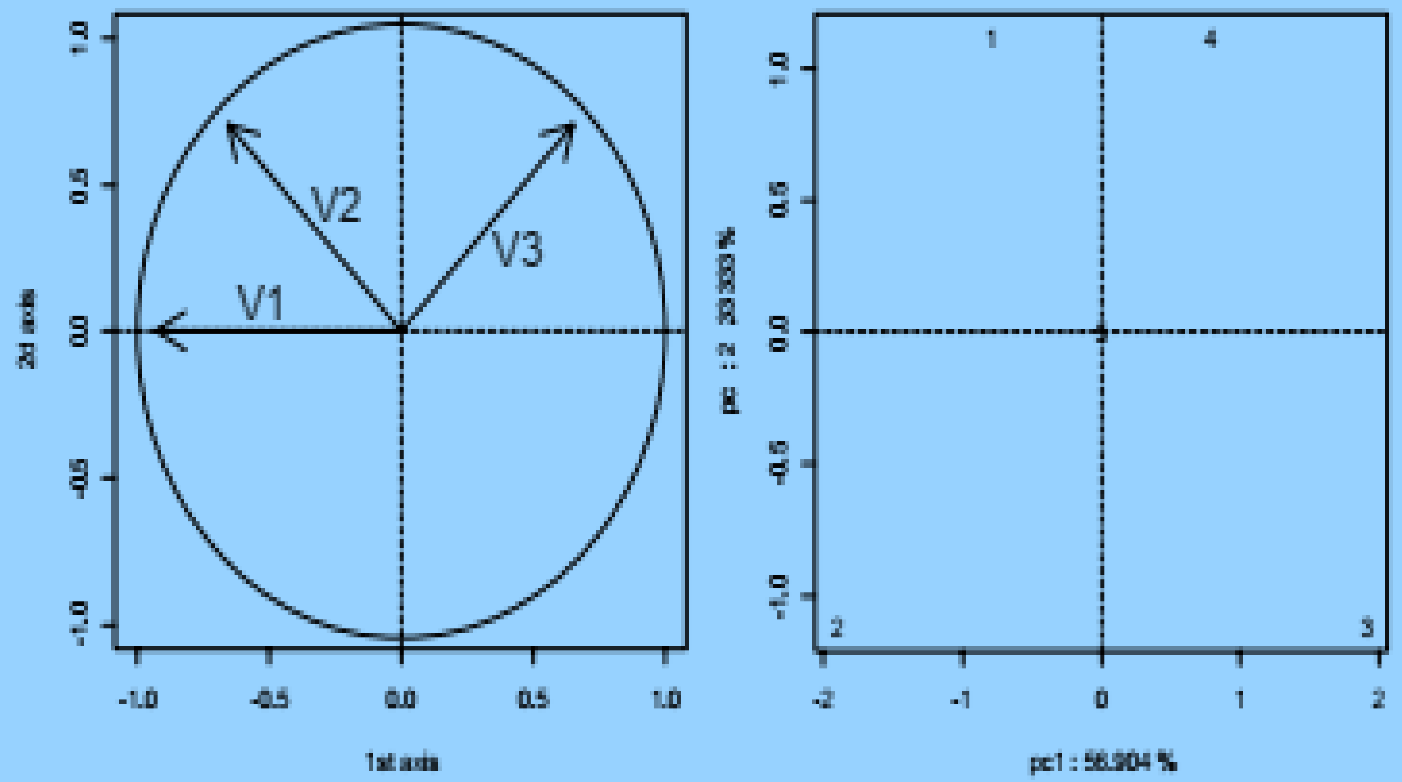
et la seconde $\mathbf{c}^2 = \frac{\sqrt{10}}{2} \times \frac{1}{\sqrt{2}}(1, -1, -1, 1, 0)'$. Les propriétés de ces composantes montrent qu'elles sont orthogonales deux à deux

$$\langle \mathbf{c}^i, \mathbf{c}^j \rangle_{\mathbf{M}} = \frac{1}{5}\mathbf{c}^{i\prime}\mathbf{c}^j = 0,\ \forall i \neq j$$

et que leur norme est reliée à chaque valeur propre par

$$\|\mathbf{c}^j\|^2_{\mathbf{M}} = \frac{1}{5}\mathbf{c}^{j\prime}\mathbf{c}^j = \lambda_j,\ j = 1, 2.$$

5. représentation des individus dans le plan factoriel $(1, 2)$.



