Analyse des Données

Travaux Dirigés

2017-2018

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 :

(i) P est idempotent,
(ii) $\mathrm{Ker}P = (I - P)(E)$,
(iii) P est un projecteur.

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, autrement 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* :

$$\begin{cases} XMy = XMX'b \\ P\,y = X'b \end{cases}$$

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

Chaque individu i est muni d'un poids p_i tel que :

$$\sum_{i \in I} p_i = 1, \text{ avec } I = \{1, \ldots, n\}$$

Les poids p_i sont généralement égaux à $\dfrac{1}{n}$.

Rappelons les définitions suivantes :

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

$$v_{jj'} = \sum_{i \in I} p_i (x_i^j - \overline{x}^j)(x_i^{j'} - \overline{x}^{j'}) = \sum_{i \in I} p_i \, y_i^j \, y_i^{j'}.$$

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

On note :

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

vecteur des moyennes — individu i après centrage — variable j 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_i, X la matrice $(n \times p)$ des données x_i^j et Y la matrice $(n \times p)$ des données centrées y_i^j.

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

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

3. De même, à partir de la relation : $y_i^j = x_i^j - \overline{x}^j$ avec j variant de 1 à p, montrer que y_i est l'image de x_i, 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 $v_{jj'}, s_j^2$ et $r_{jj'}$.

5. Soit V la matrice variance empirique des p variables, c'est-à-dire la matrice $p \times p$ dont le terme général est $v_{jj'}$. 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 [†] :

$$\forall z \in \mathbb{R}^p, \quad u^\star(z) = \sum_{j=1}^{p} u_j \, z_j$$

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

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

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_i, peut être identifié à l'ensemble formé par les n vecteurs $x_i \in E$ représentant ces points :

$$\mathcal{M} = \{x_i \mid i = 1, ..., n\},$$

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 :

$$g = \frac{1}{p} \sum_{i=1}^{n} p_i \, x_i, \text{ avec } p = \sum_{i=1}^{n} p_i.$$

Dans tout le texte, on suppose que $g = 0$ et que l'inertie totale du nuage $\mathcal{M}$ peut s'écrire sous la forme :

$$I_T(\mathcal{M}) = \sum_{i=1}^{n} p_i \, \|x_i - g\|^2 = \sum_{i=1}^{n} p_i \, \|x_i\|^2 .$$

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

$$I_W(\mathcal{M}) = \sum_{i=1}^{n} p_i \, \|x_i - P_W(x_i)\|^2.$$

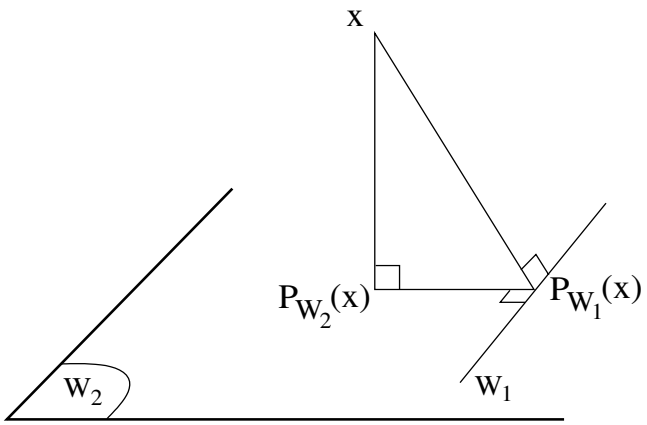
En utilisant le fait que l'application linéaire $I - P_W$ est le projecteur M-orthogonal sur $W^\perp$, montrer les relations suivantes :

a) $I_{W^\perp}(\mathcal{M}) = I_T(P_W(\mathcal{M}))$,
b) $I_T(\mathcal{M}) = I_W(\mathcal{M}) + I_{W^\perp}(\mathcal{M})$.

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 :

(i) $W_1 \subseteq W_2$,
(ii) $W_2^\perp \subseteq W_1^\perp$,
(iii) $P_{W_1} = P_{W_1} \circ P_{W_2}$.

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 :

$I \setminus J$	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 I de six individus. On suppose que chaque individu i de I $(1 \leq i \leq 6)$ est muni de la masse $1/6$. On note X le tableau associé.

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 $\mathrm{INR}(i)$, $\mathrm{CTR}_\alpha(i)$ et $COR_\alpha(i)$.

