Analyse des Données

Travaux Dirigés

2015-2016

TD 1

Rappels d'algèbre linéaire

I. Projecteurs

Soit E un espace vectoriel et soient E_1 et E_2 deux sous-espaces vectoriels de E. On rappelle que E_1 et E_2 sont dits *supplémentaires*, et on note $E = E_1 \oplus E_2$, si pour tout x de E, il existe de façon unique deux vecteurs $x_1 \in E_1$ et $x_2 \in E_2$ tels que :

$$x = x_1 + x_2 \,.$$

On rappelle que le *projecteur* P sur E_1 parallèlement à E_2, est l'application qui à tout vecteur x associe le vecteur x_1. Par la suite, l'application identité est notée I.

I. 1. Montrer que tout projecteur est linéaire et idempotent (i.e. $P\,2 = P$).

I. 2. Si P est un endomorphisme, montrer que les propriétés suivantes sont équivalentes :

$$\begin{aligned} &\text{(i)} \quad P \text{ est idempotent,} \\ &\text{(ii)} \quad \mathrm{Ker}P = (I - P)(E), \\ &\text{(iii)} \quad P \text{ est un projecteur.} \end{aligned}$$

I. 3. De la relation $P^2 = P$, déduire que les valeurs propres sont 1 ou 0.

II. Projecteurs M-orthogonaux

Soit E un espace vectoriel euclidien muni d'une métrique M, autremant dit le produit scalaire de deux vecteurs x et y de E, s'écrit :

$$M(x, y) = x'M\,y = y'M\,x\,.$$

On considère un sous-espace vectoriel F de E, et on note $F^\perp$ *l'orthogonal* de F selon la métrique M, c'est-à-dire le sous-espace vectoriel défini par :

$$F^\perp = \{y \in E \mid \forall x \in F,\ M(x, y) = 0\}.$$

Rappelons que le projecteur M-orthogonal sur F est le projecteur sur F parallèlement à $F^\perp$. Dans la suite de ce texte, P désigne le projecteur M-orthogonal sur F.
On considère une partie génératrice $\{x^1, ..., x^p\}$ à p éléments de F, et on note X' la matrice dont les colonnes sont formées des x^i, qui peut donc s'écrire sous la forme :

$$X' = (x^1, ..., x^p).$$

II. 1. Montrer que pour tout $i \leq p$, on a :

$$\forall y \in E, \quad (x^i)'M(y - P\,y) = 0\,.$$

et en déduire :

$$\forall y \in E, \quad XMy = XMPy\,. \tag{1}$$

II. 2. Déduire de (1) les équations dites *normales* :

$$\begin{cases} XMy = XMX'b \\ Py = X'b \end{cases}$$

où b désigne un vecteur ayant p composantes.

II. 3. Montrer que les propriétés suivantes sont équivalentes :

(a) F est de dimension p,

(b) X' est injective,

(c) X' est de rang p,

(d) la forme quadratique XMX' est définie positive,

(e) XMX' est un isomorphisme.

II. 4. Montrer que si $x^1, \ldots, x^p$ sont linéairement indépendants, alors :

$$P = X'(XMX')^{-1}XM\,.$$

II. 5. Une application linéaire A est dite M-symétrique si pour tout x et y de E, on a la relation $M(A\,x,\ y) = M(x, A\,y)$; à partir de cette définition, montrer qu'un projecteur (quelconque) est M-orthogonal si et seulement si il est M-symétrique.

II. 6. Montrer que le projecteur M-orthogonal sur F peut également être défini comme étant l'application qui à tout x de E associe l'unique vecteur $\widehat{x}$ défini par la condition (c) :

$$(c) \begin{cases} \| x - \widehat{x} \|_M^2 \text{ est minimum,} \\ \widehat{x} \in F\,. \end{cases}$$

(Pour cela, on pourra utiliser le théorème de Pythagore . . .).

TD 2
Interprétation géométrique de la moyenne et de la covariance empiriques

Dans ce texte, on considère p variables dont on connait les valeurs sur un échantillon de n individus.

Définitions et notations

On notera x_i^j la valeur de la variable i $(1 \leq i \leq p)$ pour l'individu j $(1 \leq j \leq n)$. Il en résulte qu'une variable i est caractérisée par le vecteur x_i de $F = \mathbb{R}^n$, vecteur dont les composantes sont les x_i^j pour $1 \leq j \leq n$. De même, un individu j est caractérisé par le vecteur x^j de $E = \mathbb{R}^p$, vecteur dont les composantes sont les x_i^j pour $1 \leq i \leq p$.

Chaque individu j est muni d'un poids p_j tel que :

$$\sum_{j \in J} p_j = 1,$$

avec $J = \{1, \ldots, n\}$. Les poids p_j sont généralement égaux à $\dfrac{1}{n}$.

Rappelons les définitions suivantes :

- Moyenne (empirique) de la variable i : $\overline{x}_i = \sum_{j \in J} p_j \, x_i^j$.
- Variable i centrée : $y_i^j = x_i^j - \overline{x}_i$
- Covariance (empirique) entre les variables i et i' :

$$v_{ii'} = \sum_{j \in J} p_j (x_i^j - \overline{x}_i)(x_{i'}^j - \overline{x}_{i'}) = \sum_{j \in J} p_j \, y_i^j \, y_{i'}^j.$$

- Variance (empirique) de la variable i : $s_i^2 = v_{ii}$
- Corrélation (empirique) entre les variables i et i' : $r_{ii'} = \dfrac{v_{ii'}}{s_i \, s_{i'}}$.

On note :

$$\overline{x} = \begin{pmatrix} \overline{x}_1 \\ \vdots \\ \overline{x}_p \end{pmatrix} \qquad y^j = \begin{pmatrix} y_1^j \\ \vdots \\ y_p^j \end{pmatrix} \qquad y_i = \begin{pmatrix} y_i^1 \\ \vdots \\ y_i^n \end{pmatrix}$$

vecteur des moyennes — individu j après centrage — variable i centrée

On note enfin j_n le vecteur de $\mathbb{R}^n$ dont toutes les composantes sont égales à 1, D_p la matrice $(n \times n)$ diagonale des poids p_j, X la matrice $(p \times n)$ des données x_i^j et Y la matrice $(p \times n)$ des données centrées y_i^j.

1. Pour quelle métrique N de $\mathbb{R}^n$ la moyenne $\overline{x}_i$ peut-elle être considérée comme l'abscisse de la projection de x_i sur j_n ?

2. A partir du résultat obtenu en **1.** et de la relation : $y_i^j = x_i^j - \overline{x}_i$ avec j variant de 1 à n, montrer que y_i est l'image de x_i, selon une transformation géométrique de $\mathbb{R}^n$ que l'on précisera.

3. De même, à partir de la relation : $y_i^j = x_i^j - \overline{x}_i$ avec i variant de 1 à p, montrer que y^j est l'image de x^j, selon une transformation géométrique de $\mathbb{R}^p$ que l'on précisera.

4. Interpréter à l'aide du produit scalaire de $\mathbb{R}^n$ défini par D_p, les quantités $v_{ii'}, s_i^2$ et $r_{ii'}$.

5. Soit V la matrice variance empirique des p variables, c'est-à-dire la matrice pxp dont le terme général est $v_{ii'}$. Montrer que V définit une forme bilinéaire symétrique positive.

6. On sait qu'à tout vecteur u de $\mathbb{R}^p$ on peut associer un unique élément $u^\star$ de $(\mathbb{R}^p)^\star$, c'est-à-dire une forme linéaire $u^\star$ de $\mathbb{R}^p$ qui est définie par [†] :

$$\forall z \in \mathbb{R}^p, \quad u^\star(z) = \sum_{i=1}^{p} u_i \, z_i$$

Par conséquent, tout vecteur u définit une nouvelle variable qui est combinaison linéaire des variables x_i et qui vaut $u^\star(x^j)$ pour l'individu j , c'est-à-dire :

$$\sum_{i=1}^{p} u_i \, x_i^j$$

Montrer que $u'Vu$ est égal à la variance de la variable ainsi associée au vecteur u de $\mathbb{R}^p$. De même, si w désigne un deuxième vecteur de $\mathbb{R}^p$, montrer que $u'Vw$ est égal à la covariance empirique des variables associées aux vecteurs u et w. En déduire que V peut être considérée comme une forme quadratique semi-définie positive sur le dual $(\mathbb{R}^p)^\star$.

7. Montrer que V peut s'écrire sous la forme matricielle suivante :

$$V = Z \, N \, Z',$$

où N est la métrique définie en **1.** et Z est une matrice à préciser. Si V est une matrice définie, on dit alors que V est la métrique induite par la métrique N et par l'application linéaire Z.

†. cf. Rappels d'Algèbre Linéaire

TD 3
Application du théorème des trois perpendiculaires
à l'analyse en composantes principales

Soit E un espace vectoriel de dimension finie, muni d'une métrique M. Par la suite W désigne un sous-espace vectoriel de E et l'on note P_W le projecteur M-orthogonal sur W.

1. Inertie d'un nuage de points. On rappelle qu'un nuage $\mathcal{M}$ de n points munis de masses p_j, peut être identifié à l'ensemble formé par les n vecteurs $x_j \in E$ représentant ces points :

$$\mathcal{M} = \{x^j \mid j = 1, ..., n\},$$

où E désigne ici l'espace vectoriel associé à l'espace affine $\mathcal{E}$ contenant les n points. On rappelle que le centre de gravité g de ce nuage est défini par :

$$g = \frac{1}{p} \sum_{j=1}^{n} p_j \, x^j,$$

avec $p = \sum_{j=1}^{n} p_j$. Dans tout le texte, on suppose que $g = 0$ et que l'inertie totale du nuage $\mathcal{M}$ peut s'écrire sous la forme :

$$I_T(\mathcal{M}) = \sum_{j=1}^{n} p_j \, \|xj - g\|^2 = \sum_{j=1}^{n} p_j \, \|xj\|^2 .$$

De plus, on définit l'inertie du nuage $\mathcal{M}$ par rapport au sous espace vectoriel W comme étant :

$$I_W(\mathcal{M}) = \sum_{j=1}^{n} p_j \, d^2(x^j, P_W(x^j)).$$

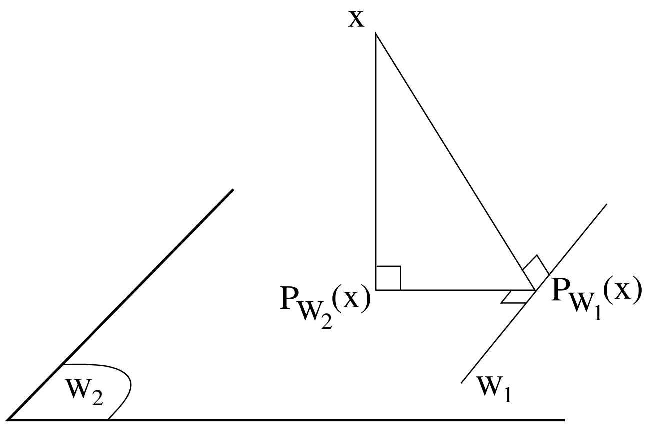
En utilisant le fait que l'application linéaire $I - P_W$ est le projecteur M-orthogonal sur $W^\perp$, montrer les relations suivantes :

$$\begin{aligned} &\text{a)} \quad I_{W^\perp}(\mathcal{M}) = I_T(P_W(\mathcal{M})), \\ &\text{b)} \quad I_T(\mathcal{M}) = I_W(\mathcal{M}) + I_{W^\perp}(\mathcal{M}). \end{aligned}$$

2. Théorème des trois perpendiculaires.
Si W_1 et W_2 désignent deux sous-espaces vectoriels de E, montrer que les trois conditions suivantes sont équivalentes :

$$\begin{aligned} &\text{(i)} \quad W_1 \subseteq W_2, \\ &\text{(ii)} \quad W_2^\perp \subseteq W_1^\perp, \\ &\text{(iii)} \quad P_{W_1} = P_{W_1} \circ P_{W_2}. \end{aligned}$$

Remarque : cette propriété est appelée "théorème des trois perpendiculaires" ; en effet, la condition (iii) s'interprète géométriquement par l'existence de trois angles droits, comme le montre l'exemple suivant :



3. Inertie du nuage projeté

3.1. Soit $\mathcal{N}$ le nuage $\mathcal{M}$ projeté sur W, c'est-à-dire $\mathcal{N} = P_W(\mathcal{M})$, et soit Δu un axe du sous-espace vectoriel W. Montrer[†] que l'inertie de $\mathcal{N}$ par rapport à $(\Delta u)^\perp$ est égale à l'inertie du nuage $\mathcal{M}$ par rapport à $(\Delta u)^\perp$. En déduire que le premier axe principal du nuage $\mathcal{N}$ est l'axe de W à inertie minimum pour le nuage $\mathcal{M}$.

3.2. Montrer que :

$$I_{\Delta u}(\mathcal{N}) = I_{\Delta u}(\mathcal{M}) + I_T(\mathcal{N}) - I_T(\mathcal{M}),$$

avec :

- $I_{\Delta u}(\mathcal{N})$: inertie de $\mathcal{N}$ par rapport à Δu,
- $I_{\Delta u}(\mathcal{M})$: inertie de $\mathcal{M}$ par rapport à Δu,
- $I_T(\mathcal{N})$ inertie totale de $\mathcal{N}$,
- $I_T(\mathcal{M})$ inertie totale de $\mathcal{M}$.

†. on utilisera de préférence le théorème des trois perpendiculaires.

TD 4
Exemple d'analyse en composantes principales

On considère le tableau de données suivant, noté X :

$I \setminus J$	1	2	3	4	5	6
x	1	0	0	2	1	2
y	0	0	1	2	0	3
z	0	1	2	1	0	2

associé aux résultats de trois variables x, y et z mesurées sur un échantillon J de six individus. On suppose que chaque individu j de J ($1 \leq j \leq 6$) est muni de la masse $1/6$.

1re Partie

On désire effectuer l'Analyse en composantes principales (A.C.P.) de X sur matrice variance (i.e. en supposant $\mathbb{R}^3$ muni de la métrique identité). On dit encore que l'on effectue une A.C.P. non normée.

1. Quel est le tableau Y centré associé à X ?

2. Donner la matrice variance V associée au tableau X.

3. Montrer que le vecteur $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$ est vecteur propre de V relatif à la valeur propre nulle. Qu'en déduit-on pour la représentation de J ?

4. Calculer les axes factoriels non triviaux (i.e. relatifs à une valeur propre non nulle) associés au tableau X, les valeurs propres et les pourcentages d'inertie correspondants.

5. A.C.P. (non normée) du nuage des individus.

5.1. Résultats numériques. Dans un tableau dont les colonnes représentent les 6 individus, indiquer les résultats suivants :

a) les valeurs des variables centrées ; on indiquera sur 2 colonnes supplémentaires les coordonnées des deux axes factoriels non triviaux (ces résultats sont donnés sur les 3 premières lignes),

b) les valeurs des deux composantes principales (2 lignes suivantes),

c) la valeur du coefficient INR (1 ligne),

d) les valeurs des contributions CTR et COR (4 lignes).

On rappellera la définition et l'intérêt des coefficients $\mathrm{INR}(j)$, $\mathrm{CTR}_\alpha(j)$ et $COR_\alpha(j)$.

5.2. Résultats graphiques. Donner la représentation graphique du nuage des individus dans le plan euclidien des deux premiers axes principaux.

6. A.C.P. (non normée) du nuage des variables. On se limitera à calculer les covariances et les corrélations des trois variables x, y et z avec les deux facteurs non triviaux, et on donnera la représentation graphique de ces trois variables sur le cercle de corrélations.

2^e Partie

On désire maintenant faire l'A.C.P. normée des données, c'est-à-dire on désire effectuer l'A.C.P. sur matrice de corrélation.

7. Calculer la matrice de corrélation R.

8. Donner le tableau centré réduit Y associé à R.

9. A.C.P. normée. On donnera les résultats suivants :

a) pour le nuage des individus, calculer les axes principaux d'inertie et les valeurs propres associées,

b) pour le nuage des individus et celui des variables, donner les composantes principales,

c) représenter les individus sur le plan des deux premiers axes factoriels et représenter également les variables sur le cercle des corrélations.

Pour les calculs, on pourra adopter la présentation de la question **5.1**. On comparera les résultats obtenus à ceux des questions **5.** et **6.**

3^e Partie

On désire à présent faire l'A.C.P. des données en utilisant la métrique de $\mathbb{R}^3$ dont la matrice est $M = \text{Diag}(a, b, c) = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$. On dit encore que l'on effectue une A.C.P. avec la métrique M.

10. Quelles conditions doivent vérifier les trois réels a, b, c pour que M soit une métrique ?

11. Préciser la matrice dont les vecteurs propres sont les axes factoriels de l'A.C.P.

12. Calculer cette matrice en fonction de a, b, c et montrer qu'elle est singulière. Que peut-on en déduire pour la représentation du nuage des points ? Montrer que les valeurs propres sont toutes distinctes.

Par la suite, sauf indication contraire, on prendra $a = c$.

13. Calculer en fonction des réels a et b les axes factoriels de l'A.C.P. ainsi que la part d'inertie du nuage qu'ils expliquent.

14. Calculer en fonction des réels a et b les contributions CTR des variables sur les axes factoriels.

15. Calculer les corrélations entre chaque variable et chaque composante principale ainsi que les contributions COR.

16. Quelle est la formule qui permet de calculer les composantes principales ?

17. Déduire des questions précédentes les valeurs des composantes principales du nuage de points, et calculer les CTR et les COR pour les individus.

18. Application Numérique. Effectuer la représentation graphique du nuage des points dans le premier plan principal en supposant que $M = \mathrm{Diag}(a = 1, b = 2, c = 1)$.

19. Pour quelles valeurs convenables des réels a, b et c, peut-on déduire des questions précédentes les résultats de l'A.C.P. normée des données proposées ?

20. Comment peut-on procéder pour appliquer les résultats de cet exercice lorsque la métrique M n'est pas définie par une matrice diagonale ?

TD 5

A.F.T.D. : Analyse Factorielle d'un Tableau de Distances

On considère un nuage de n points, que l'on note $\mathcal{N} = \{P_1, \ldots, P_n\}$. Chacun de ces n points P_j est muni du poids p_j. On suppose que pour tout j et j', on connaît le carré de la distance, noté $d^{jj'}$, entre les points P_j et $P_{j'}$. Par conséquent :

$$d^{jj'} = d^{j'j} \geq d^{jj} = 0,$$

pour tout $j, j' \in \{1, \ldots, n\}$. Si M désigne la métrique associée à cette distance, on en déduit :

$$(P_jP_{j'})'\mathrm{M}(P_jP_{j'}) = \|P_jP_{j'}\|^2 = d^{jj'}.$$

On adopte les notations suivantes où G désigne le centre de gravité du nuage $\mathcal{N}$, et p la masse totale, c'est-à-dire la somme de tous les p_j :

$$\text{Pour tout } j, \qquad d^{\,\bullet j} = \sum_{j'=1}^{n} \frac{p_{j'}}{p} d^{jj'}, \quad d^{\,\bullet\bullet} = \sum_{j=1}^{n} \frac{p_j}{p} d^{\,\bullet j} \quad \text{et} \quad \mathbf{D_p} = \mathrm{Diag}(p_j).$$

Par la suite, on désire effectuer une *Analyse Factorielle sur Tableau de Distances*, c'est-à-dire représenter (le mieux possible) le nuage $\mathcal{N}$ en ayant pour seules données les valeurs $d^{jj'}$, c'est-à-dire les carrés des distances entre les points. Rappelons que pour une A.C.P., les données sont constituées par les valeurs prises par les variables sur les n individus.

1$^{\text{ère}}$ partie : Etude dans le cas général. Dans cette question, on suppose que l'on connaît la matrice $\mathbf{Y}$ (centrée) du type varibles $\times$ individus et que l'on affectué l'ACP du nuage $\mathcal{N}$ avec la métrique M. On rappelle que compte tenu des notations précédentes, la matrice $\mathbf{Y}$ peut s'écrire sous la forme $\mathbf{Y} = \big(GP_1, \ldots, GP_n\big)$.