Le premier axe oppose les variations de V1, V2 avec V3. Le second est un axe de taille. Les individus 2 et 3 présentent de faibles valeurs de V2 et V3, l'individu 2 étant caractérisé par une forte valeur de V1. Les individus 1 et 4 sont attachés aux variables V2 et V3 respectivement. L'individu 5 est le plus consensuel puisque confondu avec le centre de gravité de l'ACP.

Exercice *02:* Soit X la matrice $\begin{pmatrix} 0 & 30 & 4000 \\ 0 & 10 & 5000 \\ 4 & 30 & 5000 \\ 4 & 10 & 6000 \end{pmatrix}$, représentant les valeurs de 3 variables pour 4 individus. Nous allons en réaliser l'ACP.

1. Centrer et réduire X. On appellera cette nouvelle matrice de données $\tilde{X}$.
2. Calculer Φ, la matrice de variance-covariance empirique de $\tilde{X}$. C'est aussi la matrice de corrélation de X.
3. Identifier les valeurs propres de Φ et l'écrire sous la forme $U\Lambda U^\intercal$ où Λ est une matrice diagonale. Que représentent Λ et U dans l'ACP ?
4. Donner les composantes principales, représenter le graphique des individus pour les 2 premiers axes et l'interpréter. Que se passe-t-il pour le troisième axe ?
5. Calculer les variances des composantes principales. Que vérifiez-vous ?
6. Calculer les pourcentages d'inertie représentés par chaque axe.
7. Représenter le cercle des corrélations, représentant les corrélations entre composantes principales et variables initiales. Comment l'interprétez-vous ?

Exercice :

Trois notes ont été relevées sur cinq étudiants. Elle figurent dans le tableau suivant :
Mathématiques ; Français ; Economie

$$T = \begin{pmatrix} 11 & 18 & 14 \\ 8 & 10 & 12 \\ 10 & 6 & 10 \\ 9 & 14 & 8 \\ 9.5 & 12 & 11 \end{pmatrix}$$

1. Déterminer les moyennes et écart-types des trois notes (les cinq étudiants ont un poids identique), afin d'effectuer une A.C.P. normée sur ces données.
2. Calculez Z issue du centrage et de la réduction des données du tableau X.
3. Quelle matrice doit-on diagonaliser pour obtenir les vecteurs directeurs des axes principaux ? Comment s'exprime-t-elle en fonction de Z si les poids sont uniformes ?
4. Vérifiez que ω1 et ω2 sont bien des vecteurs propres de R
5. En déduire trois valeurs propres de R. Quelle est la marge de manœuvre dans le choix de u1,u2 et u3, base orthonormée de vecteurs directeurs des axes principaux ?
6. Représentez les trois variables dans le cercle des corrélations

$$\omega_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\omega_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

1. les moyennes et écart-types des trois notes

Matière	Moyenne	Ecart-type
Mathématiques	9.5	1
Français	12	4
Economie	11	2

2. Calculez Z

$$Z = \begin{pmatrix} -1.5 & 1.5 & 1.5 \\ -1.5 & -0.5 & 0.5 \\ 0.5 & -1.5 & -0.5 \\ -0.5 & 0.5 & -1.5 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\boldsymbol{R} = \frac{\mathbf{1}}{\mathbf{5}}\boldsymbol{Z'Z} = \begin{pmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 \\ 0.4 & 0.4 & 1 \end{pmatrix}$$

3. La matrice à diagonaliser est la **matrice des corrélations**
4. On vérifie bien que $\boldsymbol{R\omega_1 = 1.8\omega_1}$ et $\boldsymbol{R\omega_2 = 0.6\omega_2}$.
5. La troisième valeur propre est égale à $3 - (1.8 + 0.6) = 0.6$ $\quad u_k = \frac{\omega_k}{\|\omega_k\|}$
6.