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.

2e 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.**

3e Partie

On désire à présent faire l'A.C.P. des données en utilisant la métrique de $\mathbb{R}^3$ dont la matrice est $M = \mathrm{Diag}(a, b, c) = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$. 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 $M = \text{Diag}(a = 1, b = 2, c = 1)$.

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_i est muni du poids p_i. On suppose que pour tout i et i', on connaît le carré de la distance, noté $d_{ii'}$, entre les points P_i et $P_{i'}$. Par conséquent :

$$d_{ii'} = d_{i'i} \geq d_{ii} = 0,$$

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

$$(P_iP_{i'})'\mathrm{M}(P_iP_{i'}) = \|P_iP_{i'}\|^2 = d_{ii'}.$$

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_i :

$$\text{Pour tout } i, \qquad d_{i\,\bullet} = \sum_{i'=1}^{n} \frac{p_{i'}}{p}\, d_{ii'}, \quad d_{\bullet\,\bullet} = \sum_{i=1}^{n} \frac{p_i}{p}\, d_{i\,\bullet} \quad \text{et} \quad \mathbf{D_p} = \mathrm{Diag}(p_i).$$

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 $d_{ii'}$, 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.

1ère partie : Etude dans le cas général. Dans cette question, on suppose que l'on connaît la matrice $\mathbf{Y}$ (centrée) du type individus $\times$ variables 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 :

$$\text{a)} \quad \mathbf{Y}'\,\mathbf{D_p}\mathbf{Y}\,\mathrm{M}\,\mathbf{u} = \lambda\,\mathbf{u},$$
$$\text{b)} \quad \mathbf{F} = \mathbf{Y}\mathrm{M}\,\mathbf{u},$$

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 :

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

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

1.3. De la relation $\sum_{i=1}^{n} p_i \, (GP_i) = 0$, 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 $d_{ii'}$.

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 :

$$p \, d_{\bullet i} = I_G + p \, \|GP_i\|^2$$

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 $i, i' \in \{1, \ldots, n\}$, on a :

$$(GP_i)' \mathrm{M} \, (GP_{i'}) = \frac{1}{2} \big(d_{i\bullet} + d_{i'\bullet} - d_{\bullet\bullet} - d_{ii'} \big).$$

2 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

Dans ce TP, nous utilisons le logiciel Factorminer pour réaliser une ACP.

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".

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

On commence par charger la librairie Factominer par la commande

```
library(FactoMineR)
```

puis on importe les données

```
depenses <- read.table("/Users/dp/Dropbox/mido/analyse-des-donnees-M1/tp/depense-men
```

1. La commande summary donne les statistiques de base pour chaque variable.

```
summary(depenses)
```

Commenter brièvement le tableau des statistiques sommaires.

2. La commande cor donne les statistiques de base pour chaque variable.

```
cor(depenses)
```

Que peut-on dire des corrélations entre variables ?

3. Pour effectuer une analyse en composantes principales, il faut préciser les variables qualitatives utilisées et les variables supplémentaires. Dans notre cas, les variables qualitatives vont de la colonne 1 à la colonne 7 et on choisit de mettre les variables de la colonne 8 à la colonne 9 en variables supplémentaires. La commande pour une ACP normalisée est

```
res.pca <- PCA(depenses,quanti.sup=9,quali.sup=8)
```

A la suite de cette commande, les nuages des individus et des variables apparaissent. Dans le nuage des individus, on peut retirer les variables qualitatives avec la commande

```
plot(res.pca, invisible="quali")
```

4. Pour afficher les valeurs propres, on utilise la commande

```
res.pca\$eig
```

Calculer la contribution relative $\mathrm{INR}(j)$ de chaque variable x^j à l'inertie totale I_T. Combien d'axes factoriels pensez-vous qu'il faille retenir ? que vaut l'inertie totale ?

5. Pour afficher les coordonnées des variables, on utilise la commande

```
round(res.pca$var$coord,2)
```

Puis pour afficher les contributions des variables, on utilise la commande

```
round(res.pca$var$contrib,2)
```

Enfin pour afficher le carré des corrélations variables facteurs, on utilise la commande

```
round(res.pca$var$cos2,2)
```

Quelles sont les variables qui contribuent le plus au premier axe factoriel, puis au deuxième ? Calculer la qualité de la représentation de chaque variable dans le plan factoriel 1×2. Interpréter le cercle des corrélations.