1. Equation vérifiée par les composantes principales F.

1.1. Soit $\mathbf{u}$ le vecteur normé dirigeant l'axe factoriel associé à la valeur propre non nulle λ et soit $\mathbf{F}$ la composante principale qui lui est associée. Rappeler pourquoi l'on a :

$$\begin{array}{ll} \text{a)} & \mathbf{Y}\,\mathbf{D_p}\mathbf{Y}'\,\mathrm{M}\,\mathbf{u} = \lambda\,\mathbf{u}, \\ \text{b)} & \mathbf{F} = \mathbf{Y}'\mathrm{M}\,\mathbf{u}, \end{array}$$

et en déduire que : $\left\{ \begin{array}{l} \mathbf{Y}'\mathrm{M}\,\mathbf{Y}\,\mathbf{D_p}\,\mathbf{F} = \lambda\,\mathbf{F} \\ \text{avec } \mathbf{D_p}(\mathbf{F},\mathbf{F}) = \lambda \end{array} \right.$

1.2. Déduire de **1.1.** que $\mathbf{F}$ est solution de l'équation, notée (1), définie par :

$$\text{Pour tout } j \in \{1, \ldots, n\}, \qquad \sum_{j'=1}^{n} p_{j'}(GP_j)'\,\mathrm{M}\,(GP_{j'})\,\mathbf{F}(j') = \lambda\,\mathbf{F}(j), \tag{1}$$

$$\text{avec} \sum_{j=1}^{n} p_j\,\mathbf{F}^2(j) = \lambda.$$

1.3. De la relation $\sum_{j=1}^{n} p_j\,(GP_j) = 0$, déduire que (1) possède une solution associée à la valeur propre $\lambda = 0$.

2. Calcul des termes de l'équation (1) en fonction des valeurs des $d^{jj'}$.

2.1. En appliquant le théorème de Huyghens †, et en notant I_G l'inertie du nuage $\mathcal{N}$ par rapport à G, montrer que :

$$p\,d^{\,\bullet j} = I_G + p\,\|GP_j\|^2$$

2.2. Déduire de **2.1**. que :

$$I_G = \frac{1}{2}\,p\,d^{\,\bullet\bullet}$$

2.3. En utilisant **2.1.**, **2.2.** et le théorème de Pythagore généralisé ††, montrer que pour tout $j, j' \in \{1, \ldots, n\}$, on a :

$$(GP_j)'\mathrm{M}\,(GP_{j'}) = \frac{1}{2}\big(d^{\,\bullet j} + d^{\,\bullet j'} - d^{\,\bullet\bullet} - d^{jj'}\big).$$

2$^{\text{ème}}$ partie : Application.

On suppose $n = 4$, et que $\begin{cases} d^{12} = d^{23} = d^{34} = d^{41} = a^2 \\ d^{13} = d^{24} = b^2 \\ p_1 = p_2 = p_3 = p_4 = 1 \end{cases}$

1. Montrer qu'outre la valeur propre nulle, (1) admet une valeur propre simple et une valeur propre double.

2. Donner la représentation du nuage en projection sur l'axe factoriel correspondant à la valeur propre simple, et sur le plan factoriel correspondant à la valeur propre double.

3. Indiquer les relations que doivent vérifier les nombres a et b pour que le nuage $\mathcal{N}$ soit représentable :

α) dans un espace euclidien,
β) dans un plan,
γ) sur une droite.

†. i.e. la relation $I_A(\mathcal{N}) = I_G(\mathcal{N}) + p\,\|GA\|^2$, où A désigne un point quelconque.
††. i.e. la relation $\|BC\|^2 = \|AB\|^2 + \|AC\|^2 - 2\,(AB)'\mathrm{M}\,(AC)$ vérifiée pour tout triangle ABC.

TD 6
Interprétation d'une Analyse en Composantes Principales normée

Le tableau des données, ci-dessous, indique les dépenses annuelles moyennes de consommation de douze catégories socio-professionnelles. Les douze individus du tableau, qui sont ici présentés en ligne, désignent des catégories socio-professionnelles caractérisées par la CSP variable qualitative à trois modalités (1 = manoeuvre, 2 = employé et 3 = cadre) et le nombre d'enfants (2, 3, 4 ou 5). Par exemple, l'individu MA2 désigne la catégorie "manoeuvre avec deux enfants", EM2 la catégorie "employé avec deux enfants", CA2 la catégorie "cadre avec deux enfants", etc. Les sept premières variables, qui sont présentées ici en colonne, sont les variables actives. Ces variables désignent respectivement les dépenses annuelles moyennes de consommation en pain, légumes, fruits, viande, volaille, lait et vin. On effectue l'ACP normée (ACPN) de ce tableau en considérant comme supplémentaires les variables "CSP" et "enfants". Les résultats ont été obtenus à l'aide du logiciel *R* et sont présentés en Annexe.

IDEN	pain	légumes	fruits	viande	volaille	lait	vin	CSP	enfants
MA2	332	428	354	1437	526	247	427	1	2
EM2	293	559	388	1527	567	239	258	2	2
CA2	372	767	562	1948	927	235	433	3	2
MA3	406	563	341	1507	544	324	407	1	3
EM3	386	608	396	1501	568	319	363	2	3
CA3	438	843	689	2345	1148	243	341	3	3
MA4	534	660	367	1620	638	414	407	1	4
EM4	460	699	484	1856	762	400	416	2	4
CA4	385	789	621	2366	1149	304	282	3	4
MA5	655	776	423	1848	759	495	486	1	5
EM5	584	995	548	2056	893	518	319	2	5
CA5	515	1097	887	2630	1167	561	284	3	5

1. Commenter brièvement le tableau des statistiques sommaires.

2. Que peut-on dire des corrélations entre variables ?

3. Que vaut l'inertie totale dans l'ACP normée effectuée ?

4. Calculer la contribution relative $\mathrm{INR}(i)$ de chaque variable x_i à l'inertie totale I_T.

5. Combien d'axes factoriels pensez-vous qu'il faille retenir ?

6. Calculer la contribution $\mathrm{CTR}_\alpha(i)$ de chaque variable x_i au $\alpha^{\text{ème}}$ facteur ($\alpha = 1, 2$).

7. Calculer le carré de la corrélation $\mathrm{COR}_\alpha(i)$ de chaque variable x_i avec le $\alpha^{\text{ème}}$ facteur ($\alpha = 1, 2$), puis la qualité de la représentation associée $\mathrm{QLT}_2(i)$ dans le plan factoriel 1×2.

8. Quel est l'intérêt de ranger les variables par corrélation décroissante avec le premier facteur (ou première composante principale) pour examiner la matrice de corrélation ?

9. Quelles sont les variables caractéristiques du premier axe factoriel ?

10. Quelles sont les variables caractéristiques du deuxième axe factoriel ?

11. Interpréter le cercle des corrélations.

12. Quelles sont les observations caractéristiques du premier axe factoriel (on se basera sur les contributions $\text{CTR}_1(j)$ de chaque observation j au premier axe factoriel).

13. Donner la qualité de représentation de chaque observation j sur le plan des deux premiers axes factoriels.

14. Quelles sont les observations caractéristiques du deuxième axe factoriel ?

15. Interpréter le plan des deux premiers axes factoriels (on aura intérêt à joindre tous les points associés à une famille ayant le même nombre d'enfants et, de même, tous les points associés à une même CSP).

ANNEXE

A1 — *Statistiques élémentaires.*

Le logiciel R permet de générer les statistiques élémentaires univariées et bivariées. Le premier tableau contient les statistiques univariées suivantes : valeur minimale, premier quartile, médiane, moyenne, troisième quartile et valeur maximale. Le second contient les corrélations entre les sept variables actives.

	pain	legum.	fruits	viande	volaille	lait	vin
Min.	293.0	428.0	341.0	1437	526.0	235.0	258.0
Qu. 1	381.8	596.8	382.8	1522	567.8	246.0	310.2
Med.	422.0	733.0	453.5	1852	760.5	321.5	385.0
Mean	446.7	732.0	505.0	1887	804.0	358.2	368.6
Qu. 3	519.8	802.5	576.8	2128	982.2	434.2	418.8
Max.	655.0	1097.0	887.0	2630	1167.0	561.0	486.0
s-dv.	102.59	181.13	158.06	378.90	238.10	112.14	68.73

TABLE 1 – Statistiques élémentaires univariées sur les variables actives

	pain	legumes	fruits	viande	volaille	lait	vin	CSP	enfants
pain	1.00	0.59	0.20	0.32	0.25	0.86	0.30	-0.22	0.88
legumes	0.59	1.00	0.86	0.88	0.83	0.66	-0.36	0.60	0.71
fruits	0.20	0.86	1.00	0.96	0.93	0.33	-0.49	0.82	0.40
viande	0.32	0.88	0.96	1.00	0.98	0.38	-0.44	0.78	0.53
volaille	0.25	0.83	0.93	0.98	1.00	0.23	-0.40	0.82	0.42
lait	0.86	0.66	0.33	0.38	0.23	1.00	0.01	-0.12	0.93
vin	0.30	-0.36	-0.49	-0.44	-0.40	0.01	1.00	-0.57	-0.05
CSP	-0.22	0.60	0.82	0.78	0.82	-0.12	-0.57	1.00	0.00
enfants	0.88	0.71	0.40	0.53	0.42	0.93	-0.05	0.00	1.00

TABLE 2 – Corrélations entre toutes les variables (actives ou passives)

A2 — *Résultats de l'ACP normée.*

Les principaux résultats de l'ACP générés par le package "FactoMineR" du logiciel R sont :
- les valeurs propres
- Pour les variables : coordonnées, corrélation variable-facteur et contributions ;
- Pour les individus : coordonnées, qualité de représentation et contributions.

a) *Valeurs propres*

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	4.3337	61.9100	61.9100
comp 2	1.8310	26.1576	88.0676
comp 3	0.6301	9.0012	97.0688
comp 4	0.1281	1.8295	98.8982
comp 5	0.0573	0.8185	99.7168
comp 6	0.0187	0.2678	99.9845
comp 7	0.0011	0.0155	100.0000

b) *Corrélation variable-facteur*

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
pain	0.50	0.84	-0.01	-0.19	0.01
legumes	0.97	0.13	-0.05	-0.01	-0.19
fruits	0.93	-0.28	0.12	0.20	-0.02
viande	0.96	-0.19	0.16	-0.02	0.10
volaille	0.91	-0.27	0.28	-0.12	0.05
lait	0.58	0.71	-0.35	0.16	0.08
vin	-0.43	0.65	0.62	0.11	-0.02

c) *Coordonnées des variables*

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
pain	0.50	0.84	-0.01	-0.19	0.01
legumes	0.97	0.13	-0.05	-0.01	-0.19
fruits	0.93	-0.28	0.12	0.20	-0.02
viande	0.96	-0.19	0.16	-0.02	0.10
volaille	0.91	-0.27	0.28	-0.12	0.05
lait	0.58	0.71	-0.35	0.16	0.08
vin	-0.43	0.65	0.62	0.11	-0.02

d) *Contributions des variables*

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
pain	5.73	38.71	0.01	29.65	0.17
legumes	21.70	0.97	0.38	0.07	65.54
fruits	19.93	4.20	2.14	29.86	0.44
viande	21.36	1.98	4.31	0.28	17.23
volaille	19.17	3.89	12.57	10.38	4.83
lait	7.87	27.32	19.67	20.43	11.30
vin	4.24	22.93	60.92	9.32	0.49

e) *Coordonnées des individus*

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
MA2	-2.99	-0.38	0.42	0.38	0.23
EM2	-1.98	-1.87	-1.37	-0.17	-0.10
CA2	-0.12	-0.76	1.48	0.20	-0.47
MA3	-2.13	0.34	-0.11	0.11	0.01
EM3	-1.75	-0.18	-0.52	0.15	-0.18
CA3	1.77	-1.42	1.04	-0.45	-0.07
MA4	-0.98	1.43	-0.29	-0.27	0.10
EM4	-0.27	0.66	0.28	0.30	0.17
CA4	1.67	-1.81	0.10	-0.42	0.44
MA5	0.23	2.90	0.59	-0.26	0.13
EM5	2.04	1.18	-1.03	-0.34	-0.34
CA5	4.51	-0.10	-0.59	0.75	0.08

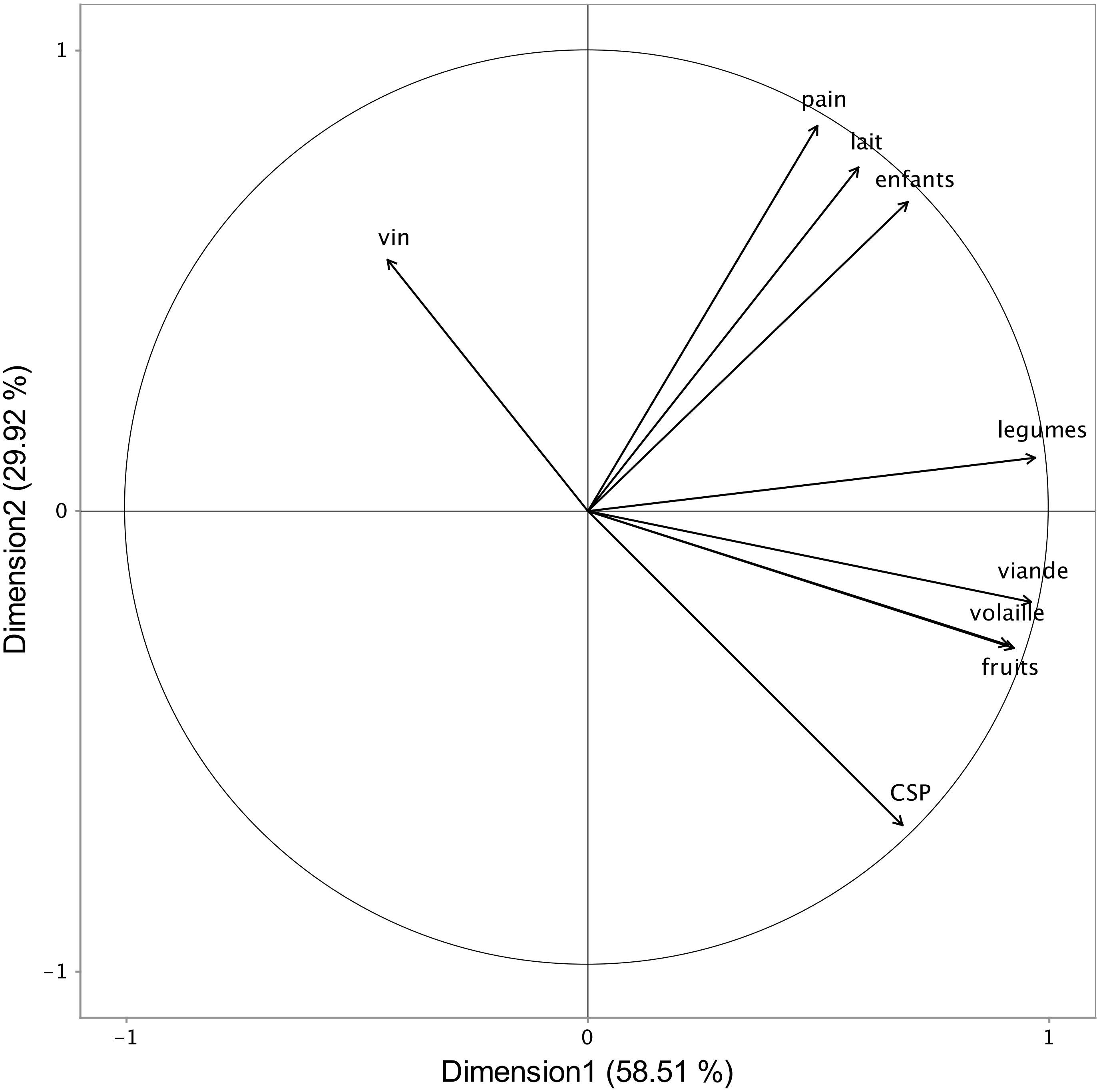
f) *Qualité de représentation des individus* ($\cos^2$)

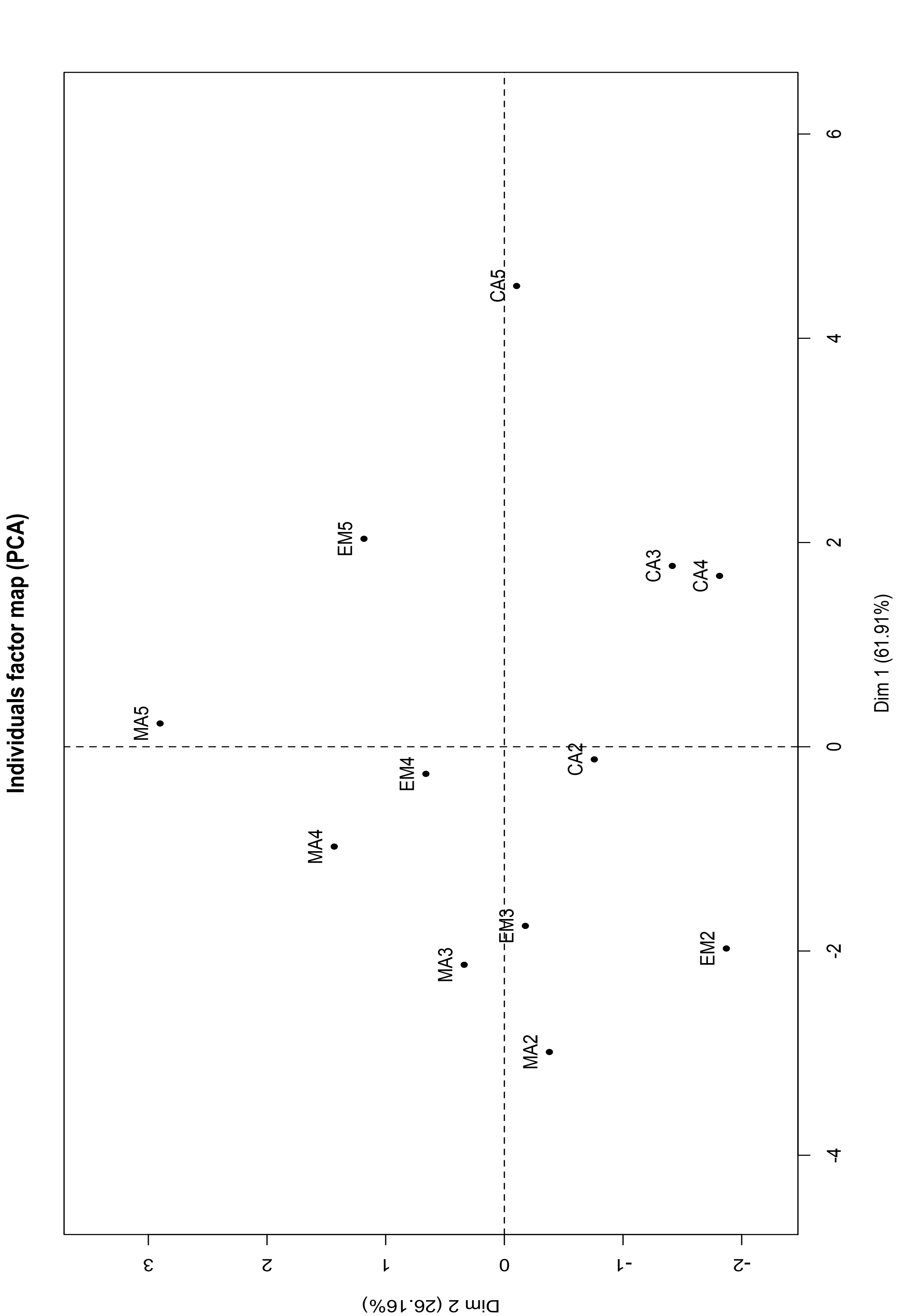
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
MA2	0.94	0.02	0.02	0.02	0.01
EM2	0.42	0.38	0.20	0.00	0.00
CA2	0.00	0.19	0.72	0.01	0.07
MA3	0.97	0.02	0.00	0.00	0.00
EM3	0.89	0.01	0.08	0.01	0.01
CA3	0.48	0.31	0.17	0.03	0.00
MA4	0.30	0.64	0.03	0.02	0.00
EM4	0.10	0.61	0.11	0.13	0.04
CA4	0.43	0.50	0.00	0.03	0.03
MA5	0.01	0.95	0.04	0.01	0.00
EM5	0.60	0.20	0.16	0.02	0.02
CA5	0.96	0.00	0.02	0.03	0.00

g) *Contributions des individus*

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
MA2	17.18	0.66	2.32	9.40	7.92
EM2	7.50	15.93	24.73	1.77	1.39
CA2	0.03	2.61	29.14	2.68	31.46
MA3	8.76	0.52	0.17	0.79	0.01
EM3	5.92	0.14	3.62	1.52	4.56
CA3	6.03	9.12	14.28	13.19	0.81
MA4	1.84	9.35	1.11	4.86	1.39
EM4	0.14	1.99	1.07	5.91	4.01
CA4	5.38	14.96	0.14	11.31	28.22
MA5	0.10	38.30	4.63	4.28	2.40
EM5	7.98	6.37	14.13	7.45	16.90
CA5	39.14	0.05	4.65	36.83	0.91