On a : $(\boldsymbol{c_k}) = \boldsymbol{\sqrt{\lambda_k} u_k \; \forall j = 1, 2, 3}$, d'où :

$$(c_1) = \sqrt{0.6}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \text{ et } (c_2) = \sqrt{0.3}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

Analyse Factorielle des Correspondances

Une étude a été menée sur 592 femmes dans le but d'analyser une éventuelle liaison entre les variables “couleur des yeux” et “couleur des cheveux”. Dans la suite, on désignera par X la variable “couleur des yeux” et par Y la variable “couleur des cheveux”. Les résultats de l'étude sont contenus dans le tableau suivant :

X \ Y	**Brun**	**Châtain**	**Roux**	**Blond**
Marron	68	119	26	7
Noisette	15	54	14	10
Vert	5	29	14	16
Bleu	20	84	17	94

1. Le tableau suivant donne la distribution bivariée en pourcentage, ainsi que les distributions marginales elles-mêmes exprimées en pourcentage. Compléter alors ce tableau.

	Brun	**Châtain**	**Roux**	**Blond**	$f_{i.}$
Marron	11.49	20.11	4.39	1.18	·
Noisette	·	9.12	2.36	1.69	15.70
Vert	0.84	4.90	2.36	2.70	10.80
Bleu	3.38	14.19	2.87	15.89	36.33
$f_{.j}$	·	48.32	11.98	21.46	·

2. Le tableau suivant donne les profils-lignes. Compléter ce tableau.

%	Brun	Châtain	Roux	Blond	
Marron	·	54.11	11.81	3.17	·
Noisette	16.11	·	15.03	10.76	·
Vert	7.78	45.37	21.85	25.00	·
Bleu	9.30	·	7.90	43.74	·
G_ℓ	·	·	·	·	100

3. Indiquer dans quel espace sera représenté le nuage des profils-lignes. Ces profils vont en fait appartenir à un sous-espace. Préciser la dimension de ce dernier, puis indiquer la raison de cette propriété.

4. Le tableau suivant donne les profils-colonnes. Compléter ce tableau.

%	Brun	Châtain	Roux	Blond	G_c
Marron	62.99	41.62	36.64	5.50	·
Noisette	13.87	·	19.70	7.88	·
Vert	4.60	10.14	19.70	12.58	·
Bleu	·	29.37	23.96	·	·
	·	·	·	·	

5. Indiquer dans quel espace sera représenté le nuage des profils-colonnes. Ces profils vont en fait appartenir à un sous-espace. Préciser la dimension de ce dernier, puis indiquer la raison de cette propriété.
6. Pour chacun des nuages de profils, préciser la matrice des pondérations (i.e. D_n et D_p), ainsi que les matrices associées aux métriques (i.e. M_p et M_n).

7. Rappeler la formule matricielle permettant de calculer la matrice à diagonaliser, notée S, pour réaliser l'ajustement du nuage des profils-lignes. Pourquoi est-il inutile de chercher à écrire, puis à diagonaliser la matrice associée aux profils-colonnes ?

8. L'opération de diagonalisation de la matrice S par un logiciel de mathématiques a permis d'obtenir les valeurs propres suivantes :

$$\lambda_0 = 1 \qquad \lambda_1 = 0.20877 \qquad \lambda_2 = 0.02223 \qquad \lambda_3 = 0.00260$$

Que pouvez-vous dire sur la valeur propre λ_0 ? Sur quelles valeurs propres allez-vous concentrer votre analyse ?

9. Déterminer l'inertie associée à chacun des axes factoriels non-triviaux. En déduire le pourcentage d'inertie associé à chacun des axes, puis l'inertie cumulée.

10. Sachant que l'on veut conserver au moins 90% de l'inertie, indiquer le nombre d'axes factoriels que l'on doit sélectionner.

11. Les deux premiers vecteurs propres issus de la diagonalisation de la matrice S sont respectivement

$$u_1 = (-0.201436,\ -0.156799,\ -0.033982,\ 0.392135)$$

$$u_2 = (-0.262634,\ 0.105975,\ 0.256799,\ -0.100131)$$

Après avoir rappelé la formule permettant d'obtenir les coordonnées des profils-lignes dans la nouvelle base, compléter les matrices colonnes suivantes :

$$\Psi_1 = \begin{pmatrix} . \\ -0.2125 \\ 0.1617 \\ 0.5474 \end{pmatrix} \qquad \text{et} \qquad \Psi_2 = \begin{pmatrix} -0.0883 \\ 0.1673 \\ . \\ -0.0831 \end{pmatrix}$$

12. Rappeler la formule permettant d'obtenir la contribution d'un profil-ligne à la formation d'un axe. Compléter alors le tableau suivant :

Profils-lignes	$Cr_1(i)$	$Cr_2(i)$
1	.	13.04
2	3.40	19.79
3	1.35	.
4	52.14	11.26

13. Rappeler la formule permettant d'obtenir la qualité de représentation d'un profil-ligne sur un axe. Comment obtient-on alors facilement la qualité de représentation d'un profil-ligne sur le plan factoriel principal. Compléter alors le tableau suivant :

Profils-lignes	$Qual_1(i)$	$Qual_2(i)$	$Qual_{1\times2}(i)$
1	96.70	2.78	.
2	.	33.58	87.82
3	17.59	76.77	94.36
4	97.75	2.25	.