Graphiquement, il est possible de représenter les variables dont la qualité de représentation dépasse un certain seuil. Pour cela, on utilise la commande

```
plot.PCA(res.pca,choix="var",select="cos2 0.9")
```

6. Pour afficher les coordonnées des individus, on utilise la commande

```
round(res.pca$ind$coord,2)
```

Puis pour afficher les contributions des individus, on utilise la commande

```
round(res.pca$ind$contrib,2)
```

Enfin pour afficher le carré des corrélations individus facteurs, on utilise la commande

```
round(res.pca$ind$cos2,2)
```

Quelles sont les individus qui contribuent le plus au premier axe factoriel, puis au deuxième ? Calculer la qualité de la représentation de chaque individu dans le plan factoriel 1×2. 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).

Graphiquement, il est possible de représenter les variables dont la qualité de représentation dépasse un certain seuil. Pour cela, on utilise la commande

```
plot.PCA(res.pca,choix="ind",select="cos2 0.9")
```

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.

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

	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)

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

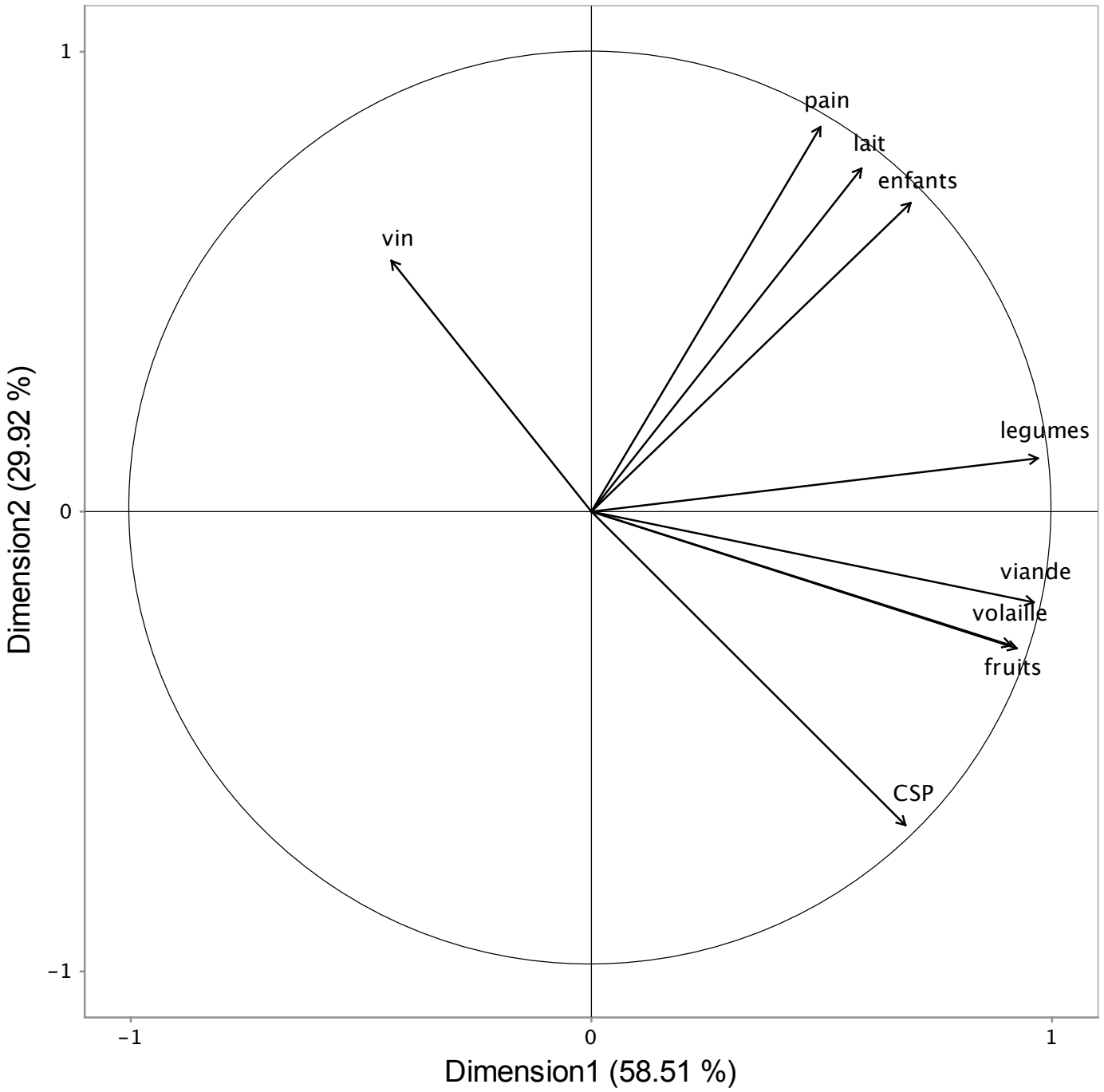
f) *Qualité de représentation des individus* ($\cos^2$)