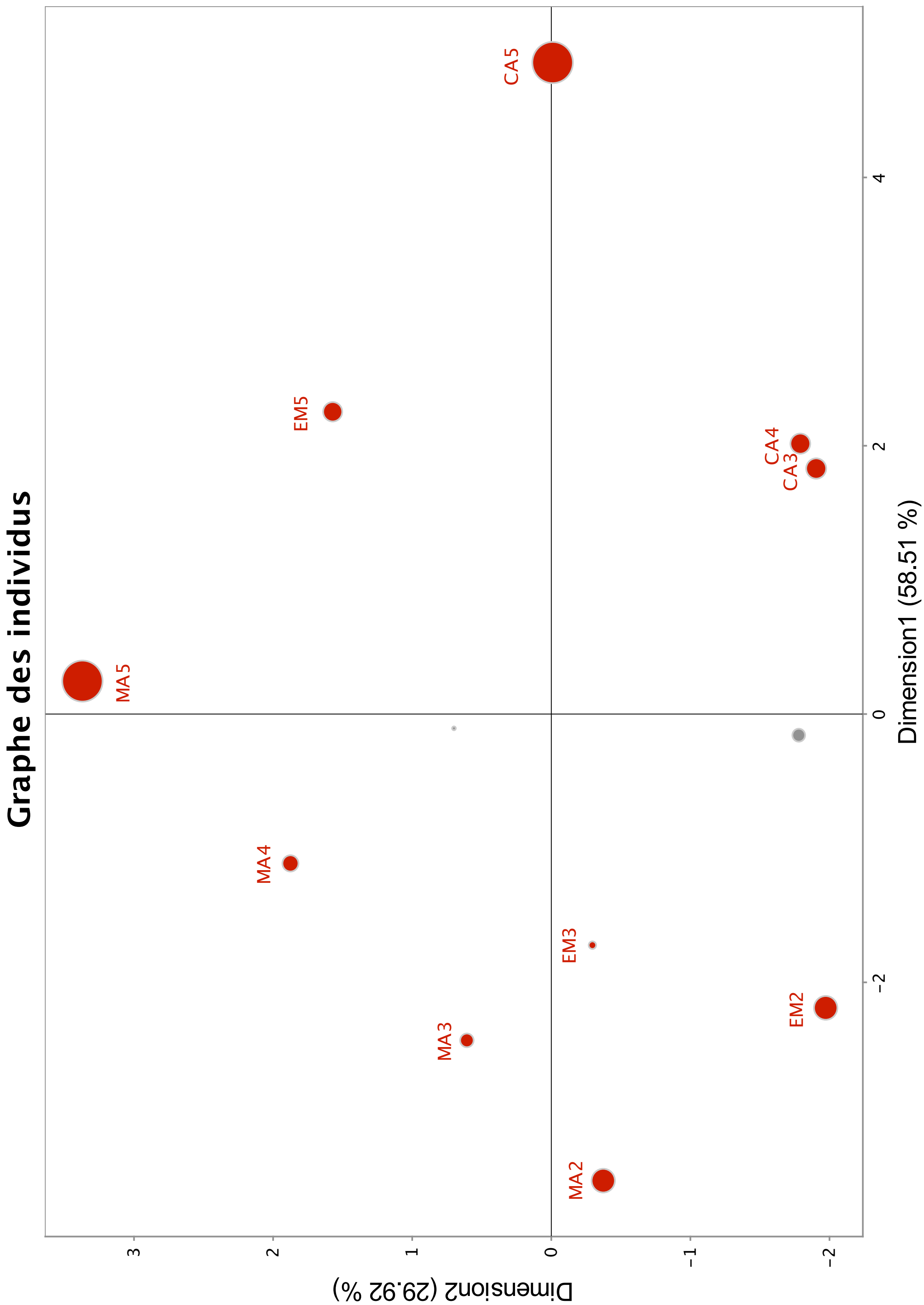
Graphe des variables





Individuals factor map (PCA)
Dim 1 (61.91%)
Dim 2 (26.16%)
-4
-2
0
2
4
6
-2
-1
0
1
2
3
MA2
MA3
EM3
EM2
MA4
EM4
CA2
MA5
CA3
CA4
EM5
CA5

Affichage des individus ayant un indice $QLT_2 > .75$ par des disques proportionnels à leurs CTR_{12}.



A3 – *Programme en R.*

```
> library(FactoMineR)
> depense <- read.table("depense.txt")
> library(dynGraph)
> res.pcadep <- PCA(depense)
> dynGraph(res.pcadep)
```

TD 7

AFC − Analyse Factorielle des Correspondances

1$^{\text{ère}}$ Partie : rappels et généralités

Rappels. L'Analyse Factorielle des Correspondances (AFC) a pour but d'analyser (et même de visualiser) un tableau K de nombres positifs. Nous nous plaçons dans le cas usuel où K est un *tableau de contingence* : étant donné deux ensembles, notés $I = \{1, 2, \ldots, p\}$ et $J = \{1, 2, \ldots, q\}$, chacun d'eux décrivant les modalités prises par une variable qualitative, le terme général k_{ij} de K est égal à l'effectif des individus ayant pris simultanément les modalités i et j relativement aux deux variables. Le principe de l'AFC consiste à effectuer deux ACPs, l'une sur le nuage $\mathcal{N}(I)$ constitué des profils lignes de K, l'autre sur le nuage $\mathcal{N}(J)$ constitué des profils colonnes de K. Le $i^{\text{ème}}$ individu du nuage $\mathcal{N}(I)$, appelé *profil de la $i^{\text{ème}}$ ligne de K*, est égal à la distribution (empirique) de J lorsque l'on suppose que la modalité i de l'autre variable est réalisée. Ce profil, qui est donc un vecteur de $\mathbb{R}^q$, est noté f_J^i et s'obtient en divisant les q effectifs de la $i^{\text{ème}}$ ligne de K par le total, noté $k_{i\,\bullet}$, de cette $i^{\text{ème}}$ ligne. Le profil f_J^i est muni du poids $f_{i\,\bullet} = \dfrac{k_{i\,\bullet}}{k}$, le nombre k étant le total des termes du tableau K, i.e. l'effectif total des individus. Par conséquent, le poids $f_{i\,\bullet}$ est aussi la probabilité (empirique) que la $i^{\text{ème}}$ modalité soit réalisée. Rappelons que le nuage $\mathcal{N}(I)$ est muni de la métrique $D_{1/f_J} = \mathrm{Diag}\big(1/f_{\bullet j}\big)_{j\in J}$, c'est-à-dire la métrique, appelée *métrique du Khi-deux*, qui est définie par la matrice diagonale d'ordre q dont le terme général est $\dfrac{1}{f_{\bullet j}}$. Les caractéristiques du nuage $\mathcal{N}(J)$ sont définies de façon similaire : le $j^{\text{ème}}$ individu est le *profil de la $j^{\text{ème}}$ colonne de K*, noté f_I^j, et les coordonnées de ce vecteur de $\mathbb{R}^p$ sont obtenues en divisant les p effectifs de la $j^{\text{ème}}$ colonne de K par le total, noté $k_{\bullet j}$, de cette $j^{\text{ème}}$ colonne. Le poids associé à ce profil est égal à $f_{\bullet j} = \dfrac{k_{\bullet j}}{k}$, qui s'interprète comme la probabilité que j soit réalisée. La métrique du nuage, appelée aussi *métrique du Khi-deux*, est la métrique associée à la matrice $D_{1/f_I} = \mathrm{Diag}\big(1/f_{i\,\bullet}\big)_{i\in I}$.

1. Calculer les coordonnées du centre de gravité, noté g_J, du nuage $\mathcal{N}(I)$; sans faire de calcul, donner par symétrie les coordonnées du centre de gravité, noté g_I, du nuage $\mathcal{N}(J)$.

2. On note $d^2(i, i')$ la distance (du Khi-deux) entre i et i', c'est-à-dire la distance entre les profils lignes i et i' selon la métrique D_{1/f_J}. Exprimer $d^2(i, i')$ en fonction des quantités $f_j^i, f_j^{i'}$ et $f_{\bullet j}$, où j varie de 1 à k. Par symétrie, donner sans calculs l'expression de la distance (du Khi-deux) entre j et j', c'est-à-dire la distance $d^2(j, j')$ entre les profils colonnes j et j' selon la métrique D_{1/f_I}.

3. En considérant le nuage $\mathcal{N}(I)$, calculer l'inertie totale I_T en fonction de $f_j^i, f_{i\,\bullet}$ et $f_{\bullet j}$, où i varie de 1 à p et j varie de 1 à q. Par symétrie et en considérant le nuage $\mathcal{N}(J)$, donner sans calculs une seconde expression de I_T en fonction des quantités $f_i^j, f_{i\,\bullet}$ et $f_{\bullet j}$.

4. En notant $x_\alpha(i)$ (resp. $y_\alpha(j)$) la $\alpha^{\text{ème}}$ composante principale du profil de la $i^{\text{ème}}$ ligne (resp. $j^{\text{ème}}$ colonne), et en utilisant les formules de transition, exprimer $x_\alpha(i)$ en fonction de $\sqrt{\lambda_\alpha}, f_j^i$ et $y_\alpha(j)$, j variant de 1 à q. En supposant que la valeur de λ_α est constante et égale à λ, en déduire que le profil centré de la $i^{\text{ème}}$ ligne, i.e. $f_J^i - g_J$, est une combinaison linéaire (que l'on précisera) des profils centrés des colonnes, i.e. des vecteurs $f_I^j - g_I$.

$2^{\text{ème}}$ Partie : Application 1

On désire effectuer l'AFC du tableau K_{IJ} suivant :

$I \setminus J$	A	B	C	D	E	F
i_1	1	0	0	1	1	2
i_2	0	1	0	1	2	1
i_3	0	0	2	1	1	1

1. Calculer les marges de K_{IJ}.

2. On considère le nuage des profils-colonnes de K_{IJ}.

- **a)** Déterminer le poids de chaque élément j de J.
- **b)** Quelle est la métrique dont est muni l'espace $\mathbb{R}^3$?
- **c)** Calculer le tableau des profils-colonnes de K_{IJ}, et le centre de gravité g du nuage associé.
- **d)** Dans l'espace $\mathbb{R}^3$, on considère les points H_1, H_2, H_3 qui sont les extrémités des vecteurs de la base canonique de $\mathbb{R}^3$, i.e. les points de coordonées respectives $(1,0,0)$, $(0,1,0)$ et $(0,0,1)$. Placer les profils des points A, B, C, D, E et F dans le triangle $H_1H_2H_3$ ainsi que le centre de gravité G ($\overrightarrow{OG} = g$).
- **e)** Calculer (avec la métrique du Khi-deux) la longueur des côtés du triangle ABC. Que peut-on dire de ce triangle ?
- **f)** Combien y a-t-il d'axes factoriels non triviaux ?

3. On note $F_\alpha(i)$ (resp. $G_\alpha(i)$) ($\alpha = 1, 2$) l'abscisse de la projection du profil de la $i^{\text{ème}}$ ligne (resp. $j^{\text{ème}}$ colonne) sur le $\alpha^{\text{ème}}$ axe factoriel issu de l'analyse des correspondances de K_{IJ} qui est associé à la valeur propre λ_α. De plus, la relation suivante est ici vérifiée :

$$F_1 = \sqrt{\frac{\lambda_1}{2}} \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = \sqrt{\lambda_1}\, \varphi_1^I .$$

- **a)** À l'aide de la formule de transition, déterminer les valeurs de $G_1(j)$ pour $j \in J$, et en déduire la valeur propre λ_1.
- **b)** Donner le facteur φ_2^I de variance 1 (on supposera $\varphi_2^{i_1} > 0$).
- **c)** Calculer de même les valeurs de $G_2(j)$ pour $j \in J$, et λ_2.
- **d)** Déduire de **a)**, **b)** et **c)** les valeurs de $F_1(i)$ et $F_2(i)$ pour $i \in I$.
- **e)** Rappeler la définition et l'intérêt des contributions CTR_α et calculer ces contributions pour tout $\alpha \in \{1, 2\}$ et pour tous les éléments de I et J.
- **f)** Même question qu'en **e)** en remplaçant CTR par COR.
- **g)** Calculer les contributions INR pour tous les éléments de I et J.
- **h)** Effectuer la représentation simultannée de I et J dans le plan des axes factoriels 1 et 2, et interpréter cette représentation.
- **i)** Comment aurait-on pu, à partir de considérations de symétrie dans le plan engendré par H_1, H_2 et H_3, déterminer les axes factoriels du nuage des profils des colonnes du tableau K_{IJ} et en déduire les composantes principales G_1 et G_2, puis les valeurs propres et les composantes principales F_1 et F_2 ?

$3^{\text{ème}}$ Partie : Application 2

On désire effectuer l'AFC du tableau K suivant :

$I \setminus J$	A	B	C	D	E	F	G
α	1	0	0	0	1	1	1
β	0	1	0	1	0	1	1
γ	0	0	1	1	1	0	1

1. Etude du nuage $\mathcal{N}(\mathbf{I})$

1.1. Calculer les poids associés aux profils des lignes α, β et γ, ainsi que le carré de la distance (du Khi-deux) entre α et β, β et γ, α et γ.

1.2. En déduire que :

a) les deux valeurs propres non triviales λ_1 et λ_2 issues de l'AFC de K, ont la même valeur que l'on notera par la suite λ.

b) le centre de gravité g_J, que l'on précisera, est à égale distance des profils de α, β et γ.

1.3. Calculer la valeur de l'inertie totale I_T et en déduire la valeur de λ.

2. Etude du nuage $\mathcal{N}(\mathbf{J})$

2.1. Calculer les poids des sept éléments de J, ainsi que le carré de la distance (du Khi-deux) entre A et B, B et C, C et A.

2.2. Montrer que le centre de gravité du nuage $\mathcal{N}(\mathbf{J})$ est égal au profil de la colonne G.

3. Représentation du nuage $\mathcal{N}(\mathbf{J})$

3.1. En considérant le plan engendré par les trois points A, B, C, placer les trois points A, B, C, puis situer les quatre autres points D, E, F et G par rapport à A, B, C.

3.2. Placer sur le graphique le point a centre de gravité des quatre points A, E, F, G affectés tous les quatre de la masse 1.

3.3. Donner la valeur numérique du rapport $\dfrac{d(G,a)}{d(G,A)}$, où $d(G,a)$ (resp. $d(G,A)$) désigne la distance du Khi-deux entre G et a (resp. G et A).

4. Représentation du nuage $\mathcal{N}(\mathbf{I})$

4.1. En utilisant le résultat de la question **4.** de la partie 1, calculer le profil centré $f_J^\alpha - g_J$ en fonction du profil centré $f_J^a - g_I$, i.e. le vecteur $G\alpha$ en fonction du vecteur Ga. De même, exprimer le vecteur GA en fonction du vecteur $G\alpha$. En déduire la valeur de λ.

4.2. Placer sur le graphique les points α, β et γ, et donner la valeur de la longueur $G\alpha$.

TD 8

ACM − Analyse factorielle des Correspondances Multiples

On considère un ensemble Q de questions. Pour toute question q de Q, on note J_q l'ensemble des modalités de réponse à la question q. On désigne par J l'union disjointe des J_q :

$$J = \bigcup_{q \in Q} J_q.$$

Soit I un ensemble de n individus ayant répondu à toutes les questions de Q.
Pour tout individu i de I et toute question q de Q, on suppose que l'individu i a adopté une et une seule modalité de réponse à la question q.

On rappelle que le tableau disjonctif complet k_{IJ} associé à ce questionnaire a pour terme général $k(i,j)$ défini par :

$$\forall i \in I, \;\; \forall j \in J, \;\; k(i,j) = \begin{cases} 1 & \text{si l'individu } i \text{ a adopté la modalité } j \text{ de } J_q, \\ 0 & \text{sinon.} \end{cases}$$

On note :

- k la somme des termes du tableau k_{IJ},
- $k(i)$ la somme des termes de la i ème ligne de k_{IJ},
- $k(j)$ la somme des termes de la j ème colonne de k_{IJ},
- F le tableau des fréquences f_{ij} associé au tableau k_{IJ} (i.e. $f_{ij} = k(i,j)/k$),
- F_1 la matrice des profils des colonnes de k_{IJ},
- F_2 la matrice des profils des lignes de k_{IJ}.

On rappelle que le tableau de Burt associé au tableau k_{IJ}, noté b_{IJ} ou plus simplement B est défini par :

$$\forall (j,j') \in J^2, \;\; b(j,j') = \sum_{i \in I} k(i,j)k(i,j').$$

On note :

- $b(j)$ la somme des termes de la j ème colonne ou de la j ème ligne (en effet B est symétrique),
- $p(j)$ la proportion des individus ayant choisi la modalité j,
- $p(j,j')$ la proportion des individus ayant choisi les modalités j et j',
- $p^j_{j'}$ la proportion des individus ayant choisi la modalité j' parmi ceux qui ont choisi la modalité j.

1. Calcul de la matrice des profils

1.1. Calculer $k(i)$, $k(j)$, k, $b(j,j')$ et $b(j)$ en fonction de n, card Q, $p(j)$ et $p(j,j')$.

1.2. Montrer que $B = k^2 F'F$.

1.3. Montrer que la matrice des profils de colonnes de B et la matrice des profils de lignes de B sont toutes deux égales à la matrice F_2F_1.

2. On effectue l'analyse des correspondances du tableau b_{JJ}. On note $c_\alpha(j)$ l'abscisse de la projection du profil de la ligne j de B sur le $\alpha^{\text{ème}}$ axe factoriel et $d_\alpha(j)$ l'abscisse de la projection du profil de la colonne j. On supposera que toutes les valeurs propres sont simples.

2.1. Ecrire les équations aux valeurs propres vérifiées par les vecteurs c_α et d_α. En déduire que pour tout α, il existe $\epsilon_\alpha \in \{-1, 1\}$ tel que $d_\alpha = \epsilon_\alpha c_\alpha$.

2.2. En utilisant les formules de transition, montrer que pour tout α,

$$F_1' F_2' c_\alpha = \epsilon_\alpha \sqrt{\lambda_\alpha}\; c_\alpha,$$

où λ_α désigne la valeur propre associée à l'axe α. En déduire que $d_\alpha = c_\alpha$ pour tout α. Quelles relations existe t'il entre les résultats de l'AFC de b_{JJ} et ceux de l'AFC de k_{IJ} ?

2.3. Montrer que c_α vérifie la relation :

$$(1) \qquad \forall j \in J_q, \quad \sum_{j' \in J \setminus J_q} p^j_{j'} c_\alpha(j') = (cardQ\,\sqrt{\lambda}_\alpha - 1) c_\alpha(j).$$

3. Dans toute cette question, on suppose que l'on a que deux questions, autrement dit que $CardQ = 2$ et $J = J_1 \cup J_2$.

3.1. En utilisant les équations (1), écrire pour chaque valeur de j, l'équation vérifiée par $c_\alpha(j)$.

3.2. On effectue l'analyse des correspondances du tableau $p_{J_1 J_2}$ c'est-à-dire du tableau de terme général $p(j_1, j_2)$ avec $j_1 \in J_1$ et $j_2 \in J_2$. On note μ_β la valeur propre associée à l'axe β et $F_\beta(j_1)$ (resp. $G_\beta(j_2)$) l'abscisse de la projection du profil de la ligne $j_1 \in J_1$ (resp. de la colonne $j_2 \in J_2$). En utilisant les formules de transition, écrire pour chaque valeur de j_1 appartenant à J_1, l'équation vérifiée par $F_\beta(j_1)$. Même question pour $G_\beta(j_2)$ avec j_2 appartenant à J_2.

3.3. Déduire de ce qui précède que l'on peut trouver tous les résultats de l'AFC du tableau b_{JJ} à partir de ceux de l'AFC du tableau $p_{J_1 J_2}$.

3.4. Déduire de 2.2. et de 3.3. que l'on peut trouver tous les résultats de l'AFC du tableau disjonctif complet k_{IJ} à partir de ceux de l'AFC du tableau $p_{J_1 J_2}$.