14. Indiquer la formule quasi-barycentrique permettant de calculer les coordonnées des profils-colonnes à partir des coordonnées des profils-lignes. Compléter alors les matrices colonnes suivantes :

$$\varphi_1 = \begin{pmatrix} \cdot \\ -0.1483 \\ -0.1296 \\ 0.8349 \end{pmatrix} \qquad \text{et} \qquad \varphi_2 = \begin{pmatrix} -0.2148 \\ 0.0327 \\ \cdot \\ -0.0696 \end{pmatrix}$$

15. Rappeler la formule permettant d'obtenir la contribution d'un profil-colonne à la formation d'un axe. Compléter alors le tableau suivant :

Profils-colonnes	$Cr_1(j)$	$Cr_2(j)$
1	22.25	·
2	5.09	2.32
3	0.96	55.13
4	·	4.67

16. Rappeler la formule permettant d'obtenir la qualité de représentation d'un profil-colonne sur un axe. Comment obtient-on alors facilement la qualité de représentation d'un profil-colonne sur le plan factoriel principal. Compléter alors le tableau suivant :

Profils-colonnes	$Qual_1(i)$	$Qual_2(i)$	$Qual_{1\times 2}(i)$
1	·	15.12	·
2	86.59	4.21	90.80
3	13.36	·	94.62
4	99.14	0.69	·

17. Fournir une représentation simultanée des deux nuages.

1. Tableau des fréquences exprimées en pourcentage, et distributions marginales :

	Brun	Châtain	Roux	Blond	$f_{i.}$
Marron	11.49	20.11	4.39	1.18	37.17
Noisette	2.53	9.12	2.36	1.69	15.70
Vert	0.84	4.90	2.36	2.70	10.80
Bleu	3.38	14.19	2.87	15.89	36.33
$f_{.j}$	18.24	48.32	11.98	21.46	100

Correction

2. Le tableau suivant donne les profils-lignes. Compléter ce tableau.

%	Brun	Châtain	Roux	Blond	
Marron	·	54.11	11.81	3.17	·
Noisette	16.11	·	15.03	10.76	·
Vert	7.78	45.37	21.85	25.00	·
Bleu	9.30	·	7.90	43.74	·
G_ℓ	·	·	·	·	100

2. Les profils-lignes sont les distributions conditionnelles de Y sachant X. On obtient

%	Brun	Châtain	Roux	Blond	
Marron	30.91	54.11	11.81	3.17	100
Noisette	16.11	58.10	15.03	10.76	100
Vert	7.78	45.37	21.85	25.00	100
Bleu	9.30	39.06	7.90	43.74	100
$f_{.j}$	18.24	48.32	11.98	21.46	100

3. Chaque profil-ligne ayant quatre coordonnées, les profils-lignes seront représentés dans l'espace $\mathbb{R}^4$. La somme des coordonnées étant constante et égale à 1, seules trois coordonnées sont suffisantes pour déterminer la position d'un profil. Par conséquent, les profils-lignes vont appartenir à un hyperplan, c'est-à-dire un sous-espace de dimension 3.

4. Les profils-colonnes sont les distributions conditionnelles de X sachant Y. On obtient

%	Brun	Châtain	Roux	Blond	$f_{i.}$
Marron	62.99	41.62	36.64	5.50	**37.17**
Noisette	13.87	**18.87**	19.70	7.88	**15.70**
Vert	4.60	10.14	19.70	12.58	**10.80**
Bleu	**18.54**	29.37	23.96	**74.04**	**36.33**
	100	**100**	**100**	**100**	**100**

5. Chaque profil-colonne ayant quatre coordonnées, les profils-colonnes seront représentés dans l'espace $\mathbb{R}^4$. La somme des coordonnées étant constante et égale à 1, seules trois coordonnées sont suffisantes pour déterminer la position d'un profil. Par conséquent, les profils-colonnes vont appartenir à un hyperplan, c'est-à-dire un sous-espace de dimension 3.