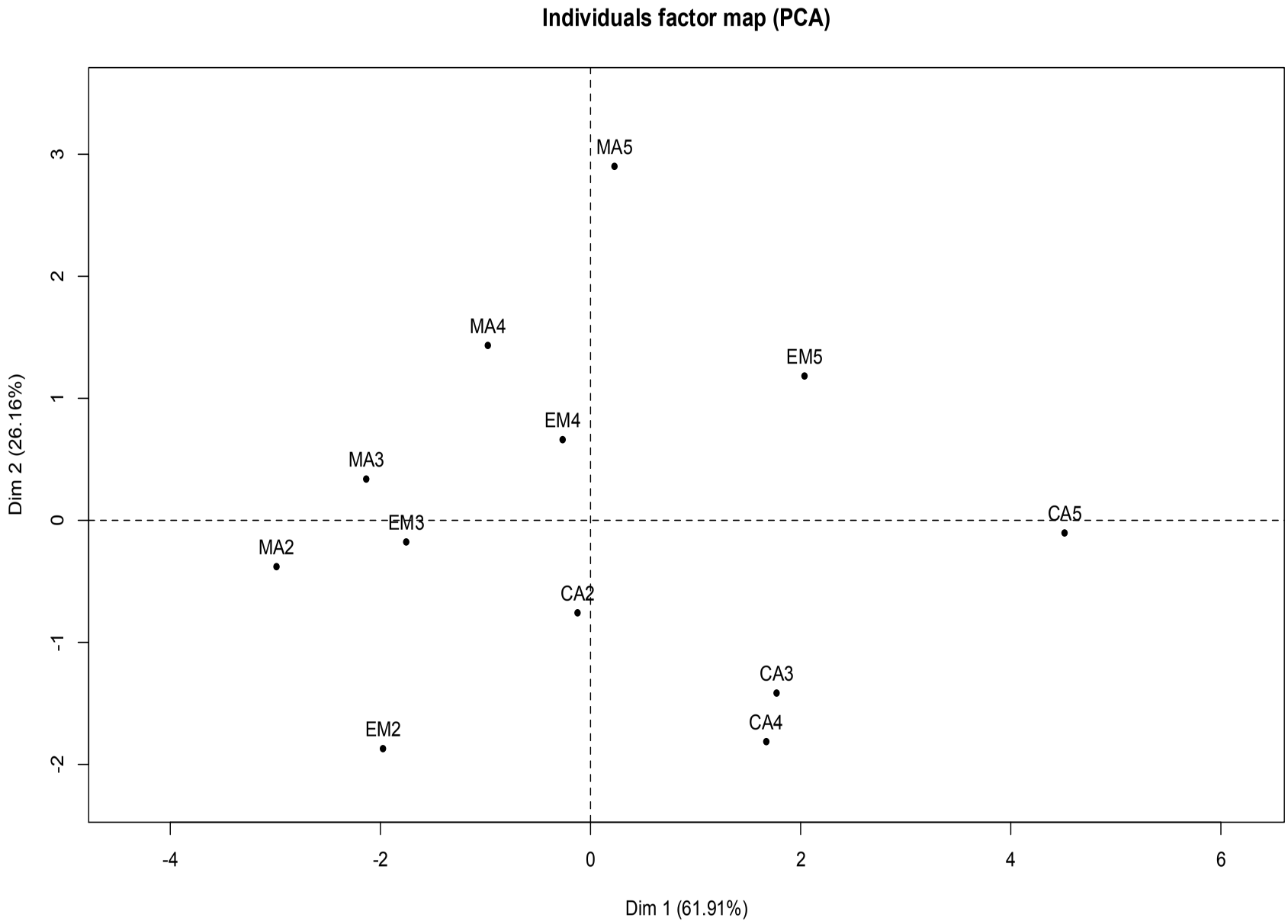
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
MA2	0.94	0.02	0.02	0.02	0.01
EM2	0.42	0.38	0.20	0.00	0.00
CA2	0.00	0.19	0.72	0.01	0.07
MA3	0.97	0.02	0.00	0.00	0.00
EM3	0.89	0.01	0.08	0.01	0.01
CA3	0.48	0.31	0.17	0.03	0.00
MA4	0.30	0.64	0.03	0.02	0.00
EM4	0.10	0.61	0.11	0.13	0.04
CA4	0.43	0.50	0.00	0.03	0.03
MA5	0.01	0.95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