4. Application numérique : sept personnes $i_1, i_2, \cdots, i_7$ ont été interrogées. Les deux questions posées étaient :

- $Q1$: Quel temps avez vous eu lors de vos dernières vacances ?
  Les réponses possibles sont : a : excellent, b : bon, c : moyen.
- $Q2$: Où avez-vous passé vos dernières vacances ?
  Les réponses possibles sont : A : à la montagne, B : à la mer.

La personne i_1 était à la montagne et le temps excellent.
La personne i_2 était à la mer et le temps bon.
La personne i_3 était à la montagne et le temps moyen.
La personne i_4 était à la montagne et le temps bon.
La personne i_5 était à la mer et le temps excellent.
La personne i_6 était à la mer et le temps moyen.
La personne i_7 était à la montagne et le temps excellent.

Faites l'AFCM du tableau disjonctif complet.

(Calculatrice autorisée, documents autorisés : 1 feuille recto-verso)

Barême approximatif : exercice 1 (8 pts) ; exercice 2 (12 pts).

Exercice 1

Soit $X = (x_i^j)_{\substack{i\in I\\ j\in J}}$ un tableau de données, avec $I = \{1,\ldots,p\}$ et $J = \{1,\ldots,n\}$. Par la suite, nous considérerons deux Analyses en Composantes Principales (ACP) :

- l'ACP du tableau X : les n individus sont alors les n vecteurs colonnes x^j ($j \in J$) de X, chacun muni du poids $1/n$, et la métrique M est alors l'identité de $\mathbb{R}^p$;
- l'ACP du tableau transposé X' : les individus sont alors les p vecteurs colonnes x_i ($i \in I$) de X', chacun muni du poids $1/p$, et la métrique M est alors l'identité de $\mathbb{R}^n$.

1) On rappelle que le tableau X est dit centré si chacune de ses variables x_i ($i \in I$) est centrée. De plus, pour tout entier positif k, on note $\mathbb{1}_k$ le vecteur de $\mathbb{R}^k$ dont toutes les coordonnées sont égales à 1. Montrer que X est centré si et seulement si $X\,\mathbb{1}_n$ est égal au vecteur nul. A quelle condition le tableau X' est-il centré ?

2) ***Par la suite, nous supposons que les tableaux X et X' sont centrés***. Montrer que l'ACP de X comporte un axe factoriel trivial (c.-à-d., un axe factoriel relatif à la valeur prore 0) pour lequel on précisera un vecteur directeur en indiquant ses coordonnées.

3) On note u^α un vecteur axial factoriel * (non trivial) de l'ACP de X qui est relatif à la α ème valeur propre λ_α. Montrer que $w^\alpha = \dfrac{1}{\sqrt{n\,\lambda_\alpha}}\, X' u^\alpha$ est un vecteur axial factoriel de l'ACP de X', et exprimer la valeur propre μ_α, relative à w^α, en fonction de n, p et λ_α.

4) Démontrer que les deux ACP de X et de X' admettent le même nombre d'axes factoriels non triviaux.

5) Soit η_i^α la covariance de la variable x_i avec la α ième composante principale ψ_α de l'ACP de X, et soit ζ_α la α ème composante principale de l'ACP de X'. Exprimer ζ_α en fonction de n, p et η^α.

---

*. On rappelle qu'un vecteur axial factoriel est un vecteur qui dirige un axe factoriel et qui est normé à 1 selon la métrique de l'ACP.

Exercice 2
On considère le tableau X de données suivant :

$$X = \begin{pmatrix} 3 & 4 & 1 & 2 & 1 & 0 & 3 & 2 \\ 3 & 4 & 1 & 2 & 5 & 4 & 3 & 2 \\ 3 & 1 & 7 & 5 & 3 & 5 & 3 & 5 \end{pmatrix},$$

où la i ème ligne désigne la variable x_i et où la j ème colonne désigne l'individu x^j. On effectue l'ACP de X sur matrice variance (i. e. avec la métrique identité de $\mathbb{R}^3$), chaque individu étant muni du poids $1/8$.

1) Calculer le tableau Y centré associé à X.

2) Exprimer la matrice variance V du tableau X en fonction de Y, puis calculer V.

3) Montrer qu'il existe un axe factoriel trivial (c.-à-d. relatif à la valeur propre 0) pour lequel on précisera un vecteur directeur : on pourra, par exemple, soit appliquer le résultat obtenu à la question 2) de l'exercice 1, soit montrer que les trois vecteurs lignes y_1, y_2 et y_3 de la matrice Y sont linéairement dépendants.

Par la suite, on admet que $\dfrac{3}{2}$ est valeur propre de V.

4) Donner la liste ordonnée des valeurs propres de V, et en déduire qu'il existe deux axes factoriels non triviaux.

5) Donner les pourcentages d'inertie expliqués par chacun des axes factoriels non triviaux.

6) Déterminer les deux axes factoriels non triviaux : pour chacun de ces deux axes, on précisera un vecteur axial factoriel qui sera choisi de façon à ce que sa première coordonnée soit négative.

7) † Le tableau suivant donne les résultats concernant le premier axe factoriel de l'ACP de X (sur matrice variance) pour tous les individus sauf pour x^2 et x^3 : déterminer les résultats portant sur ces deux individus en complétant les cases vides du tableau.

Individus	x^1	x^2	x^3	x^4	x^5	x^6	x^7	x^8
Composante principale $\psi_1(j)$	-1.225			1.225	-1.225	1.225	-1.225	1.225
Contribution $CTR_1(j)$ (%)	4.167			4.167	4.167	4.167	4.167	4.167
Cosinus carré $QLT_1(j)$	0.75			0.75	0.25	0.25	0.75	0.75

8) Calculer la valeur, en pourcentage, de la contribution $CTR_1(i)$ pour $i = 1$, c.-à-d. la contribution de la première variable à l'inertie de l'axe 1 ?

9) Calculer la qualité de représentation (ou cosinus carré) de la deuxième variable sur le deuxième axe.

†. Justifier et détailler les calculs effectués pour les questions 7), 8) et 9), en indiquant clairement les formules théoriques utilisées.

PARTIEL D'ANALYSE DES DONNÉES **M1 MMD** – 29 mars 2012

(Calculatrice autorisée, documents autorisés : 2 feuilles recto-verso)

Barême approximatif : exercice 1 (9 pts) ; exercice 2 (11 pts).

Remarque : dans chaque exercice, les analyses en composantes principales (ACP) sont effectuées sur matrice variance, donc avec la métrique M égale à l'identité.

Exercice 1

On considère un tableau X de données possédant 3 lignes et n colonnes. Le terme général x_i^j de X indique la valeur de la i ème variable prise par le j ème individu, avec $i \in I = \{1, 2, 3\}$ et $j \in J = \{1, \ldots, n\}$. On suppose que $n > 3$ et que les trois variables x_1, x_2 et x_3 vérifient la relation : $x_1 + 3\,x_2 + x_3 = 0$. On notera Y le tableau, de terme général y_i^j ($i \in I, j \in J$), qui est obtenu par centrage des données du tableau X.

Par ailleurs, on considère le tableau $Z = A\,X$, où A est la matrice définie par :

$$A = \begin{pmatrix} -1 & 2 & -2 \\ -2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}.$$

1. Montrer que $y_1 + 3\,y_2 + y_3 = 0$.

2. Déterminer un vecteur $u \neq 0$ tel que $Y'u = 0$.

3. Déduire de ce qui précède les coordonnées d'un vecteur directeur d'un axe factoriel trivial (c.-à-d. relatif à la valeur propre 0) de l'ACP de X.

4. Exprimer la matrice variance V_Z associée au tableau Z en fonction de la matrice variance V_X associée au tableau X.

5. Calculer la matrice $A'A$.

6. On note u_α (resp. w_α) le α ème vecteur axial factoriel de l'ACP de X (resp. de l'ACP de Z) et λ_α (resp. μ_α) la α ème valeur propre associée.
   Pour tout $\alpha \in \{1, 2, 3\}$, exprimer w_α en fonction de u_α (au signe près), et μ_α en fonction de λ_α.

7. Déduire, de ce qui précède, les coordonnées d'un vecteur directeur d'un axe factoriel trivial de l'ACP de Z.

Par la suite, on suppose que l'ACP de Z ne comporte qu'un seul axe factoriel trivial.

8. Pour tout $\alpha = 1, 2$, on note ζ_α (resp. Ψ_α) la α ème composante principale de l'ACP de Z (resp. de X). Exprimer ζ_α en fonction de Ψ_α.

9. Montrer que pour $\alpha = 1, 2$, on a :

$$\mathrm{cov}(z_1, \Psi_\alpha) + \mathrm{cov}(z_2, \Psi_\alpha) + 3\,\mathrm{cov}(z_3, \Psi_\alpha) = 0.$$

Exercice 2
On considère le tableau X de données suivant :

$$X = \begin{pmatrix} 3 & 2 & 5 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 & 1 & 1 \\ 0 & 5 & 2 & 3 & 3 & 5 \end{pmatrix},$$

où la i ème ligne désigne la variable x_i et la j ème colonne désigne l'individu x^j.
Dans ce qui suit, on examine les résultats générés par l'ACP de X (sur matrice variance).

1. Calculer les coordonnées du centre de gravité g du nuage constitué des vecteurs colonnes de X, et en déduire le tableau Y centré qui est associé à X.

2. Soit V la matrice variance du tableau X. Montrer que $V = \frac{1}{3}\begin{pmatrix} 9 & -2 & -5 \\ -2 & 2 & -2 \\ -5 & -2 & 9 \end{pmatrix}$.

3. Déterminer le rang de la matrice V, et en déduire que le nombre d'axes factoriels non triviaux (c.-à-d. relatifs à des valeurs propres non nulles) est égal à 2.

4. Déterminer les coordonnées d'un vecteur qui est vecteur directeur d'un axe factoriel trivial.

5. On note Ψ_1 la première composante principale. En tenant compte des valeurs particulières prises par les variables pour le 5 ème individu, déterminer la valeur de Ψ_1^5.

6. Sachant que $\Psi_1^1 = \Psi_1^3 = 2\sqrt{2}$ et $\Psi_1^2 = \Psi_1^4 = -\sqrt{2}$, en déduire la valeur de Ψ_1^6.

7. Par la suite, on note λ_1 et λ_2 (avec $\lambda_1 > \lambda_2$) les valeurs propres associées aux deux axes factoriels non triviaux. Calculer les valeurs de λ_1 et λ_2.

8. Calculer $CTR_1(j)$ pour $j = 1$ et $j = 2$, c'est-à-dire les contributions des individus 1 et 2 à l'inertie du premier axe.

9. Calculer $COR_\alpha(j)$ pour $j = 2$ et pour $\alpha \in \{1, 2\}$, c'est-à-dire la qualité de représentation de l'individu 2 sur chacun des axes factoriels non triviaux.

10. Calculer la covariance entre la première variable et la première composante principale. En déduire la valeur de $COR_\alpha(i)$ pour $i = 1$ et pour $\alpha \in \{1, 2\}$, c'est-à-dire la qualité de représentation de la première variable sur chacun des axes 1 et 2.

Calculatrice autorisée, documents autorisés : 2 feuilles recto-verso.

Barême approximatif : exercice 1 (10 pts) ; exercice 2 (10 pts).

Les exercices 1 et 2 de ce sujet peuvent être traités de façon indépendante.

Exercice 1

On considère un tableau X de données, comportant p lignes et n colonnes, et de terme général noté x_i^j avec $i \in I = \{1, \ldots, p\}$ et $j \in J = \{1, \ldots, n\}$. Pour tout $k \in \mathbb{N} \setminus \{0\}$, on note $\mathbb{1}_k$ le vecteur de $\mathbb{R}^k$ dont les coordonnées sont toutes égales à 1. On suppose que la condition suivante, notée (1), est vérifiée :

(1) Il existe une constante c telle que pour tout $j \in J$, on a $x_1^j + \ldots + x_p^j = c$.

1 - Soit $\mathcal{M}$ le nuage des n points x^j (j ème vecteur colonne de X) avec $j \in J$, chaque point x^j étant muni du poids $\dfrac{1}{n}$. On note g_X le centre de gravité de $\mathcal{M}$. Pour quelle valeur de la constante γ a-t'on $g_X = \gamma\, X\, \mathbb{1}_n$?

2 - Soit Y le tableau obtenu après avoir centré X. On rappelle l'égalité $Y = X - g_X \mathbb{1}_n'$. En utilisant cette égalité, et en notant V_X la matrice variance relative au nuage $\mathcal{M}$, montrer que $V_X = \dfrac{1}{n} XX' - g_X g_X'$.

3 - Montrer que (1) est équivalent à l'existence d'une constante c telle que $X'\mathbb{1}_p = c\, \mathbb{1}_n$.

4 - Montrer que le vecteur $\mathbb{1}_p$ dirige un axe factoriel de l'ACP sur matrice variance du tableau X. Préciser la valeur de l'inertie du nuage projeté sur cet axe.

5 - On pose $Z = Y'$. Soit $\mathcal{N}$ le nuage des p points y_i (i ème vecteur colonne de Y') avec $i \in I$, chaque point y_i étant muni du poids $\dfrac{1}{p}$. On note g_Z le centre de gravité du nuage $\mathcal{N}$. Montrer que $g_Z = 0$.

6 - Considérons les deux ACP sur matrice variance précédentes, c.-à-d. celle du tableau X et celle du tableau Z. Expliquer pourquoi ces deux ACP admettent le même nombre d'axes factoriels non triviaux.

7 - Soit u^α un vecteur axial factoriel (non trivial) de l'ACP sur matrice variance du tableau X. On note λ_α la valeur propre associée à u^α. Montrer que $Y'u^\alpha$ dirige un axe factoriel (non trivial) de l'ACP sur matrice variance du tableau Z. Préciser la valeur de l'inertie du nuage $\mathcal{N}$ projeté sur cet axe factoriel.

8 - Soit F_α (resp. G_α) la α ème composante principale de l'ACP sur matrice variance du tableau X (resp. Z). Exprimer G_α en fonction de n, λ_α, Y et F_α.

Exercice 2
On considère le tableau de données, noté X, et défini par :

$$X = \begin{array}{c|cccccc} & j_1 & j_2 & j_3 & j_4 & j_5 & j_6 \\ \hline i_1 & 1 & 1 & 2 & 0 & 3 & 1 \\ i_2 & 0 & 1 & 2 & 1 & 1 & 3 \\ i_3 & 1 & 2 & 4 & 1 & 4 & 4 \end{array}$$

où la i ème ligne désigne la variable x_i et la j ème colonne désigne l'individu x^j.
Par la suite, on considère les résultats de l'ACP sur matrice variance du tableau X.

1 - Calculer les coordonnées du centre de gravité g du nuage $\mathcal{M}$ constitué des vecteurs colonnes de X (munis du même poids 1/6), et en déduire le tableau Y centré qui est associé à X. On présentera Y sous la forme $Y = \frac{1}{3}Y_1$ où Y_1 est une matrice à coefficients entiers.

2 - Soit V la matrice variance du tableau X. Compléter les valeurs manquantes dans l'expression de la matrice V ci-dessous :

$$V = \frac{1}{18}\begin{pmatrix} 16 & 1 & 17 \\ 1 & 16 & ? \\ 17 & ? & ? \end{pmatrix}.$$

3 - Expliquer pourquoi le nombre d'axes factoriels non triviaux est égal à 2.

4 - Calculer l'inertie totale du nuage étudié.

5 - Montrer que $\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$ est un vecteur directeur d'un axe factoriel non trivial.

6 - Calculer le pourcentage d'inertie expliquée par l'axe factoriel déterminé à la question 5. Cet axe est-il le premier ou le second axe factoriel ?

7 - Déterminer les coordonnées du premier vecteur axial factoriel, noté u^1 (on choisira sa première coordonnée de façon à ce qu'elle soit positive).

8 - Calculer la première composante principale de l'individu j_2, notée $\Psi_1^{j_2}$.

9 - Calculer la contribution de l'individu j_2 à l'inertie du premier axe, notée $CTR_1(j_2)$.

10 - Calculer la qualité de représentation de l'individu j_2 sur le premier axe, notée $COR_1(j_2)$.

11 - Calculer la contribution de la variable i_1 à l'inertie du premier axe, notée $CTR_1(i_1)$.

Calculatrice autorisée, documents autorisés : 2 feuilles recto-verso.

Barême approximatif : 10 points pour chacun des deux exercices.

Exercice 1

Soit $X = \left(x_i^j\right)_{\substack{1 \leq i \leq p \\ 1 \leq j \leq n}}$ un tableau de données, comportant p lignes et n colonnes, dont le terme général x_i^j indique la valeur de la $i^{\text{ème}}$ variable pour le $j^{\text{ème}}$ individu. Le $j^{\text{ème}}$ individu est muni d'un poids p_j (avec $p_j > 0$ pour tout j et $\sum_{j=1}^{n} p_j = 1$). On note $I = \{1, \ldots, p\}$, $J = \{1, \ldots, n\}$ et r le rang de la matrice variance V_X de X. Les résultats de l'ACP de X, l'espace $\mathbb{R}^p$ étant muni d'une métrique M, sont notés comme suit :

- le $\alpha^{\text{ème}}$ vecteur axial factoriel (pour le nuage des individus) est noté u^α,
- la $\alpha^{\text{ème}}$ valeur propre est notée λ_α,
- la $\alpha^{\text{ème}}$ composante principale du nuage des individus est notée ψ_α,

avec $\alpha \in \{1, \ldots, r\}$. On considère aussi un tableau de données Y déduit de X par la formule :

$$Y = A + BX,$$

où A est matrice à p lignes et n colonnes, et B une matrice carrée d'ordre p. Par la suite, les résultats de l'ACP de $Y = A + BX$, l'espace $\mathbb{R}^p$ étant muni d'une métrique notée M_1, sont notés comme suit : v^α désigne le $\alpha^{\text{ème}}$ vecteur axial factoriel (pour le nuage des individus), μ_α la $\alpha^{\text{ème}}$ valeur propre et ξ_α la $\alpha^{\text{ème}}$ composante principale du nuage des individus, avec $\alpha \in \{1, \ldots, r_1\}$ où r_1 est le rang de la matrice variance V_Y de Y. De plus, on adopte les notations suivantes : $\mathbb{1}$ est le vecteur de $\mathbb{R}^n$ dont les coordonnées sont toutes égales à 1, $0_{p \times n}$ est la matrice à p lignes et n colonnes dont tous les termes sont égaux à 0, $\mathbb{I}_3$ est la matrice identité d'ordre p et D_b est matrice diagonale dont le $i^{\text{ème}}$ terme sur la diagonale est la $i^{\text{ème}}$ coordonnée b_i d'un vecteur $b \in \mathbb{R}^p$. Enfin, on note $\widetilde{X}$ (resp. $\widetilde{Y}$) le tableau centré déduit de X (resp. Y). On suppose par la suite que, pour tout $\alpha \in \{1, \ldots, r\}$, les valeurs propres (non nulles) λ_α sont simples.

Dans les questions 1. *et* 2., *on suppose que* $M = M_1 = \mathbb{I}_3$.

1. On suppose que $A = 0_{p \times n}$ et $B = k\mathbb{I}_3$ où k est un réel non nul.

1.a. Montrer que pour tout $i \in I$ et tout $j \in J$, on a $\widetilde{y}_i^j = k\,\widetilde{x}_i^j$.

1.b. Montrer que pour tout $\alpha \in \{1, \ldots, r\}$, on a :

$$v^\alpha = \epsilon_\alpha\, u^\alpha,\ \ \mu_\alpha = k^2 \lambda_\alpha \ \text{ et } \ \xi_\alpha = \epsilon_\alpha\, k\, \psi_\alpha \qquad (\text{avec } \epsilon_\alpha = \pm 1).$$

2. On suppose que $A = \left(a_1 \mathbb{1}, \ldots, a_p \mathbb{1}\right)'$ et $B = \mathbb{I}_3$, où les a_i ($i \in I$) sont des réels.

2.a. Déterminer l'expression de y_i^j en fonction de x_i^j, pour tout $i \in I$ et tout $j \in J$.

2.b. Comparer les tableaux centrés $\widetilde{X}$ et $\widetilde{Y}$.

2.c. Pour chaque $\alpha \in \{1, \ldots, r\}$, exprimer v^α en fonction de u^α, et μ_α en fonction de λ_α.

3. On suppose que $A = 0_{p\times n}$ et $B = D_b$ où $b \in \mathbb{R}^p$ est tel que $b_i \neq 0$ pour tout i.
Dans cette question, $M = D_{1/s^2}$, où D_{1/s^2} désigne la matrice diagonale des inverses des variances des p variables du tableau X, et $M_1 = \mathbb{I}_3$.