6. La matrice des pondérations des profils-lignes est

$$D_n = \begin{pmatrix} 0,3717 & 0 & 0 & 0 \\ 0 & 0,1570 & 0 & 0 \\ 0 & 0 & 0,1080 & 0 \\ 0 & 0 & 0 & 0,3633 \end{pmatrix}$$

La matrice des pondérations des profils-colonnes est

$$D_p = \begin{pmatrix} 0,1824 & 0 & 0 & 0 \\ 0 & 0,4832 & 0 & 0 \\ 0 & 0 & 0,1198 & 0 \\ 0 & 0 & 0 & 0,2146 \end{pmatrix}$$

A partir de ces matrices, et des relations $M_p = D_p^{-1}$ et $M_n = D_n^{-1}$, on obtient

$$M_p = D_p^{-1} = \begin{pmatrix} 5.4825 & 0 & 0 & 0 \\ 0 & 2.0695 & 0 & 0 \\ 0 & 0 & 8.3472 & 0 \\ 0 & 0 & 0 & 4.6598 \end{pmatrix} \quad \text{et} \quad M_n = D_n^{-1} = \begin{pmatrix} 2.6903 & 0 & 0 & 0 \\ 0 & 6.3694 & 0 & 0 \\ 0 & 0 & 9.2593 & 0 \\ 0 & 0 & 0 & 2.7525 \end{pmatrix}$$

7. La matrice S à diagonaliser est $S = F_\ell^t D_n F_\ell D_p^{-1}$. Sachant que l'on a $F_\ell = D_n^{-1} F$, on a aussi $S = F^t D_n^{-1} F D_p^{-1}$ Dans les deux cas, on aboutit à

$$S = \begin{pmatrix} 0.2378 & 0.1942 & 0.1825 & 0.1083 \\ 0.5146 & 0.4956 & 0.4956 & 0.4217 \\ 0.1199 & 0.1228 & 0.1348 & 0.1043 \\ 0.1274 & 0.1873 & 0.1868 & 0.3655 \end{pmatrix} \qquad T = \begin{pmatrix} 0.4649 & 0.4042 & 0.3316 & 0.2741 \\ 0.1707 & 0.1700 & 0.1591 & 0.1366 \\ 0.0963 & 0.1094 & 0.1240 & 0.1144 \\ 0.2679 & 0.3161 & 0.3851 & 0.4747 \end{pmatrix}$$

Remarque : La matrice T à diagonaliser pour l'ajustement des profils-colonnes est $T = F_c D_p F_c^t D_n^{-1}$. Sachant que l'on a $F_c = F D_p^{-1}$, on a aussi $T = F D_p^{-1} F^t D_n^{-1}$. Dans les deux cas, on aboutit à

8. D'après le cours, on sait que la première valeur propre est toujours égale à l'unité. Cette valeur propre est associée à un vecteur propre engendrant un sous-espace vectoriel de dimension un dit trivial. En pratique, on supprime cette valeur propre qui n'apporte pas d'information. On concentre alors l'analyse sur les trois dernières valeurs propres.

9. D'après le cours, on a $\quad I = \sum_{\alpha=1}^{3} I(\alpha) \qquad$ avec $I(\alpha) = \lambda_\alpha$.

Par conséquent, sachant que les valeurs propres sont $\lambda_1 = 0.20877$, $\lambda_2 = 0.02223$ et $\lambda_3 = 0.0026$, on obtient

$$I(1) = 0.20877 \qquad I(2) = 0.02223 \qquad I(3) = 0.0026 \qquad \text{avec } I = 0.23360.$$

On déduit alors que $\quad PI(1) = 89.37\% \qquad PI(2) = 9.52\% \qquad PI(3) = 1.11\%$

10. Sous la contrainte de devoir conserver au moins 90% d'information, on est amené à garder les deux premiers axes factoriels.