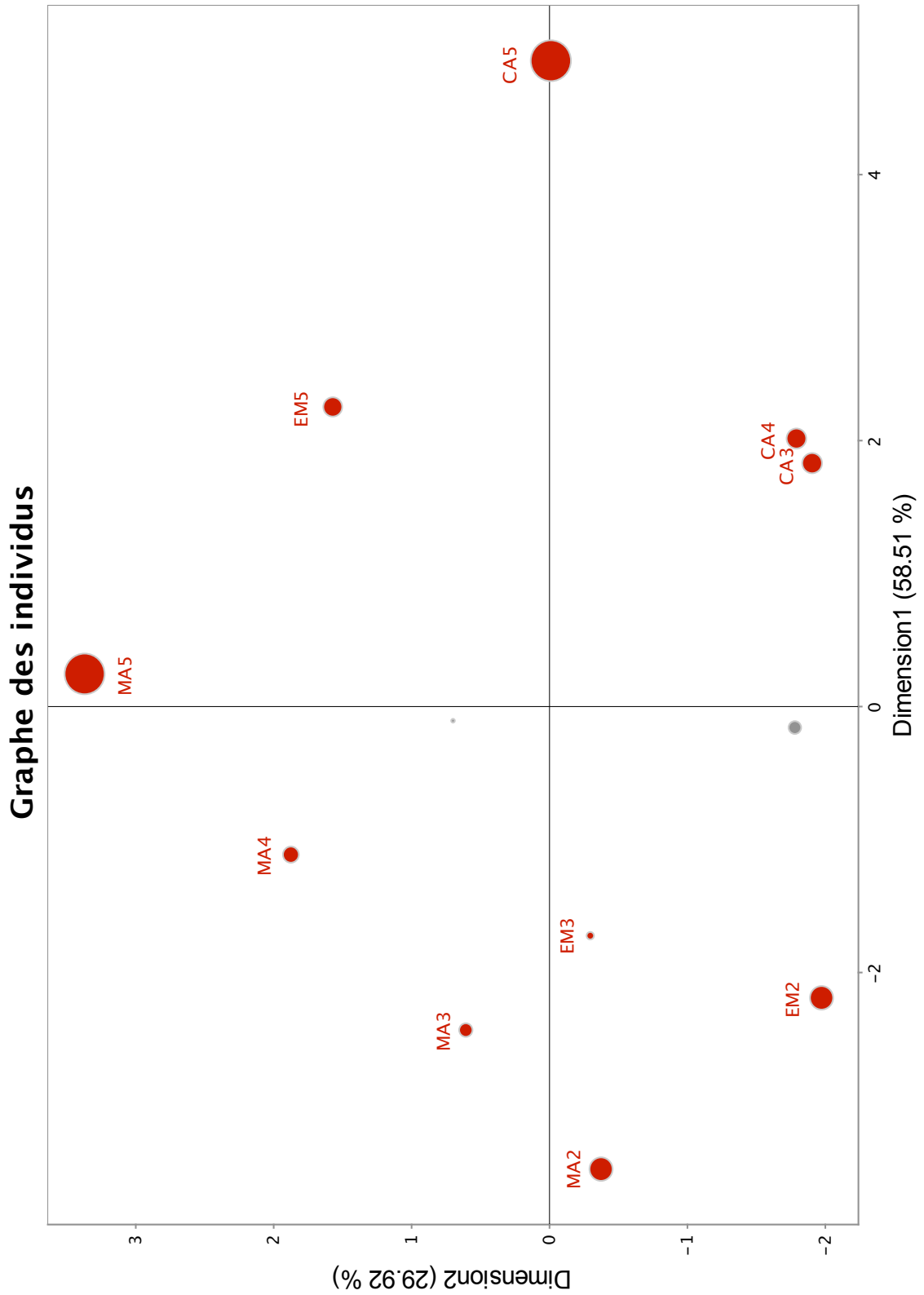
	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
MA2	17.18	0.66	2.32	9.40	7.92
EM2	7.50	15.93	24.73	1.77	1.39
CA2	0.03	2.61	29.14	2.68	31.46
MA3	8.76	0.52	0.17	0.79	0.01
EM3	5.92	0.14	3.62	1.52	4.56
CA3	6.03	9.12	14.28	13.19	0.81
MA4	1.84	9.35	1.11	4.86	1.39
EM4	0.14	1.99	1.07	5.91	4.01
CA4	5.38	14.96	0.14	11.31	28.22
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)
MA5
MA4
EM5
EM4
MA3
EM3
CA5
MA2
CA2
CA3
CA4
EM2
Dim 2 (26.16%)
Dim 1 (61.91%)
3
2
1
0
-1
-2
-4
-2
0
2
4
6