3.a. Donner l'expression de y^j en fonction de x^j, pour tout $j \in J$.

3.b. Montrer que, pour tout $j, j' \in J$, la quantité $\|y^j - y^{j'}\|^2_{M_1}$ peut s'exprimer sous la forme $\|x^j - x^{j'}\|^2_N$ où N est une métrique dont la valeur ne dépend que de b. En déduire une condition portant sur la valeur absolue de b_i ($i \in I$) qui, si elle est vérifiée, entraîne que $\xi_\alpha = \pm\, \psi_\alpha$, pour chaque $\alpha \in \{1, \ldots, r_1\}$. Par la suite, on notera (C) cette condition.

3.c. On suppose que (C) est vérifiée. Pour α quelconque, exprimer v^α en fonction de u^α.

4. On suppose $A = \big(a_1\mathbb{1}, \ldots, a_p\mathbb{1}\big)'$ et $B = D_b$ où $b \in \mathbb{R}^p$ est tel que $b_i > 0$ pour tout i. Par ailleurs, la métrique M (resp. M_1) est égale à la matrice diagonale des inverses des variances des p variables du tableau X (resp. Y). Exprimer ξ_α en fonction de ψ_α, pour chaque valeur possible de α.

Exercice 2

On considère le tableau de données, noté X, qui est défini par :

$$X = \begin{array}{c|ccccc} & j_1 & j_2 & j_3 & j_4 & j_5 \\ \hline i_1 & 1 & 2 & 3 & 4 & 5 \\ i_2 & 6 & 7 & 8 & 9 & 10 \\ i_3 & 11 & 12 & 13 & 14 & 15 \\ i_4 & 16 & 17 & 18 & 19 & 20 \\ i_5 & 21 & 22 & 23 & 24 & 25 \end{array}$$

où la $i^{\text{ème}}$ ligne désigne la variable x_i et la $j^{\text{ème}}$ colonne désigne l'individu x^j.
Par la suite, on considère les résultats de l'ACP sur matrice variance du tableau X (i.e., ACP de X avec la métrique identité et chaque individu ayant un poids égal à $1/5$).

1. Soit g le centre de gravité du nuage des cinq individus associés au tableau X. Montrer que $g = x^{j_3}$, c'est-à-dire coïncide avec l'individu j_3.

2. Calculer le tableau centré, noté Y, qui est associé à X.

3. Combien existe-t'il d'axes factoriels non triviaux ? (justifier votre réponse)

4. Pour chacun des axes factoriels non triviaux, calculer l'unique vecteur axial factoriel qui le dirige et dont la première coordonnée est positive.

5. Déterminer l'inertie totale du nuage étudié.

6. Calculer la première composante principale de l'individu j_5, notée $\Psi_1^{j_5}$.

7. Calculer la contribution de l'individu j_5 à l'inertie du premier axe, notée $CTR_1(j_5)$.

8. Déterminer la représentation des variables dans le nouveau système d'axes factoriels de l'espace des variables.

9. Dans cette question, on considère les résultats de l'ACP normée du tableau X (i.e., ACP de X lorsque $\mathbb{R}^5$ est muni de la métrique $M = D_{1/s^2}$, où D_{1/s^2} désigne la matrice diagonale des inverses des variances des variables du tableau X, et lorsque chaque individu possède un poids égal à $1/5$). Répondre à nouveau aux questions 3), 4), 5) et 6).

(Calculatrice autorisée, tous documents autorisés)

Barême approximatif : partie 1 (10 pts) ; partie 2 (10 pts).

PREMIÈRE PARTIE

Soit un tableau $\big(k(i,j)\big)_{\substack{i\in I\\ j\in J}}$ de nombres positifs, avec $I = \{1,\ldots,p\}$ et $J = \{1,\ldots,n\}$. On suppose que, pour tout $i \in I$, tous les nombres $k(i,.) = \sum_{j\in J} k(i,j)$ sont égaux entre eux et que, de plus, pour tout $j \in J$, tous les nombres $k(.,j) = \sum_{i\in I} k(i,j)$ sont égaux entre eux également. On note $a = k(1,.)$ et $b = k(.,1)$.

1) On note $k = \sum_{i\in I}\sum_{j\in J} k(i,j)$. Quelles relations existe-t'il entre les nombres a, b et k ?

2) On effectue l'analyse des correspondances du tableau k_{IJ}.

2-a) Quel est le poids attribué à chaque colonne j de J ?

2-b) Pourquoi la métrique du chi-deux utilisée pour calculer la distance entre les profils colonnes de deux colonnes j et j' est-elle la métrique usuelle à une constante de proportionnalité près ?

2-c) Quel lien existe-t'il entre le tableau des profils colonnes de k_{IJ} et le tableau k_{IJ} ?

2-d) Calculer le carré de la distance du chi-deux (que l'on notera $d^2_{\chi^2}(j,j')$) entre les profils de deux colonnes j et j' du tableau k_{IJ}, et comparer $d^2_{\chi^2}(j,j')$ au carré de la distance usuelle $d^2(j,j')$ entre les colonnes j et j' du tableau k_{IJ} :

$$d^2(j,j') = \sum_{ß\in I}(k(i,j) - k(i,j'))^2.$$

3) On effectue l'ACP sur matrice variance de k_{IJ}, I étant l'ensemble des variables et J celui des individus. Déduire de 2) que l'on peut obtenir les résultats de cette analyse à partir de ceux de l'analyse des correspondances du tableau k_{IJ}. On précisera l'équivalence entre ces deux analyses en indiquant avec précision les relations entre les valeurs propres, les inerties totales, les composantes principales sur J et les coordonnées qui donnent les représentations de l'ensemble I dans les deux analyses.

4) On effectue l'ACP sur matrice de corrélation de k_{IJ}, I étant l'ensemble des variables et J celui des individus. Cette analyse est elle équivalente aux analyses définies en 2) et 3) ?

Au cours d'une enquête effectuée auprès de lycéens et relative à la perception du travail en milieu scolaire, on a posé la question suivante à N lycéens :

> Classez de 1 à 3, les $p = 3$ propositions ci-dessous, en donnant le numéro 1 à celle qui, pour vous, 5 ans après avoir terminé vos études serait la condition la meilleure, et le numéro 3 à celle qui, pour vous, 5 ans après avoir terminé vos études serait la condition la moins bonne.
>
> - *SÉCURITE : avoir un travail stable, la sécurité de l'emploi,*
> - *TEMPS LIBRE : avoir assez de temps libre,*
> - *INTÉRÊT : avoir un travail intéressant.*

On désigne par k_{IJ} le tableau de contingence qui est défini par la condition (1) suivante :

$$\forall i \in I, \forall j \in J,\ k(i,j) = \text{ nombre de lycéens ayant classé la proposition } i \text{ au rang } j. \quad (2)$$

1) Que peut-on dire de la marge sur I du tableau k_{IJ} (*i. e.*, des nombres $k(i,.)$) ?

2) Que peut-on dire de la marge sur J du tableau k_{IJ} (*i. e.*, des nombres $k(.,j)$) ?

3) Par la suite, on suppose que l'on a interrogé $N = 8$ lycéens (ou sujets) $S_1, \ldots, S_8$, et l'on désignera toujours par I l'ensemble des trois propositions, par J l'ensemble des rangs $1, 2, 3$. Le tableau brut des classements est donné dans le tableau 1 ci-dessous.

SUJETS

rangs	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
1	1	3	1	2	3	2	3	2
2	2	1	3	1	1	1	1	1
3	3	2	2	3	2	3	2	3

TABLE 3 – Tableau des classements : par exemple S_2 classe la proposition 3 en premier, la proposition 1 en second et la proposition 2 à la dernière place.

3-a) Construire le tableau k_{IJ}.

3-b) Appliquer le principe d'équivalence distributionnelle pour simplifier le tableau k_{IJ}. En déduire le tableau des profils-lignes ne comportant que deux profils.

3-c) Pour chaque profil-ligne défini en 3-b), calculer la distance du khi-deux entre le profil-ligne considéré et le centre de gravité du nuage des profils-lignes.

3-d) Effectuer l'analyse des correspondances du tableau k_{IJ}. On donnera le nombre d'axes factoriels non triviaux, les composantes principales sur I et J, la ou les valeurs propres non triviales et la représentation simultanée associée.

3-e) Donner les résultats de l'ACP sur matrice variance du tableau k_{IJ}, I étant l'ensemble des variables et J celui des individus (on pourra appliquer certains résultats de la première partie).

EXAMEN D'ANALYSE DES DONNÉES **M1 MMD** – 31 mai 2012

Calculatrice autorisée, documents autorisés : 3 feuilles recto-verso.

Barême : Exercice 1 (13 pts) ; Exercice 2 (7 pts).

N. B. *Dans l'exercice 2, les questions 2 et 3 sont indépendantes des questions 1-b et 1-c.*

EXERCICE 1

Soit le tableau k_{IJ} suivant, où $I = \{i_1, i_2, i_3\}$ et $J = \{A, B, C, D, E\}$.

$I \setminus J$	A	B	C	D	E
i_1	2	2	0	0	0
i_2	0	2	2	0	0
i_3	0	0	0	2	2

Dans ce qui suit, on effectue l'Analyse Factorielle des Correspondances (AFC) de k_{IJ}. On notera ψ_α^i (resp. ψ_α^j) l'abscisse de la projection du profil de la ligne i (resp. colonne j) sur le α ème axe factoriel issu de l'AFC de k_{IJ} et associé à la valeur propre λ_α.

1) Calculer les lois marginales f_I et f_J.

2) Déterminer la matrice F_1 (resp. F_2) des profils-colonnes (resp. profils-lignes) de k_{IJ}.

3) Donner le centre de gravité, noté g, du nuage $\mathcal{N}(I)$ des profils-lignes de k_{IJ}.

4) On note $\rho(i)$ la distance du khi-deux entre g et le profil de la ligne i.
Calculer $\rho^2(i_3)$. Par la suite on admettra que $\rho^2(i_1) = \rho^2(i_2) = \dfrac{5}{4}$.

5) Calculer l'inertie totale du nuage $\mathcal{N}(I)$.

6) Calculer les vecteurs propres de la matrice $F_2'F_1' = (F_1F_2)'$. En déduire le nombre r d'axes factoriels non triviaux ainsi que ψ_α^i pour $i \in I$, $\alpha \leq r$ (en supposant $\psi_\alpha^{i_1} > 0$).

7) Pour les questions 7-a, 7-b, 7-c et 7-d, les solutions seront données sans effectuer le calcul des valeurs propres et des vecteurs propres de la matrice $F_2'F_1' = (F_1F_2)'$.

7-a) En supposant le facteur φ_1^I de la forme $(a, a, b)'$ avec $a > 0$, calculer a et b.

7-b) Déduire du résultat précédent les valeurs de ψ_1^j pour $j \in J$.

7-c) Déduire de ce qui précède la valeur de λ_1, puis déterminer ψ_1^i pour $i \in I$.

7-d) Déterminer la valeur de λ_2, puis calculer ψ_2^i pour $i \in I$ (on supposera $\psi_2^{i_1} > 0$).

8) Montrer que $\mathrm{COR}_1(i_1) = \dfrac{2}{5}$, puis calculer $\mathrm{COR}_2(i_1)$, $\mathrm{COR}_1(i_2)$ et $\mathrm{COR}_2(i_2)$.

EXERCICE 2

Soit I un ensemble de n individus ayant répondu à un questionnaire comportant un ensemble Q de questions. On désigne par J_q l'ensemble des modalités de réponse à la question $q \in Q$ et par $J = \bigcup_{q \in Q} J_q$ l'ensemble des modalités de réponse à toutes les questions. Chaque individu i choisit une et une seule modalité de réponse à chaque question $q \in Q$, et pour chaque modalité $j \in J$, on pose :

$$k(i,j) = \begin{cases} 1 & \text{si } i \text{ a choisi la modalité } j, \\ 0 & \text{sinon.} \end{cases}$$

On note $k(i) = \sum_{j \in J} k(i,j)$ et $k(j) = \sum_{i \in I} k(i,j)$. De plus, $p(j)$ désigne la proportion d'individus ayant choisi la modalité j.

1) On se place dans le cadre de l'AFC du tableau k_{IJ}. On note $\mathcal{N}(J)$ le nuage des profils-colonnes, et g le centre de gravité de $\mathcal{N}(J)$.

1-a) Montrer que $g = \frac{1}{n} \mathbb{1}_n$.

1-b) Soit $\rho(j)$ la distance du Khi-deux entre g et le profil de la j ème colonne de k_{IJ}. Montrer que $\rho^2(j) = \dfrac{1 - p(j)}{p(j)}$.

1-c) En déduire que l'inertie totale I_T de $\mathcal{N}(J)$ est égale à $\sum_{q \in Q} \dfrac{(\text{Card} J_q) - 1}{\text{Card} Q}$.

2) On considère le tableau k_{IJ_q} qui est le sous-tableau de k_{IJ} obtenu en ne considérant que les colonnes de l'ensemble J_q des modalités de la question q. On note $\mathcal{N}(J_q)$ le sous-nuage de $\mathcal{N}(J)$ défini comme l'ensemble des profils-colonnes de ce sous-tableau k_{IJ_q}. De plus, on note $\widetilde{f_j}$ le poids de la colonne j (avec $j \in J_q$) dans le sous-tableau k_{IJ_q}. De façon générale, si x est une caractéristique (par ex. marge, centre de gravité, etc) définie pour le tableau k_{IJ}, elle sera notée $\widetilde{x}$ lorsqu'elle est définie pour le tableau k_{IJ_q}.

2-a) Exprimer $\widetilde{f_j}$ en fonction de $p(j)$.

2-b) On note $\widetilde{g}$ le centre de gravité de $\mathcal{N}(J_q)$. Montrer que $\widetilde{g} = g$.

3) On suppose que l'on a $\text{Card} J \leq n$, et on désigne par r le nombre d'axes factoriels non triviaux de l'AFC de k_{IJ}. Quel majorant de r peut-on déduire du résultat obtenu à la question 2-b) ?

EXAMEN D'ANALYSE DES DONNÉES M1 MMD, 3 juin 2014

Calculatrice autorisée, documents autorisés : 3 feuilles A4 recto-verso.

Barème : problème 1 (11 pts), problème 2 (9 pts). Les deux problèmes sont indépendants, mais il est suggéré de débuter par le problème 1 qui introduit le cadre général traité dans le problème 2.

Problème 1.

On considère le tableau de données bloc-diagonal, noté X, et défini par :

$$X = \begin{array}{c|cccccc} & j_1 & j_2 & j_3 & j_4 & j_5 & j_6 \\ \hline i_1 & 2 & 0 & 3 & 0 & 0 & 0 \\ i_2 & 0 & 4 & 1 & 0 & 0 & 0 \\ i_3 & 0 & 0 & 0 & 2 & 0 & 3 \\ i_4 & 0 & 0 & 0 & 0 & 4 & 1 \end{array}$$

On note $I = \{i_1, \ldots, i_4\}$ et $J = \{j_1, \ldots, j_6\}$. Pour tout $i \in I$ et $j \in J$, on note f_J^i et f_I^j les profils respectifs de la $i^{\text{ème}}$ ligne et de la $j^{\text{ème}}$ colonne de X. Par la suite, on considère l'AFC du tableau X.

1) Quelle métrique M est utilisée pour analyser le nuage $\mathcal{N}(J)$ des profils-colonnes ?

2) Soit H_1 (resp. H_2) la matrice des profils-colonnes (resp. profils-lignes) du tableau X. Calculer la matrice $H_1 H_2$.

3) Montrer que la droite Δv engendrée par le vecteur $v = f_I^{j_2} - f_I^{j_1}$ est un axe factoriel de l'AFC de X associé à une valeur propre λ dont on précisera la valeur.

4) Déterminer un vecteur $w \in \mathbb{R}^4$ qui est M-orthogonal à v et tel que la droite Δw engendrée par w soit un axe factoriel de l'AFC de X associé à la valeur propre λ

5) Montrer que l'inertie totale du nuage $\mathcal{N}(J)$ est égale à $\dfrac{12}{5}$.

6) Montrer que 1 et λ sont les seules valeurs propres non triviales (i.e. non nulles) de l'AFC de X.

7) On choisit Δv comme deuxième axe factoriel de l'AFC de X. Déterminer les coordonnées du vecteur axial factoriel, noté u_I, qui dirige cet axe : on supposera que la deuxième coordonnée de u_I est positive afin de fixer l'orientation de l'axe.

8) Soit ψ_2^J la $2^{\text{ème}}$ composante principale de $\mathcal{N}(J)$. Montrer que pour tout $j \in J$, on a :

$$\psi_2^j = \sqrt{2}(f_{i_2}^j - f_{i_1}^j).$$

9) Soit ψ_2^I la $2^{\text{ème}}$ composante principale de $\mathcal{N}(I)$. Montrer que pour tout $i \in I$, on a :

$$\psi_2^i = 2\sqrt{\frac{14}{5}}\, u_i.$$

10) Calculer la contribution $\mathrm{COR}_2(j)$, pour $j \in \{j_1, j_2\}$.

11) Donner la représentation simultannée de $\mathcal{N}(I)$ et $\mathcal{N}(J)$ sur le 2 ème axe factoriel.

Problème 2.
On considère un tableau X bloc-diagonal de la forme : $X = \begin{pmatrix} K & 0 \\ 0 & K \end{pmatrix}$, où K désigne une matrice de nombres positifs, à p lignes et q colonnes (avec $p, q \geq 1$), et où 0 est la matrice, à p lignes et q colonnes, dont tous les termes sont nuls. On note I' et J' les ensembles des indices relatifs, respectivement, aux $2p$ lignes et $2q$ colonnes de X. On utilisera les notations conventionnelles en AFC pour le tableau K : k (somme du tableau K), I et J (ensembles des indices désignant, respectivement, les lignes et les colonnes de K), F_1 et F_2 (matrices des profils colonnes et des profils lignes de K, respectivement), et f_I et f_J (profils respectifs de la colonne de marge et de la ligne de marge de K). Pour tout $u, v \in \mathbb{R}^l$ (avec $l \geq 1$), on notera $\begin{pmatrix} u \\ v \end{pmatrix}$ le vecteur de $\mathbb{R}^{2l}$ dont les l premières coordonnées sont celles de u et les l suivantes celles de v. On notera 0_p le vecteur nul de $\mathbb{R}^p$. Par la suite, on se place dans le cadre de l'AFC du tableau X. Par la suite, on notera H_1 et H_2 les matrices respectives des profils-colonnes et des profils-lignes du tableau X.

1) Montrer que $\begin{pmatrix} f_I \\ f_I \end{pmatrix}$ est un vecteur directeur d'un axe factoriel trivial de l'AFC de X.

2) Montrer que la matrice H_1H_2 s'exprime sous la forme : $H_1H_2 = \begin{pmatrix} F_1F_2 & 0 \\ 0 & F_1F_2 \end{pmatrix}$.

3) Dans cette question, u_I désigne un vecteur axial factoriel (non trivial) de l'AFC de K pour le nuage $\mathcal{N}(J)$, qui est associé à une valeur propre notée λ. Déterminer $z \in \mathbb{R}^p$ tel que les vecteurs $\begin{pmatrix} u_I \\ z \end{pmatrix}$ et $\begin{pmatrix} z \\ u_I \end{pmatrix}$ dirigent des axes factoriels de l'AFC de X qui soient distincts et chacun associés à la valeur propre λ.

4) Montrer que $v_{I'} = \begin{pmatrix} f_I \\ 0_p \end{pmatrix}$ est vecteur propre de H_1H_2 associé à la valeur propre 1.

5) Quelle condition, non vérifiée par $v_{I'}$ (défini en 4)), empêche le vecteur $v_{I'}$ d'être un vecteur directeur d'un axe factoriel de l'AFC de X ?

6) Déterminer un vecteur directeur d'un axe factoriel de $\mathcal{N}(J')$ associé à la valeur propre égale à 1 : on exprimera ce vecteur directeur en fonction de f_I.

7) Dans cette question, on considère plus généralement un tableau bloc-diagonal, noté T, dont les r ($r \geq 2$) blocs sur la diagonale sont tous égaux à la matrice K définie au début de ce problème. On suppose de plus que les valeurs propres de l'AFC de K sont distinctes de 1. Montrer que 1 est valeur propre l'AFC de T avec un ordre de multiplicité égal à $r-1$. Afin de simplifier la formulation de la preuve sans entraîner une perte réelle de généralité, on se limitera à donner le principe détaillé de la démonstration de ce résultat dans le cas particulier où $p \leq q$ et où les valeurs propres de l'AFC de K sont simples.