11. D'après le cours, les coordonnées des profils-lignes sur les deux premiers axes factoriels (les facteurs) se calculent via le produit matriciel suivant : $\Psi_\alpha = F_\ell D_p^{-1} u_\alpha$. Ainsi, on obtient

$$\Psi_1 = \begin{pmatrix} \mathbf{-0.4925} \\ -0.2125 \\ 0.1617 \\ 0.5474 \end{pmatrix} \qquad \text{et} \qquad \Psi_2 = \begin{pmatrix} -0.0883 \\ 0.1673 \\ \mathbf{0.3391} \\ -0.0831 \end{pmatrix}$$

12. D'après le cours, on a

$$Cr_\alpha(i) = f_{i.} \frac{\Psi_\alpha^2(i)}{\lambda_\alpha}$$

On obtient alors pour les profils-lignes

Profils-lignes	$Cr_1(i)$	$Cr_2(i)$
1	**43.18**	13.04
2	3.40	19.79
3	1.35	**55.83**
4	52.14	11.26

13. on a $Qual_\alpha(i) = \dfrac{\Psi_\alpha^2(i)}{\|\vec{G_\ell E_i}\|^2_{M_p}}$

On obtient alors pour les profils-lignes

Profils-lignes	$Qual_1(i)$	$Qual_2(i)$	$Qual_{1\times 2}(i)$
1	96.70	2.78	**99.48**
2	**54.24**	33.58	87.82
3	17.59	76.77	94.36
4	97.75	2.25	**100.00**

On peut observer que l'ensemble des profils-lignes présente une très bonne qualité de représentation.

14. on a la relation $\sqrt{\lambda_\alpha}\varphi_\alpha(j) = \sum^{4} f_{i/j}\Psi_\alpha(i), \qquad \forall\ j = 1, \cdots, 4.$

On obtient alors

$$\varphi_1 = \begin{pmatrix} \mathbf{-0.5046} \\ -0.1483 \\ -0.1296 \\ 0.8349 \end{pmatrix} \quad \text{et} \quad \varphi_2 = \begin{pmatrix} -0.2148 \\ 0.0327 \\ \mathbf{0.3196} \\ -0.0696 \end{pmatrix}$$

15. on a $Cr_\alpha(j) = f_{.j}\dfrac{\varphi_\alpha^2(j)}{\lambda_\alpha}$ On obtient alors pour les profils-colonnes

Profils-colonnes	$Cr_1(j)$	$Cr_2(j)$
1	22.25	**37.88**
2	5.09	2.32
3	0.96	55.13
4	**71.7**	4.67

16. on a $Qual_\alpha(j) = \dfrac{\varphi_\alpha^2(j)}{\|\vec{G_c V_j}\|^2_{M_n}}$ on obtient

Profils-colonnes	$Qual_1(j)$	$Qual_2(j)$	$Qual_{1\times 2}(j)$
1	**83.70**	15.12	**98.28**
2	86.59	4.21	90.80
3	13.36	**81.26**	94.62
4	99.14	0.69	**99.83**

On peut observer que l'ensemble des profils-colonnes présente une très bonne qualité de représentation.