Affichage des individus ayant un indice $\mathrm{QLT}_2 > .75$ par des disques proportionnels à leurs CTR_{12}.



TD 7
Analyse Factorielle des Correspondances

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 $i^{ème}$ 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 $j^{ème}$ 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 $f_I^j - g_I$.

2ème 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. $G_\alpha(j)$) ($\alpha = 1, 2$) 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 ?

3ème 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 $F_2 F_1$.

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 :

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

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

$$B_{JJ} = \begin{pmatrix} B_{J_1 J_1} & B_{J_1 J_2} \\ B_{J_2 J_1} & B_{J_2 J_2} \end{pmatrix},$$

où $B_{J_1 J_1}$ est le tableau de Burt en croisant J_1 avec lui-même, et de même pour $B_{J_1 J_2}$, $B_{J_2 J_1}$, $B_{J_2 J_2}$.

3.1. Montrer que les matrices $B_{J_1 J_1}$ et $B_{J_2 J_2}$ sont diagonales et on déterminera les termes de la diagonale.

3.2. On effectue l'analyse des correspondances du tableau $B_{J_1 J_2}$ c'est-à-dire du tableau de terme général $b(j_1, j_2)$ avec $j_1 \in J_1$ et $j_2 \in J_2$.

On note $B^{J_2}_{J_1}$ le profil colonne de $B_{J_1 J_2}$ et $B^{J_1}_{J_2}$ le profil ligne.

On note μ_β la valeur propre non nulle associée à l'axe β et F_β (resp. G_β) la composante principale associée au nuage des profils lignes (resp. profils colonnes).

En utilisant les formules de transition, écrire les relations entre F_β et G_β.

3.3. On étudie l'analyse des correspondances du tableau B_{JJ}.

Exprimer le profil colonne B^J_J comme une matrice par blocs en utilisant $B^{J_2}_{J_1}$, $B^{J_1}_{J_2}$ et les matrices identités I_{q_1} et I_{q_2}.

Montrer que $\begin{pmatrix} F_\beta \\ G_\beta \end{pmatrix}$ et $\begin{pmatrix} F_\beta \\ -G_\beta \end{pmatrix}$ sont des vecteurs propres de $((B^J_J)')^2$ dont on précisera les valeurs propres en fonction de μ_β.

Soit $u \in \ker(B^{J_1}_{J_2})'$, montrer que $\begin{pmatrix} 0 \\ u \end{pmatrix}$ est un vecteur propre de $((B^J_J)')^2$.

De même soit $v \in \ker(B^{J_2}_{J_1})'$, montrer que $\begin{pmatrix} v \\ 0 \end{pmatrix}$ est un vecteur propre de $((B^J_J)')^2$.

3.3. On note q_1 le cardinal de J_1 et q_2 celui de J_2. On note r le rang de $B^{J_2}_{J_1}$.

Rappeler pourquoi r est aussi le rang de $B^{J_1}_{J_2}$, de $B^{J_2}_{J_1} B^{J_1}_{J_2}$ et de $B^{J_1}_{J_2} B^{J_2}_{J_1}$.

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 $B_{J_1 J_2}$.

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 $B_{J_1J_2}$.

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.