Calculatrice autorisée, documents autorisés : 3 feuilles recto-verso.

Exercice 1 *(10 pts)*
On considère le tableau de contingence k_{IJ} défini par :

$$k_{IJ} = \begin{array}{c|cccc} I \backslash J & j_1 & j_2 & j_3 & j_4 \\ \hline i_1 & 1 & 2 & 3 & 4 \\ i_2 & 5 & 6 & 7 & 8 \\ i_3 & 9 & 10 & 11 & 12 \\ i_4 & 13 & 14 & 15 & 16 \end{array}$$

où $I = \{i_1, i_2, i_3, i_4\}$ et $J = \{j_1, j_2, j_3, j_4\}$ désignent respectivement les ensembles de modalités de deux variables qualitatives. Par la suite, on utilise les notations usuelles en AFC, et on considère les résultats de l'AFC de ce tableau K.

1.[†] Calculer la matrice F_2 des profils-lignes de K.

2. Montrer que la métrique du Khi-deux, notée M, dont est muni le nuage $\mathcal{N}(I)$ des profils-lignes de K, est définie par $M = \mathrm{Diag}\big(u_m\big)_{m=1,\ldots,4}$ avec $u_m = 34/(6+m)$.

3. Pour tout $i \in I$, on note k_{iJ} le vecteur de $\mathbb{R}^4$ dont les coordonnées sont les nombres k_{ij} pour j variant dans J. On note aussi $\mathbb{1}_4$ le vecteur de $\mathbb{R}^4$ dont chaque coordonnée vaut 1. Déterminer les deux entiers α et β tels que l'on ait :

$$k_{i_m J} = \alpha(m+\beta)\mathbb{1}_4 + k_{i_1 J},$$

pour tout $m \in \{1,2,3,4\}$. En déduire l'expression de $k_{i_m\cdot}$ en fonction de m.

4. En déduire que chaque profil-ligne de K est un barycentre des deux vecteurs $\frac{1}{4}\mathbb{1}_4$ et $f_J^{i_1}$, et exprimer les deux coefficients de ce barycentre seulement en fonction de m.

5. Montrer que l'AFC de K admet un seul axe factoriel non trivial.

6. Montrer que l'axe factoriel non trivial de $\mathcal{N}(I)$ est dirigé par $w = \big(-3, -1, 1, 3\big)'$.

7. Soient ψ^I la composante principale et φ^J le facteur relatifs à l'axe factoriel non trivial du nuage $\mathcal{N}(I)$. Montrer que pour tout $m \in \{1,2,3,4\}$, on a :

$$\psi^{i_m} = \frac{5\,\psi^{i_1} + 2\,(m-1)\,\delta}{5 + 8\,(m-1)} \quad \text{avec} \quad \delta = \sum_{j \in J} \varphi^j.$$

8.[‡] Sachant que l'inertie totale de l'AFC de K est égale à 0.00936, calculer la coordonnée de la modalité $j_1 \in J$ sur l'axe factoriel non trivial.

---

†. Ce calcul se limitera à l'évaluation de chaque terme de F_2 par une fraction de deux entiers.
‡. La valeur de ψ^{j_1} sera calculée avec une précision d'au moins 3 décimales.

Exercice 2 (10 pts)
On considère un tableau de contingence arbitraire noté $K = k_{IJ}$. On adopte les notations usuelles en AFC, en particulier les notations $k = k_{\cdot\cdot}$ (somme des termes de K), $k_{i\cdot}$, $k_{\cdot j}$, f_I, f_J, f_J^i, f_I^j (avec $i \in I$ et $j \in J$), $F_1 = f_I^J$ et $F_2 = f_J^I$, où I et J désignent les ensembles de modalités de deux variables qualitatives avec $p = \text{card } I$ et $q = \text{card } J$.
Soit alors le tableau $X = x_{IJ}$ déduit de K par la transformation suivante :

$$x_{ij} = \beta_1 k_{ij} + \beta_2 \frac{k_{i\cdot} k_{\cdot j}}{k},$$

pour tout $(i, j) \in I \times J$, et où β_1 et β_2 sont deux réels positifs arbitraires de somme β non nulle. Pour tout $(i, j) \in I \times J$, on note $h_{ij} = \dfrac{x_{ij}}{x_{\cdot\cdot}}$ où $x_{\cdot\cdot} = x$ désigne la somme de tous les termes de X. Par ailleurs, on note $H_1 = H_I^J$ (resp $H_2 = H_J^I$) la matrice des profils-colonnes (resp. profils-lignes) du tableau X. Par la suite, on pose $r_1 = \dfrac{\beta_1}{\beta}$, $r_2 = \dfrac{\beta_2}{\beta}$, $r = r_1 + r_2$, et on note $\mathbb{1}_p$ (resp. $\mathbb{1}_q$) le vecteur de $\mathbb{R}^p$ (resp. $\mathbb{R}^q$) dont toutes les coordonnées sont égales à 1. On suppose que les valeurs propres de l'AFC de K sont simples et de somme non nulle.

1. *Calculs préliminaires.*

1.a. Calculer le centre de gravité des profils colonnes du tableau X en fonction de f_I.

1.b. Calculer $f_I' D_{1/f_I}$, où D_{1/f_I} désigne la matrice $\text{Diag}(1/f_i)_{i \in I}$.

1.c. Montrer que les deux matrices $f_I \mathbb{1}_q' F_2$ et $F_1 f_J \mathbb{1}_p'$ sont égales et exprimer leur valeur commune en fonction de f_I et $\mathbb{1}_p'$.

1.d. Pour $j \in J$, on note h_I^j le profil de la $j^{\text{ème}}$ colonne du tableau X. Calculer h_I^j en fonction de r_1, r_2, f_I^j et f_I, puis calculer la matrice H_1 en fonction de r_1, r_2, f_I, F_1 et $\mathbb{1}_q$.

1.e. Calculer la matrice H_2 en fonction de r_1, r_2, f_J, F_2 et $\mathbb{1}_p$.

2. Soit u_α (resp. v_α) le $\alpha^{\text{ème}}$ vecteur axial factoriel non trivial du nuage des profils-colonnes issu de l'AFC de K (resp. X). Soit λ_α (resp. μ_α) la $\alpha^{\text{ème}}$ valeur propre de l'AFC de K (resp. X). De plus, on note ψ_α^J (resp. ξ_α^J) la $\alpha^{\text{ème}}$ composante principale du nuage des profils-colonnes issue de l'AFC de K (resp. X).

2.a. En résolvant l'équation aux vecteurs propres dont est solution v_α, donner l'expression de v_α en fonction de u_α, ainsi que l'expression de μ_α en fonction de λ_α.

2.b. En déduire l'expression de ξ_α^J en fonction de ψ_α^J.

3. On note h_I le profil de la marge en colonne du tableau X, autrement dit, on désigne par h_I le vecteur de $\mathbb{R}^p$ dont les coordonnées sont les nombres $h_i = \dfrac{x_{i\cdot}}{x}$ pour $i \in I$.

3.a. Calculer l'expression de $h_I^j - h_I$ en fonction de $f_I^j - f_I$.

3.b. Par un argument géométrique simple, en déduire les résultats des questions 2.a. et 2.b.

Exercice 1

1) X est centré $\Longleftrightarrow \overline{x}_i = 0$, pour tout $i \in I \Longleftrightarrow \sum_j x_i^j = 0$, pour tout $i \in I \Longleftrightarrow X\mathbb{1}_n = 0$. Donc X est centré ssi $X\mathbb{1}_n = 0$, et par conséquent X' est centré ssi $X'\mathbb{1}_p = 0$.

2) Chaque axe factoriel de l'ACP de X est dirigé par un vecteur u tel que $V_X\, u = \lambda\, u$, ce qui équivaut à $X\,X'u = n\,\lambda\,u$. Or $X'\,\mathbb{1}_p = 0$ car X' est centré. Donc V_X admet $\mathbb{1}_p$ comme vecteur propre relatif à la valeur propre 0. Donc il existe un axe factoriel dirigé par $\mathbb{1}_p$.

3) On a $V_{X'}\,w^\alpha = \dfrac{1}{p\sqrt{n\,\lambda_\alpha}}\,X'X\,X'u^\alpha = \dfrac{n}{p\sqrt{n\,\lambda_\alpha}}\,X'V_X\,u^\alpha = \dfrac{\sqrt{n}\,\lambda_\alpha}{p\sqrt{\lambda_\alpha}}\,X'u^\alpha = \dfrac{\sqrt{n\,\lambda_\alpha}}{p}\,X'u^\alpha$.
D'où, $V_{X'}\,w^\alpha = \dfrac{n}{p}\,\lambda_\alpha\,w^\alpha$. Vérifions, de plus, que w^α est bien normé à 1 :
$\|w^\alpha\|^2_{I_n} = \dfrac{1}{n\,\lambda_\alpha}\|X'u^\alpha\|^2_{I_n} = \dfrac{1}{n\,\lambda_\alpha}(u^\alpha)'X\,X'u^\alpha = \dfrac{1}{\lambda_\alpha}(u^\alpha)'V_X\,u^\alpha = (u^\alpha)'u^\alpha = 1.$
Par conséquent w^α est un vecteur axial factoriel de l'ACP de X' relatif à la valeur propre $\mu^\alpha = n\,\lambda_\alpha/p$.

4) D'après 4) et par symétrie, si w^α est le α ème vecteur axial factoriel (non trivial) de l'ACP de X' relatif à la α ème valeur propre μ_α, alors $\dfrac{1}{\sqrt{p\,\mu_\alpha}}\,X\,w^\alpha$ est vecteur axial factoriel de l'ACP de X', et il est relatif à la valeur propre $p\,\mu_\alpha/n$. Donc les axes factoriels de ces deux ACP se correspondent de façon bijective, et leurs nombres sont identiques.
On peut aussi raisonner de la manière suivante : les vecteurs axiaux factoriels non triviaux sont, respectivement, les vecteurs propres normés relatifs aux valeurs propres non nulles des matrices $X\,X'$ et $X'X$. Or ces matrices ont même rang, donc le nombre d'axes factoriels des deux ACPs sont identiques.

5) La définition de η_i^α est ici la suivante : $\eta_i^\alpha = \dfrac{1}{n}\,x_i'\,\psi_\alpha$ et donc $\eta^\alpha = \dfrac{1}{n}\,X\,X'u^\alpha = \lambda_\alpha\,u^\alpha$.
On obtient alors :

$$\zeta_\alpha = X\,w^\alpha = \frac{1}{\sqrt{n\,\lambda_\alpha}}\,X\,X'\,u^\alpha = \frac{n}{\sqrt{n\,\lambda_\alpha}}\,V_X\,u^\alpha = \sqrt{n}\,\sqrt{\lambda_\alpha}\,u^\alpha = \sqrt{\frac{n}{\lambda_\alpha}}\,\eta^\alpha.$$

Exercice 2

1) On a $y_i^j = x_i^j - \overline{x}_i$ avec $\overline{x}_1 = 2, \overline{x}_2 = 3$ et $\overline{x}_3 = 4$. D'où :

$$Y = \begin{pmatrix} 1 & 2 & -1 & 0 & -1 & -2 & 1 & 0 \\ 0 & 1 & -2 & -1 & 2 & 1 & 0 & -1 \\ -1 & -3 & 3 & 1 & -1 & 1 & -1 & 1 \end{pmatrix}.$$

2)

$$V = \frac{1}{8}\,Y\,Y' = \frac{1}{8}\begin{pmatrix} 12 & 0 & -12 \\ 0 & 12 & -12 \\ -12 & -12 & 24 \end{pmatrix} = \frac{3}{2}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

3) On remarque que Y' est centré car la somme de chaque colonne est nulle. Donc $\mathbb{1}_3$ est vecteur directeur d'un axe factoriel trivial d'après l'exercice 1 - 2). On peut aussi remarquer que $y_1 + y_2 + y_3 = 0$ car la somme de chaque colonne est nulle. Donc, pour tout $j \in J$, $y_1^j + y_2^j + y_3^j = 0$, ce qui montre que $Y'\,\mathbb{1}_3$ est nul, et donc que $\mathbb{1}_3$ est vecteur propre de $V_Y = V_X$ relatif à la valeur propre 0. D'où le résultat.

4) L'inertie totale vaut $\text{trace}(V) = \dfrac{3}{2} + \dfrac{3}{2} + 3$. Donc $\lambda_1 = \dfrac{9}{2}$, $\lambda_2 = \dfrac{3}{2}$ et $\lambda_3 = 0$.

5) Soit τ_i le taux d'inertie expliqué par le i° axe factoriel : $\tau_1 = \dfrac{9/2}{6} = \dfrac{3}{4}$, $\tau_2 = \dfrac{3/2}{6} = \dfrac{1}{4}$.

6) Si $u = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ est un vecteur axial factoriel non trivial, alors u est orthogonal à $\mathbb{1}_3$, *i. e.* $u_1 + u_2 + u_3 = 0$. On remarque que :

$$\frac{3}{2}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \frac{3}{2}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}.$$

Donc le vecteur $u^2 = \dfrac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$ est vecteur propre de V associé à $\lambda_2 = \dfrac{3}{2}$. Par conséquent, u^2, qui est de norme 1, est un vecteur axial factoriel qui dirige le deuxième axe factoriel. Soit $\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$ le premier axe factoriel. On a :

$$w_1 + w_2 + w_3 = 0 \quad \text{et} \quad w_1 - w_2 = 0.$$

Donc $w_3 = -2w_1$, et par conséquent, $u^1 = \dfrac{1}{\sqrt{6}}\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$, qui est de norme 1, est un vecteur axial factoriel qui dirige le premier axe factoriel.

7) On a :

$$\psi_1(2) = {y^2}'\,u^1 = \frac{1}{\sqrt{6}}\begin{pmatrix} 2 & 1 & -3 \end{pmatrix}\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} = -\frac{9}{\sqrt{6}} \approx -3.674.$$

Comme ψ_1 est centrée, on en déduit que $\psi_1(3) \approx 3.674$.

Par ailleurs, $\text{CTR}_1(j) = p_j \dfrac{\psi_1^2(j)}{\lambda_j}$. D'où, $\text{CTR}_1(2) = \text{CTR}_1(3) \approx \dfrac{1}{8}\dfrac{3.674^2}{4.5} \approx 37.5\%$.

On a aussi $\text{QLT}_1(j) = \dfrac{\psi_1^2(j)}{\|x^j\|^2}$, donc $\text{QLT}_1(2) = \text{QLT}_1(3) \approx \dfrac{3.674^2}{2^2 + 1^2 + 3^2} \approx 0.964$, car x^2 et x^3 ont la même valeur de composante principale sur l'axe 1 et ont la même norme.

8) On a $\text{CTR}_1(i) = (u_i^1)^2$. Donc $\text{CTR}_1(1) = 1/6 \approx 16.667\%$.

9) On a $\text{QLT}_2(i) = \dfrac{(\eta_i^2)^2}{v_{ii}} = \dfrac{(\sqrt{\lambda_2}\,u_i^2)^2}{v_{ii}} = \lambda_2 \dfrac{(u_i^2)^2}{v_{ii}}$. D'où $\text{QLT}_2(2) = \dfrac{3}{2}\dfrac{1}{2}\dfrac{2}{3} = \dfrac{1}{2}$.

Exercice 1

1) Comme $x_1 + 3\,x_2 + x_3 = 0$, on a $x_1^j + 3\,x_2^j + x_3^j = 0$ pour tout $j \in J$. Il en résulte $\overline{x}_1 + 3\,\overline{x}_2 + \overline{x}_3 = 0$. Par soustraction avec la première égalité, on a $y_1 + 3\,y_2 + y_3 = 0$.

2) L'égalité $y_1 + 3\,y_2 + y_3 = 0$ équivaut à $Y'u = 0$ avec $u' = (1, 3, 1)$.

3) Soit V_X la matrice variance associée au tableau X. On a : $V_X\, u = \frac{1}{n} Y\, Y'u = 0$, ce qui montre que u dirige un axe factoriel trivial de l'ACP de X.

4) $V_Z = \mathrm{Var}(AX) = A\,\mathrm{Var}(X) A' = A\, V_X A'$, d'après une proposition du cours.

5) Il est immédiat de vérifier que $A'A = 9\,\mathbb{I}_3$ où $\mathbb{I}_3$ désigne la matrice identité de $\mathbb{R}^3$.

6) On a $V_Z A\, u^\alpha = A\, V_X A' A\, u^\alpha = 9 A\, V_X\, u^\alpha = 9\lambda_\alpha A\, u^\alpha$. Par ailleurs $\|u^\alpha\|^2 = 1$, donc $\|A\, u^\alpha\|^2 = (A\, u^\alpha)' A\, u^\alpha = (u^\alpha)' A' A\, u^\alpha = 9$. Par conséquent $\|\frac{1}{3}\, A\, u^\alpha\|^2 = 1$. Donc $\frac{1}{3}\, A\, u^\alpha$ est le α ème vecteur axial factoriel de l'ACP de Z (au signe près). Il en résulte que $w^\alpha = \frac{1}{3}\, A\, u^\alpha$ et $\mu_\alpha = 9\,\lambda_\alpha$.

7) D'après ce qui précède, $\frac{1}{3}\, A\, u$ dirige un axe factoriel trivial de l'ACP de Z. On a :

$$\frac{1}{3}\, A\, u = \frac{1}{3} \begin{pmatrix} -1 & 2 & -2 \\ -2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 3 \\ 3 \\ 9 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}.$$

8) Rappelons une propriété qui a été utilisée en TD : si Y est le tableau centré associé à X, alors le tableau centré associé à $Z = AX$ est égal à $T = AY$ (cf. TD4bis). On peut d'ailleurs montrer cette propriété de la façon suivante. Soit z_i le vecteur qui contient la i ème ligne de $Z = AX$. Donc z_i est la i ème colonne de $X'A'$, d'où $z_i = X'a_i = a_i^1\, x_1 + a_i^2\, x_2 + a_i^3\, x_3$. Par conséquent $\overline{z}_i = a_i^1\, \overline{x}_1 + a_i^2\, \overline{x}_2 + a_i^3\, \overline{x}_3$. D'où $t_i = z_i - \overline{z}_i = a_i^1\, y_1 + a_i^2\, y_2 + a_i^3\, y_3 = Y'a_i$. Donc, pour tout i, t_i est la i ème colonne de $Y'A'$, soit $T = AY$. Donc la propriété est vraie. On en déduit :

$$\zeta_\alpha = T'\, w^\alpha = \frac{1}{3}\, T' A\, u^\alpha = \frac{1}{3}\, Y' A' A\, u^\alpha = 3 Y' \mathbb{I}_3\, u^\alpha = 3\, \Psi_\alpha.$$

9) On a $\mathrm{cov}(z_i, \frac{\zeta_\alpha}{\sqrt{\mu_\alpha}}) = \sqrt{\mu_\alpha}\, w_i^\alpha$. Donc $\mathrm{cov}(z_i, \zeta_\alpha) = \mu_\alpha\, w_i^\alpha$ et par conséquent, on en déduit que $\mathrm{cov}(z_i, \Psi_\alpha) = \dfrac{1}{3} \mu_\alpha\, w_i^\alpha$. Or w^α étant un vecteur axial non trivial de l'ACP de Z, on a $w^\alpha \perp (1, 1, 3)'$ au sens de la métrique $\mathbb{I}_3$, d'où l'on déduit la relation demandée.

Exercice 2

1) On a $g = \begin{pmatrix} \overline{x}_1 = 2 \\ \overline{x}_2 = 1 \\ \overline{x}_3 = 3 \end{pmatrix}$. Or $y_i^j = x_i^j - g_i$, d'où : $Y = \begin{pmatrix} 1 & 0 & 3 & -2 & 0 & -2 \\ 1 & -1 & -1 & 1 & 0 & 0 \\ -3 & 2 & -1 & 0 & 0 & 2 \end{pmatrix}$.

2) Le calcul de V est une application directe de la formule $V = \dfrac{1}{6}\, Y\, Y'$.

3) Notons v^1, v^2 et v^3 les trois colonnes de V. On remarque que l'on a $v^1 + 2v^2 + v^3 = 0$. Donc V est au plus de rang égal à 2. Comme V n'est pas de rang 1, sinon toutes les colonnes seraient colinéaires, on en déduit que V est de rang 2. Donc le nombre d'axes factoriels non triviaux est égal à 2.

4) De $v^1 + 2v^2 + v^3 = 0$, on déduit que $Vu = 0$ avec $u = (1, 2, 1)'$, ce qui montre que u est un vecteur directeur d'un axe factoriel trivial.