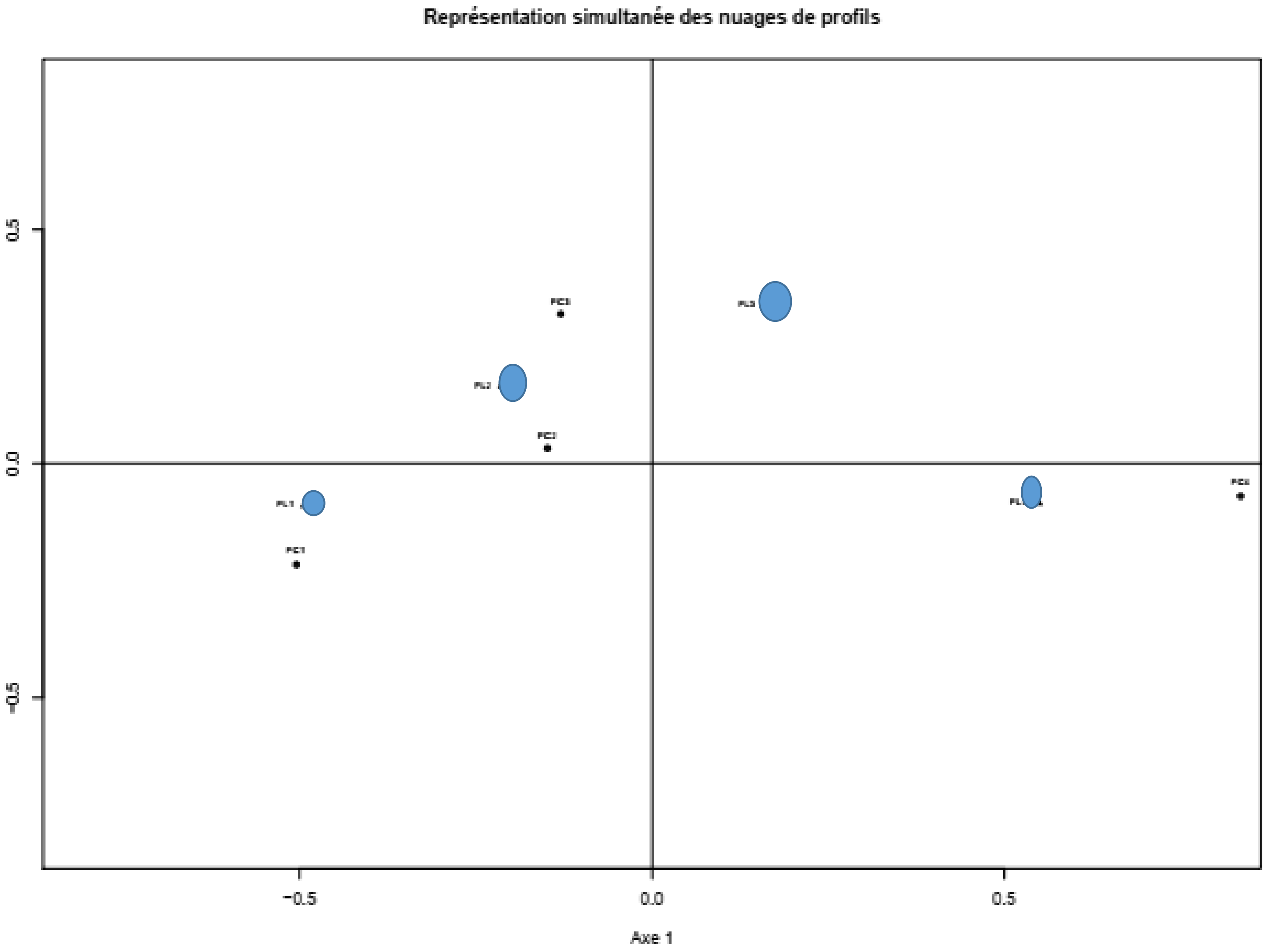


Fig. 1 - **Représentation simultanée**

Pour obtenir ce graphique sous R, le programme suivant a été écrit :

- u1=c(-0.225,-0.0971,0.0739,0.2501)
- u2=c(-0.0132,0.0249,0.0506,-0.0124)
- z1=c(-0.2306,-0.0678,-0.0592,0.3815)
- z2=c(-0.0320,0.0048,-0.0104,-0.032)
- plot(u1,u2,type="p",main="Représentation simultanée des deux nuages de profils",xlab=" ",ylab=" ", xlim=c(-0.4,0.4),ylim=c(-0.1,0.1))
- abline(h=0) ;abline(v=0)
- text(u1,u2,labels=c("PL1","PL2","PL3","PL4"),cex=0.6,pos="2")
- points(z1,z2)
- text(z1,z2,labels=c("PC1","PC2","PC3","PC4"),cex=0.6,pos="3")

Exercice n^o 2

Une enquête menée en 1911 auprès de 1725 élèves d'écoles primaires avait pour objectif de vérifier l'existence ou non d'une relation entre "l'intelligence" et l'origine sociale de l'élève au travers de sa qualité vestimentaire. Les résultats observés ont été les suivants :

X \ Y	Y1	Y2	Y3	Y4	Y5	Y6
X1	30	51	115	207	193	40
X2	40	100	203	253	140	15
X3	39	55	73	64	30	4
X4	17	15	20	10	10	1

où l'on a posé

$X1 =$ Très bien habillé
$X2 =$ Bien habillé
$X3 =$ Passablement habillé
$X4 =$ Mal habillé

$Y1 =$ Lent et stupide
$Y2 =$ Lent
$Y3 =$ Lent mais intelligent
$Y4 =$ Assez intelligent
$Y5 =$ Intelligent
$Y6 =$ Très intelligent