5) On observe que $x^5 = g$, donc $\Psi_1^5 = 0$.

6) Comme Ψ_1 est centrée, on en déduit $2\sqrt{2} - \sqrt{2} + 2\sqrt{2} - \sqrt{2} + 0 + \Psi_1^6 = 0$. D'où $\Psi_1^6 = -2\sqrt{2}$.

7) On a $\|\Psi_1\|^2_{D_p} = \lambda_1$, donc $\lambda_1 = \dfrac{1}{6}\,(8 + 2 + 8 + 2 + 8) = \dfrac{14}{3}$.
Comme $I_T = \mathrm{tr}(VM) = \dfrac{20}{3} = \lambda_1 + \lambda_2$, on a $\lambda_2 = \dfrac{20}{3} - \dfrac{14}{3} = 2$.

8) On a $\mathrm{CTR}_1(j) = p_j \dfrac{(\Psi_1^j)^2}{\lambda_1}$. D'où $\mathrm{CTR}_1(j = 1) = \dfrac{8}{6}\,\dfrac{3}{14} = \dfrac{2}{7}$.
De même $\mathrm{CTR}_1(j = 2) = \dfrac{2}{6}\,\dfrac{3}{14} = \dfrac{1}{14}$.

9) On a $\mathrm{COR}_1(j) = \dfrac{(\Psi_1^j)^2}{\|y^j\|^2_M}$. Donc $\mathrm{COR}_1(j = 2) = \dfrac{2}{1^2 + 2^2} = \dfrac{2}{5}$ et $\mathrm{COR}_2(j = 2) = \dfrac{3}{5}$.

10) On a $\mathrm{cov}(x_1, \Psi_1) = \mathrm{cov}(y_1, \Psi_1) = \dfrac{1}{6}\, y_1' \Psi_1$. Donc :

$$\mathrm{cov}(x_1, \Psi_1) = \frac{1}{6}\,(2\sqrt{2} + 6\sqrt{2} + 2\sqrt{2} + 4\sqrt{2}) = \frac{7}{3}\sqrt{2}.$$

Par ailleurs, $\quad \mathrm{COR}_1(i = 1) = \dfrac{(\eta_i^1)^2}{\|y_1\|^2} = \dfrac{\mathrm{cov}\Big(y_1, \dfrac{\Psi_1}{\sqrt{\lambda_1}}\Big)^2}{\|y_1\|^2} = \dfrac{\mathrm{cov}(y_1, \Psi_1)^2}{\lambda_1\, \|y_1\|^2}$.

Or :

$$\mathrm{cov}(y_1, \Psi_1)^2 = \frac{49 \times 2}{9} \quad ; \quad \lambda_1 = \frac{14}{3} \quad ; \quad \|y_1\|^2_{D_p} = v_{11} = 3.$$

Par conséquent,

$$\begin{aligned} \mathrm{COR}_1(i = 1) &= \frac{49 \times 2}{9}\,\frac{3}{14}\,\frac{1}{3} = \frac{7}{9}, \\ \mathrm{COR}_2(i = 1) &= 1 - \mathrm{COR}_1(i = 1) = \frac{2}{9}. \end{aligned}$$

EXERCICE 1

1) $\gamma = \dfrac{1}{n}$

2) $$V_X = \frac{1}{n}YY' = \frac{1}{n}\big(X - g_X\mathbb{1}_n'\big)\big(X' - \mathbb{1}_n\, g_X'\big) = \frac{1}{n}\big(XX' - g_X\mathbb{1}_n'X' - X\mathbb{1}_n\, g_X' + g_X\mathbb{1}_n'\mathbb{1}_n\, g_X'\big)$$
$$= \frac{1}{n}XX' - g_Xg_X' - g_Xg_X' + g_Xg_X' = \frac{1}{n}XX' - g_Xg_X'.$$

3) (1) équivaut à $\exists c$ tel que $x_1 + \ldots + x_p = c\,\mathbb{1}_n$. Or $x_1 + \ldots + x_p = X'\mathbb{1}_p$, d'où le résultat.

4) $$V_X\,\mathbb{1}_p = \frac{1}{n}XX'\mathbb{1}_p - g_Xg_X'\mathbb{1}_p = \frac{c}{n}X\,\mathbb{1}_n - \frac{1}{n}X\,\mathbb{1}_n\frac{1}{n}\,\mathbb{1}_n'\,X'\,\mathbb{1}_p = c\,g_X - \frac{1}{n}X\,\mathbb{1}_n\frac{1}{n}\,\mathbb{1}_n'\,c\,\mathbb{1}_n$$
$$= c\,g_X - \frac{1}{n}\,X\,\mathbb{1}_n\,c = 0.$$

5) $$g_Z = \frac{1}{p}Z\,\mathbb{1}_p = \frac{1}{p}Y'\,\mathbb{1}_p = \frac{1}{p}\big(X' - \mathbb{1}_n\,g_X'\big)\mathbb{1}_p = \frac{1}{p}X'\,\mathbb{1}_p - \frac{1}{p}\,\mathbb{1}_n\left(\frac{1}{n}X\,\mathbb{1}_n\right)'\mathbb{1}_p$$
$$= \frac{c}{p}\,\mathbb{1}_n - \frac{1}{p}\,\mathbb{1}_n\,\frac{1}{n}\mathbb{1}_n'X'\,\mathbb{1}_p = \frac{c}{p}\,\mathbb{1}_n - \frac{1}{p}\,\mathbb{1}_n\,\frac{1}{n}\,\mathbb{1}_n'\,c\,\mathbb{1}_n = \frac{c}{p}\,\mathbb{1}_n - \frac{1}{p}\,\mathbb{1}_n\,c = 0.$$

6) Le nombre d'axes factoriels non triviaux de l'ACP de X (resp. Z) est égal au rang de $Y\,Y'$ (resp. $Y'\,Y$) . Le résultat vient alors du fait que les matrices $Y'\,Y$ et $Y\,Y'$ ont même rang, cette propriété résultant de ce que $\mathrm{rg}(Y'\,Y) = \mathrm{rg}(Y) = \mathrm{rg}(Y')$ [‡] (cf. page 5 du polycopié *Rappels d'Algèbre Linéaire*)

7) $V_Z\big(Y'u^\alpha\big) = \dfrac{1}{p}\,Y'YY'u^\alpha = \dfrac{n}{p}\,Y'V_Xu^\alpha = \dfrac{n}{p}\,Y'\lambda_\alpha\,u^\alpha = \dfrac{n}{p}\,\lambda_\alpha\big(Y'u^\alpha\big)$. Donc $Y'u^\alpha$ dirige un axe factoriel (non trivial) de l'ACP de Z et l'inertie du nuage $\mathcal{N}$ projeté sur cet axe vaut $\dfrac{n}{p}\,\lambda_\alpha$.

8) Soit w^α le α ème axe factoriel de l'ACP de Z. D'après la question 7, on a :
$w^\alpha = \dfrac{Y'u^\alpha}{\|Y'u^\alpha\|_{M=I_n}}$ et $\|Y'u^\alpha\|^2_{I_n} = \big(u^\alpha\big)'Y\,Y'\,u^\alpha = n\,\big(u^\alpha\big)'V_X\,u^\alpha = n\,\big(u^\alpha\big)'\lambda_\alpha\,u^\alpha = n\,\lambda_\alpha$.
Donc $G_\alpha = Y\,w_\alpha = Y\,\dfrac{Y'u^\alpha}{\sqrt{n\,\lambda_\alpha}} = \dfrac{Y\,F_\alpha}{\sqrt{n\,\lambda_\alpha}}$.

‡. Pour rappel, on peut démontrer que $\mathrm{rg}(Y'\,Y) = \mathrm{rg}(Y)$ de la façon suivante. De $u'(Y'\,Yu) = \|Y\,u\|^2$, on déduit que $Y'\,Yu = 0 \iff Y\,u = 0$. Donc $\mathrm{Ker}(Y'\,Y) = \mathrm{Ker}(Y)$, ce qui prouve que $Y'\,Y$ et Y ont même rang, d'après la relation $\dim E = \dim(\mathrm{Ker}(f)) + \mathrm{rg}(f)$.

EXERCICE 2

1) $g' = \begin{pmatrix} 4/3 & 4/3 & 8/3 \end{pmatrix}'$. D'où $Y = \dfrac{1}{3}\begin{pmatrix} -1 & -1 & 2 & -4 & 5 & -1 \\ -4 & -1 & 2 & -1 & -1 & 5 \\ -5 & -2 & 4 & -5 & 4 & 4 \end{pmatrix}$.

2) $V = \dfrac{1}{54}\begin{pmatrix} 48 & 3 & 51 \\ 3 & 48 & 51 \\ 51 & 51 & 102 \end{pmatrix} = \dfrac{1}{18}\begin{pmatrix} 16 & 1 & 17 \\ 1 & 16 & 17 \\ 17 & 17 & 34 \end{pmatrix}$.

3) La troisième ligne de Y est la somme des deux premières, donc le rang de Y est au plus 2. Par ailleurs, les lignes n'étant pas toutes proportionnelles, ce rang n'est pas égal à 1. Donc le rang de Y, qui est aussi celui de V, vaut 2. Donc le nombre d'axes factoriels non triviaux vaut 2.

4) $I_T = \mathrm{tr}(VM) = \mathrm{tr}(V) = \dfrac{16+16+34}{18} = \dfrac{66}{18} = \dfrac{11}{3} \approx 3.6667$.

5) $VM\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \dfrac{1}{18}\begin{pmatrix} 51 \\ 51 \\ 102 \end{pmatrix} = \dfrac{1}{6}\begin{pmatrix} 17 \\ 17 \\ 34 \end{pmatrix} = \dfrac{17}{6}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$.

6) Soit τ ce pourcentage. D'après ce qui précède, nous avons $\tau = \dfrac{17/6}{11/3} = \dfrac{17}{22} \approx 0.7727$. Comme $\tau > 0.5$, il s'agit du premier axe factoriel.

7) $u^1 = \dfrac{1}{\sqrt{1^2+1^2+2^2}}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \dfrac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$

8) $\psi_1^{j_2} = (y^{j_2})'\,u^1 = \dfrac{1}{3\sqrt{6}}\begin{pmatrix} -1 & -1 & -2 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = -\dfrac{\sqrt{6}}{3} = -\sqrt{\dfrac{2}{3}} \approx -0.8165$.

9) $\mathrm{CTR}_1(j_2) = \dfrac{1}{6}\dfrac{\left(\psi_1^{j_2}\right)^2}{\lambda_1} = \dfrac{1}{6} \times \dfrac{6}{17} \times \dfrac{2}{3} = \dfrac{2}{51} \approx 0.0392 = 3.92\,\%$.

10) $\mathrm{COR}_1(j_2) = \dfrac{\left(\psi_1^{j_2}\right)^2}{\|y^{j_2}\|^2} = \dfrac{2/3}{(1/3)^2+(1/3)^2+(2/3)^2} = \dfrac{2/3}{6/9} = 1$

11) $\mathrm{CTR}_1(i_1) = \dfrac{\left(\eta_i^1\right)^2}{\lambda_1} = \dfrac{\left(\sqrt{\lambda_1}\,u_1^1\right)^2}{\lambda_1} = (u_1^1)^2 = \dfrac{1}{6} \approx 0.1667 = 16.67\,\%$.

PREMIÈRE PARTIE

1) $k = \sum_{i\in I} k(i,.) = p\,a = \sum_{j\in J} k(.,j) = nb.$

2-a) $f_j = \dfrac{k(.,j)}{k} = b/nb = 1/n.$

2-b) Cette métrique est D_{1/f_I} avec $f_i = a/pa = 1/p$. Donc $D_{1/f_I} = p\mathbb{I}_{\mathrm{d}}$.

2-c) Soit $F_1 = (f_I^1, \dots, f_I^n)$ le tableau des profils colonnes.
De $f_i^j = k(i,j)/k(.,j) = k(i,j)/b$, on déduit $F_1 = \dfrac{1}{b}\,k_{IJ}$.

2-d)

$$\begin{aligned} d^2_{\chi^2}(j,j') &= \|f_I^j - f_I^{j'}\|^2_{D_{1/f_I}} = p\|f_I^j - f_I^{j'}\|^2_{\mathbb{I}_{\mathrm{d}}} = p\sum_i \Big(\frac{k(i,j)}{b} - \frac{k(i,j')}{b}\Big)^2 \\ d^2_{\chi^2}(j,j') &= \frac{p}{b^2}\sum_i \big(k(i,j) - k(i,j')\big)^2 = \frac{p}{b^2}\,d^2(j,j') \end{aligned}$$

3) On a $d^2(j,j') = \dfrac{b^2}{p} d^2_{\chi^2}(j,j')$, donc les distances en ACP sont égales à celles en AFC multipliées par $b/\sqrt{p}$. Par ailleurs, le poids de j est identique dans les deux analyses et égal à $1/n$. D'après les résultats sur l'analyse d'un tableau de distances (AFTD), et en notant avec une astérisque $\star$ les quantités relatives à l'ACP sur matrice variance de k_{IJ}, on a $W^\star = \dfrac{b^2}{p} W$. Donc les valeurs propres de $W^\star$ sont celles de W multipliées par b^2/p, les vecteurs propres étant identiques. D'où :

$$\lambda^\star_\alpha = \frac{b^2}{p}\lambda_\alpha, \quad I^\star_T = \frac{b^2}{p} I_T, \quad (\psi^J_\alpha)^\star = \frac{\epsilon\, b}{\sqrt{p}}\psi^J_\alpha, \qquad \text{avec } \epsilon \in \{-1, 1\}.$$

Soit η^α_I le vecteur des coordonnées des variables pour l'ACP sur matrice variance :

$$\eta^\alpha_I = Y\,D_{1/n}\frac{(\psi^J_\alpha)^\star}{\sqrt{\lambda^\star_\alpha}} = Y\,D_{1/n}(\varphi^J_\alpha)^\star = \big(k_{IJ} - \frac{a}{n}\mathbb{1}_p\mathbb{1}'_p\big)D_{1/n}(\varphi^J_\alpha)^\star$$

Or $k_{IJ} = a\,F_2'$, donc $\eta^\alpha_I = a\,F_2' D_{1/n}(\varphi^J_\alpha)^\star = \dfrac{a}{n}\,F_2'(\varphi^J_\alpha)^\star = \dfrac{a}{n}\,\psi^I_\alpha.$

4) Comme $M = \mathrm{Diag}(1/v_{ii})$ pour l'ACP sur matrice de corrélation, les deux analyses ne sont pas équivalentes en général, sauf si les variances v_{ii} sont constantes.

DEUXIÈME PARTIE

1) et **2)** On a $k(i,.) = N$ et $k(.,j) = N$, par construction.

3-a)

	j	rangs		
	i	1	2	3
	1	2	6	0
prop.	2	3	1	4
	3	3	1	4

3-b)

j	rangs		
i	1	2	3
1	2	6	0
2	6	2	8

<u>Profils-lignes</u>

$$F_2 = \begin{pmatrix} 1/4 & 3/8 \\ 3/4 & 1/8 \\ 0 & 1/2 \end{pmatrix}$$

3-c) Le centre de gravité de $\mathcal{N}(I)$ est égal à $f_J = (1/3, 1/3, 1/3)'$. Les poids des profils-lignes f_J^1 et f_J^2 sont respectivement $f_{1.} = 1/3$ et $f_{2.} = 2/3$. D'où les distances :

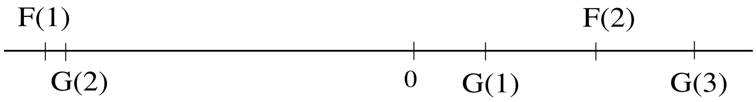
$$\begin{aligned}
\|f_J^1 - f_J\|^2_{D_{1/f_J}} &= 3\Big[\Big(\frac{1}{4}-\frac{1}{3}\Big)^2 + \Big(\frac{3}{4}-\frac{1}{3}\Big)^2 + \frac{1}{3^2}\Big] = \frac{7}{8} \\
\|f_J^2 - f_J\|^2_{D_{1/f_J}} &= 3\Big[\Big(\frac{1}{8}-\frac{1}{3}\Big)^2 + \Big(\frac{1}{8}-\frac{1}{3}\Big)^2 + \Big(\frac{1}{2}-\frac{1}{3}\Big)^2\Big] = \frac{7}{32}
\end{aligned}$$

3-d) Un seul axe factoriel (car 2 profils-lignes), avec $\lambda = \frac{1}{3} \times \frac{7}{8} + \frac{2}{3} \times \frac{7}{32} = \frac{7}{16}$. Pour cet unique axe factoriel, notons F et G les composantes principales pour I et J, respectivement. On a : $F(i) = \epsilon_i \sqrt{d^2(i, G)}$ avec $\epsilon_1, \epsilon_2 \in \{-1, 1\}$. F étant centrée, on pourra donc prendre $F(1) = -\frac{1}{2}\sqrt{\frac{7}{2}}$ et $F(2) = \frac{1}{4}\sqrt{\frac{7}{2}}$. Pour le calcul de G, on se sert des formules de transition :

$$G(j) = \frac{1}{\sqrt{\lambda}} \sum_{i=1}^{2} f_i^j F(i) = \tfrac{4}{\sqrt{7}}\Big(-f_1^j \tfrac{1}{2}\sqrt{\tfrac{7}{2}} + f_2^j \tfrac{1}{4}\sqrt{\tfrac{7}{2}}\Big) = \tfrac{1}{\sqrt{2}}(f_2^j - 2f_1^j).$$

Or $f_I^J = \begin{pmatrix} 1/4 & 3/4 & 0 \\ 3/4 & 1/4 & 1 \end{pmatrix}$, d'où $G(1) = \frac{1}{\sqrt{2}}(\frac{3}{4} - \frac{2}{4}) = \frac{1}{4\sqrt{2}} \approx 0.18$.
De même, $G(2) = \frac{1}{\sqrt{2}}(\frac{1}{4} - \frac{6}{4}) = -\frac{5}{4\sqrt{2}}$ et $G(3) = \frac{1}{\sqrt{2}}$.

Représentation simultanée.
$F(1) \approx -0.935, F(2) \approx 0.468, G(1) \approx 0.18, G(2) \approx -0.88, G(3) \approx 0.71.$



Remarque. On peut aussi procéder de la façon suivante pour répondre à la question 3-d). On commence par observer qu'il n'existe qu'un axe factoriel non trivial puisqu'il n'y a que deux profils-lignes. Puis on détermine le facteur φ associé à cet axe on résout le système suivant : $f_{1.}\varphi^1 + f_{2.}\varphi^2 = 0$ et $f_{1.}(\varphi^1)^2 + f_{2.}(\varphi^2)^2 = 1$.
Au signe près, il n'existe que la solution suivante pour le vecteur φ : $\varphi^1 = -\frac{2}{\sqrt{2}}$ et $\varphi^2 = \frac{1}{\sqrt{2}}$. Par ailleurs, d'après les formules de transition,

$$G(j) = f_I^j \varphi^I = \frac{1}{\sqrt{2}}\Big(f_2^j - 2f_1^j\Big).$$

D'où $G(1) = \frac{1}{4\sqrt{2}}$, $G(2) = -\frac{5}{4\sqrt{2}}$ et $G(3) = \frac{1}{\sqrt{2}}$. Comme $\mathrm{Var}(G) = \lambda$, on en déduit que :

$$\lambda = \frac{1}{3}\big(G(1)^2 + G(2)^2 + G(3)^2\big) = \frac{7}{16}.$$

Finalement, en utilisant la formule $F(i) = \sqrt{\lambda}\,\varphi^i$, on obtient les valeurs de $F(i)$ pour $i = 1, 2$.

3-e) *Indication* : k_{IJ} satisfait aux conditions définies au début de la première partie, avec $a = b = N = 8, p = n = 3$. Donc en utilisant les résultats du 3) de la première partie, on en déduit les résultats de l'ACP de k_{IJ} à partir de ceux de l'AFC ...

EXERCICE 1

1) $f_I = \dfrac{k(I)}{12} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}'$ et $f_J = \dfrac{k(J)}{12} = \begin{pmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}'$

2) $F_1 = \begin{pmatrix} 1 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$ $\quad F_2 = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}'$

3) $g_J = f_J$.

4) $\rho^2(i_3) = d^2(i_3, g_J) = \|f_J^{i_3} - g_J\|^2_{D_{1/f_J}} = 2 \times 6 \times \dfrac{1}{6^2} + 3 \times \dfrac{1}{3^2} + 2 \times 6\Big(\dfrac{1}{2} - \dfrac{1}{6}\Big)^2 = 2.$

5) $I_T = \displaystyle\sum_{i \in I} f_i\, \rho^2(i) = \frac{1}{3}\Big(2 \times \frac{5}{4} + 2\Big) = \frac{3}{2}.$

6) $(F_1F_2)' = \begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 3/4 & 0 \\ 0 & 0 & 1 \end{pmatrix}$; $\det((F_1F_2)' - \lambda \mathbb{I}_3) = (1-\lambda)[(\frac{3}{4} - \lambda)^2 - \frac{1}{16}] = (1 - \lambda)^2(\frac{1}{2} - \lambda)$. Donc 1 est à la fois valeur propre triviale (cf. cours) et non triviale, et il existe 2 valeurs propres non triviales $\lambda \in \{1, 0.5\}$, d'où $r = 2$. Soit $(x, y, z)'$ un vecteur propre de $(F_1F_2)'$ associé à λ. Si $\lambda = \frac{1}{2}$, on a $\frac{3x}{4} + \frac{y}{4} = \frac{x}{2}$, $\frac{x}{4} + \frac{3y}{4} = \frac{y}{2}$ et $z = \frac{z}{2}$, qui équivaut à $y = -x$ et $z = 0$. Si $\lambda = 1$, on a seulement $y = x$.
Par ailleurs, $\|\psi_2\|^2_{D_{f_I}} = \frac{1}{2}$, donc $\frac{1}{3}((\psi_2^{i_1})^2 + (\psi_2^{i_2})^2) = \frac{1}{2}$, *i.e.* $\psi_2^{i_1} = \frac{\sqrt{3}}{2}$, $\psi_2^{i_2} = -\frac{\sqrt{3}}{2}$ et $\psi_2^3 = 0$, car $\psi_\alpha^{i_1} > 0$. Enfin, ψ_1 vérifie $\psi_1^{i_1} + \psi_1^{i_2} + \psi_1^{i_3} = 0$ et $\frac{1}{3}((\psi_1^{i_1})^2 + (\psi_1^{i_2})^2 + (\psi_1^{i_3})^2) = \lambda = 1$, d'où $\psi_1^{i_1} = \frac{1}{\sqrt{2}}$, $\psi_1^{i_2} = \frac{1}{\sqrt{2}}$ et $\psi_1^{i_3} = -\sqrt{2}$.

7-a) On doit avoir $\overline{\varphi}_1 = 0$ et $\mathrm{Var}(\varphi_1) = 1$. D'où :

$$\frac{1}{3}(a + a + b) = 0 \quad \text{et} \quad \frac{1}{3}(a^2 + a^2 + b^2) = 1$$

On en déduit que $b = -2\,a$ et $6\,a^2 = 3$, et donc $a = \frac{1}{\sqrt{2}}$ et $b = -\sqrt{2}$ car $a > 0$.

7-b) D'après les formules de transition, $\psi_1^j = \Big(f_I^j\Big)'\varphi_1$. On en déduit que :
$\psi_1^A = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \varphi_1 = \frac{1}{\sqrt{2}}$; $\psi_1^B = \psi_1^C = \frac{1}{\sqrt{2}}$; $\psi_1^D = \psi_1^E = -\sqrt{2}$.

7-c) $\lambda_1 = \mathrm{Var}(\psi_1) = \frac{1}{6}\,(\psi_1^A)^2 + \frac{1}{3}\,(\psi_1^B)^2 + \frac{1}{6}\,(\psi_1^C)^2 + \frac{1}{6}\,(\psi_1^D)^2 + \frac{1}{6}\,(\psi_1^E)^2 = 1.$
Donc $\psi_1^i = \varphi_1^i$ pour tout $i \in I$, c'est-à-dire $\psi_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & -\sqrt{2} \end{pmatrix}'$.

7-d) Remarquons qu'il y a deux valeurs propres non triviales car $|I| = 3$ et $\lambda_1 < I_T$. Donc $\lambda_2 = I_T - \lambda_1 = \frac{3}{2} - 1 = \frac{1}{2}$. Les relations vérifiées par ψ_2 sont donc :
$\overline{\psi}_2 = 0$, $\quad \psi_2 \perp_{D_{F_I}} \psi_1$ et $\mathrm{Var}(\psi_2) = \frac{1}{2}$. En posant $\psi_2 = \begin{pmatrix} x & y & z \end{pmatrix}'$, on obtient :

$$x + y + z = 0, \quad \frac{1}{\sqrt{2}}x + \frac{1}{\sqrt{2}}y - \sqrt{2}z = 0 \quad \text{et} \quad \frac{1}{3}(x^2 + y^2 + z^2) = \frac{1}{2}.$$

D'où $z = 0$, $y = -x$ et donc $x = \frac{\sqrt{3}}{2}$, $y = -\frac{\sqrt{3}}{2}$ et $z = 0$, car $\psi_2^{i_1}$ est supposé > 0.

8) Pour $\alpha \in \{1, 2\}$ et $i \in I$, on a $\mathrm{COR}_\alpha(i) = \frac{\psi_\alpha(i)^2}{\rho^2(i)}$, d'où $\mathrm{COR}_1(i_1) = \mathrm{COR}_1(i_2) = \frac{2}{5}$ et donc $\mathrm{COR}_2(i_1) = \mathrm{COR}_2(i_2) = 1 - \frac{2}{5} = \frac{3}{5}$.

EXERCICE 2

1-a) Par définition, $g = f_I = \frac{k_I}{n\,\mathrm{Card}Q}$. Or $k(i) = \mathrm{Card}Q$ pour tout $i \in I$. Donc $g = \frac{1}{n}\mathbb{1}_n$.

1-b)
$$\begin{aligned}
\rho^2(j) &= \|f_I^j - g\|^2_{D_{1/f_I}} = n \sum_{i\in I} \Big(f_i^j - \frac{1}{n}\Big)^2 = n \sum_{i\in I} \Big(\frac{k(i,j)}{k(j)} - \frac{1}{n}\Big)^2 \\
&= n \sum_{i\in I} \Big(\frac{k(i,j)}{n\,p(j)} - \frac{1}{n}\Big)^2 = \frac{1}{n} \sum_{i\in I} \Big(\frac{k(i,j)}{p(j)} - 1\Big)^2 = \frac{1}{n}\Big(\sum_{i\in I} \frac{k(i,j)}{p^2(j)} - 2\sum_{i\in I}\frac{k(i,j)}{p(j)} + n\Big) \\
&= \frac{1}{n}\Big(\frac{n\,p(j)}{p^2(j)} - 2\,\frac{n\,p(j)}{p(j)} + n\Big) = \frac{1}{p(j)} - 1 = \frac{1 - p(j)}{p(j)}.
\end{aligned}$$

1-c)
$$\begin{aligned}
I_T &= \sum_{j\in J} f_j\ \rho^2(j) = \sum_{q\in Q}\sum_{j\in J_q} f_j\ \rho^2(j) = \sum_{q\in Q}\sum_{j\in J_q} \frac{k(j)}{n\,\mathrm{Card}Q}\,\frac{1-p(j)}{p(j)} = \sum_{q\in Q}\sum_{j\in J_q}\frac{1-p(j)}{\mathrm{Card}Q} \\
I_T &= \sum_{q\in Q} \frac{(\mathrm{Card}J_q) - 1}{\mathrm{Card}Q}.
\end{aligned}$$

2-a) $\widetilde{f}_j = \frac{\widetilde{k}(j)}{n}$ car la somme des éléments du tableau k_{IJ_q} est égale à n. Or $\widetilde{k}(j) = k(j)$, d'où $\widetilde{f}_j = \frac{k(j)}{n} = p(j)$.

2-b) On a $\widetilde{g} = \sum_{j\in J_q} \widetilde{f}_j \widetilde{f}_I^j$. Donc, pour tout $i \in I$, on a : $\widetilde{g}_i = \sum_{j\in J_q} \widetilde{f}_j \widetilde{f}_i^j = \sum_{j\in J_q} p(j)\frac{k(i,j)}{k(j)}$.
Comme $k(j) = n\,p(j)$, on a $\widetilde{g}_i = \frac{1}{n}\sum_{j\in J_q} k(i,j) = \frac{\widetilde{k}(i)}{n}$. Or $\widetilde{k}(i) = 1$ car chaque individu i ne choisit qu'une modalité de réponse à chaque question q, donc $\widetilde{g}_i = \frac{1}{n}$, d'où $\widetilde{g} = g$.

3) D'après la question 2-b), il existe $\mathrm{Card}Q$ relations linéairement indépendantes entre les profils colonnes du tableau k_{IJ}. Donc la dimension du support du nuage $\mathcal{N}(J)$ est inférieure ou égale à $(\mathrm{Card}J - \mathrm{Card}Q)$, d'où $r \leq (\mathrm{Card}J - \mathrm{Card}Q)$.

PROBLÈME 1

1) Il s'agit de la métrique D_{1/f_I} avec $f_I = \Big(\dfrac{k(i)}{k}\Big)_{i\in I} = \big(1/4,\dots,1/4\big)'$. D'où $D_{1/f_I} = 4\,\mathrm{Id}$.

2) On a $H_1 = \begin{pmatrix} 1 & 0 & 3/4 & 0 & 0 & 0 \\ 0 & 1 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 3/4 \\ 0 & 0 & 0 & 0 & 1 & 1/4 \end{pmatrix}$ et $H_2 = \begin{pmatrix} 2/5 & 0 & 0 & 0 \\ 0 & 4/5 & 0 & 0 \\ 3/5 & 1/5 & 0 & 0 \\ 0 & 0 & 2/5 & 0 \\ 0 & 0 & 0 & 4/5 \\ 0 & 0 & 3/5 & 1/5 \end{pmatrix}$,

d'où, $H_1H_2 = \dfrac{1}{20}\begin{pmatrix} 17 & 3 & 0 & 0 \\ 3 & 17 & 0 & 0 \\ 0 & 0 & 17 & 3 \\ 0 & 0 & 3 & 17 \end{pmatrix}$.

3) $H_1H_2\,v = H_1H_2\,(f_I^{j_2} - f_I^{j_1}) = H_1H_2\begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \dfrac{1}{20}\begin{pmatrix} -14 \\ 14 \\ 0 \\ 0 \end{pmatrix} = \dfrac{7}{10}\,v$. Donc $\lambda = \dfrac{7}{10}$.

4) Soient a, b, c, d les coordonnées de w. On a $H_1H_2\,v = \dfrac{7}{10}\,v$ et on peut choisir w tel que $w \perp v$. D'où les cinq équations : $-a + b = 0$; $17\,a + 3\,b = 14\,a$; $3\,a + 17\,b = 14\,b$; $17\,c + 3\,d = 14\,c$ et $3\,c + 17\,d = 14\,d$. Les trois premières équations sont équivalentes à $a = b = 0$ et les deux dernières à $d = -c$. Donc un choix possible pour w est le vecteur $(0, 0, 1, -1)'$.

5) On a $I_T = \mathrm{tr}(F_1F_2) - 1$; d'où $I_T = 4 \times \dfrac{17}{20} - 1 = \dfrac{12}{5}$.
On peut aussi calculer I_T selon la formule $I_T = \sum_{j\in J} f_j\,\rho^2(j)$ avec :
$f_J' = \Big(1/10, 1/5, 1/5, 1/10, 1/5, 1/5\Big)$
et $\rho^2(j_1) = \rho^2(j_2) = \rho^2(j_4) = \rho^2(j_5) = 4\,\big(1 - \dfrac{1}{4}\big)^2 + 3 \times 4\big(\dfrac{1}{4}\big)^2 = 9/4 + 3/4 = 3$.
De même, on a : $\rho^2(j_3) = \rho^2(j_6) = 4\,\big(\dfrac{3}{4} - \dfrac{1}{4}\big)^2 + 2 \times 4\big(\dfrac{1}{4}\big)^2 = 1 + 1/2 = 3/2$.
D'où $I_T = 2 \times \dfrac{1}{10} \times 3 + 2 \times \dfrac{1}{10} \times 3 + 2 \times \dfrac{1}{10} \times \dfrac{3}{2} = \dfrac{12}{5}$.

6) Il existe au plus trois valeurs propres (car $|\,I\,| = 4 < |\,J\,|$). Deux d'entre-elles sont égales à $\dfrac{7}{10}$ (cf. 3) et 4)). Comme $I_T = \dfrac{12}{5}$, il existe une troisième val. propre égale à $\dfrac{12}{5} - \big(2 \times \dfrac{7}{10}\big) = 1$.

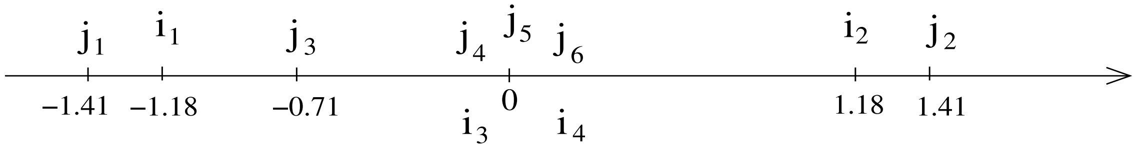
7) On a $\|v\|^2 = 4\,\big((-1)^2 + 1^2\big) = 8$. D'où $u_I = \dfrac{1}{2\sqrt{2}}(-1, 1, 0, 0)'$.

8) On a $\psi_2^j = (f_I^j)' D_{1/f_I} u_I = 4\,(f_I^j)' u_I = \sqrt{2}(f_{i_2}^j - f_{i_1}^j)$.

9) On a $\psi_2^I = \sqrt{\lambda}\,\varphi_2^I = \sqrt{\lambda}\,M\,u_I$. Donc ici $\psi_2^I = 4\sqrt{\dfrac{7}{10}}\,u_I = 2\sqrt{\dfrac{14}{5}}\,u_I$, ce qui prouve le résultat.

10) On a $\mathrm{COR}_2(j) = \dfrac{(\psi_2^j)^2}{\rho^2(j)}$ avec $\rho^2(j) = \| f_I^j - f_I \|^2 = 4\sum_{i\in I}\Big(f_i^j - \dfrac{1}{4}\Big)^2$.
D'où $\rho^2(j_1) = \rho^2(j_2) = \dfrac{9}{4} + \dfrac{3}{4} = 3$ et donc : $\mathrm{COR}_2(j_1) = \mathrm{COR}_2(j_2) = \dfrac{2}{3}$.

11) D'après 8) on a : $\psi_2^{j_1} = -\psi_2^{j_2} = -\sqrt{2} \approx -1.41$; $\psi_2^{j_3} = -\dfrac{1}{\sqrt{2}} \approx -0.71$ et $\psi_2^{j_4} = \psi_2^{j_5} = \psi_2^{j_6} = 0$.
D'après 9), $\psi_2^{i_1} = -\sqrt{\dfrac{7}{5}} \approx -1.18$; $\psi_2^{i_2} = \sqrt{\dfrac{7}{5}} \approx 1.18$ et $\psi_2^{i_3} = \psi_2^{i_4} = 0$. D'où la représentation simultannée :



PROBLÈME 2

1) D'après le cours, $f_{I'} = \begin{pmatrix} k_I/2k \\ k_I/2k \end{pmatrix} = \dfrac{1}{2}\begin{pmatrix} f_I \\ f_I \end{pmatrix}$ est vecteur directeur d'un axe factoriel trivial, et donc $\begin{pmatrix} f_I \\ f_I \end{pmatrix}$ l'est aussi.

2) Il est clair que $H_1 = \begin{pmatrix} F_1 & 0 \\ 0 & F_1 \end{pmatrix}$ et $H_2 = \begin{pmatrix} F_2 & 0 \\ 0 & F_2 \end{pmatrix}$. Donc $H_1H_2 = \begin{pmatrix} F_1F_2 & 0 \\ 0 & F_1F_2 \end{pmatrix}$.

3) Soit u_I un vecteur axial factoriel de l'AFC de K associé à la valeur propre $\lambda \neq 0$:
$H_1H_2\begin{pmatrix} u_I \\ 0_p \end{pmatrix} = \begin{pmatrix} F_1F_2u_I \\ 0_p \end{pmatrix} = \lambda\begin{pmatrix} u_I \\ 0_p \end{pmatrix}$. De même $H_1H_2\begin{pmatrix} 0_p \\ u_I \end{pmatrix} = \begin{pmatrix} 0_p \\ F_1F_2u_I \end{pmatrix} = \lambda\begin{pmatrix} 0_p \\ u_I \end{pmatrix}$. Ce qui prouve le résultat car $\begin{pmatrix} u_I \\ 0_p \end{pmatrix}$ et $\begin{pmatrix} 0_p \\ u_I \end{pmatrix}$ sont orthogonaux pour la métrique $D_{1/f_{I'}}$.

4) $H_1H_2\begin{pmatrix} f_I \\ 0_p \end{pmatrix} = \begin{pmatrix} F_1F_2f_I \\ 0_p \end{pmatrix} = \begin{pmatrix} f_I \\ 0_p \end{pmatrix}$, ce qui prouve le résultat demandé.

5) Posons $v = \begin{pmatrix} f_I \\ 0_p \end{pmatrix}$. Si v est colinéaire à un axe factoriel, alors Δv est un axe factoriel associé à la valeur propre 1, donc v serait $D_{1/f_{I'}}$-orthogonal à l'axe factoriel trivial $f_{I'}$.
Or $\langle v, f_{I'}\rangle_{D_{1/f_{I'}}} = \Big\langle \begin{pmatrix} f_I \\ 0_p \end{pmatrix}, \dfrac{1}{2}\begin{pmatrix} f_I \\ f_I \end{pmatrix} \Big\rangle_{D_{1/f_{I'}}} = 2\langle f_I, \dfrac{1}{2} f_I\rangle_{D_{1/f_I}} = \sum_{i\in I} f_{i\,\cdot}\,\dfrac{f_{i\,\cdot}}{f_{i\,\cdot}} = 1$, ce qui est contradictoire, donc v n'est colinéaire à aucun axe factoriel.

6) Soit w un vecteur directeur d'un tel axe. Donc w est vecteur propre de H_1H_2 associé à la valeur propre 1 et est nécessairement orthogonal à $f_{I'}$. Comme v et $f_{I'}$ sont des vecteurs propres de H_1H_2 associé à la valeur propre 1, on peut choisir w comme étant la projection orthogonale de v sur l'orthogonal de $f_{I'}$: $w = (\mathrm{Id} - P_{\Delta f_{I'}})(v)$. D'après 5), on a vu que $\langle v, f_{I'} \rangle_{D_{1/f_{I'}}} = 1$. De plus :

$$\langle f_{I'}, f_{I'} \rangle_{D_{1/f_{I'}}} = \left\langle \frac{1}{2}\begin{pmatrix} f_I \\ f_I \end{pmatrix}, \frac{1}{2}\begin{pmatrix} f_I \\ f_I \end{pmatrix} \right\rangle_{D_{1/f_{I'}}} = \frac{1}{2}\langle f_I, f_I \rangle_{D_{1/f_I}} + \frac{1}{2}\langle f_I, f_I \rangle_{D_{1/f_I}} = 1.$$

Donc $P_{\Delta f_{I'}}(v) = \left(\langle v, f_{I'} \rangle_{D_{1/f_{I'}}}\right) f_{I'} = f_{I'}$, et par conséquent $w = v - f_{I'} = \frac{1}{2}\begin{pmatrix} f_I \\ -f_I \end{pmatrix}$.

7) Dans ce cas, Z est de dimension $(rp \times rq)$ et l'AFC de Z comporte au plus $(rp-1)$ valeurs propres non nulles. Comme les vecteurs de la forme $(f_I', 0_p', \ldots, 0_p', -f_I', 0_p', \ldots, 0_p')'$ sont vecteurs propres associés à la valeur propre 1, il existe au moins $(r-1)$ axes factoriels. Par ailleurs, à chacun des $(p-1)$ vecteurs axiaux factoriels, noté u^α, de l'AFC de K et relatif à une valeur propre (simple) non triviale et distincte de 1, notée λ_α, correspond r axes factoriels de l'AFC de Z associés à la valeur propre λ_α et dirigés par des vecteurs directeurs de la forme $(0_p', \ldots, 0_p', (u^\alpha)', 0_p', \ldots, 0_p')'$ avec $\alpha \in \{1, \ldots, p-1\}$. Il existe donc $r(p-1)$ axes de ce type associés à des valeurs propres $\lambda_\alpha \neq 1$. Comme $rp - 1 = r - 1 + r(p-1)$, cela prouve qu'il existe exactement $r-1$ axes factoriels associés à la valeur propre 1, d'où le résultat.