Analyse Factorielle des Correspondances (AFC)

Il est important de souligner que ce polycopié est la composition des documents, ci-dessous, que nous l'avons enrichi par quelques details théoriques et applications.
Charlotte Baey (2019). Analyse de données. https://baeyc.github.io/teaching/
Alboukadel Kassambara (2019). http://www.sthda.com/french/
François Husson, Sébastien Lê et Jérôme Pagès (2009). Analyse des données avec R, Presses Universitaires de Rennes, 224 p. (ISBN 978-2-7535-0938-2)

1 Introduction

L'analyse factorielle des correspondances (AFC ou CA pour correspondence analysis en anglais) est une méthode exploratoire d'analyse des tableaux de contingence, et analyser l'association entre deux variables qualitatives (ou catégorielles). Elle a été développée essentiellement par J.-P. Benzecri durant la période 1970-1990.

L'AFC permet de résumer et de visualiser l'information contenue dans le tableau de contingence formé par les deux variables catégorielles. Le tableau de contingence contient les fréquences formées par les deux variables.

L'AFC retourne les coordonnées des éléments des colonnes et des lignes du tableau de contingence. Ces coordonnées permettent de visualiser graphiquement l'association entre les éléments de lignes et de colonnes dans un graphique à deux dimensions.

Lors de l'analyse d'un tableau de contingence, une question typique est de savoir si certains éléments lignes sont associés à certains éléments colonnes. L'analyse factorielle par correspondance est une approche géométrique pour visualiser les lignes et les colonnes d'une table de contingence dans un graphique en nuage de points, de sorte que les positions des points lignes et celles des points colonnes correspondent à leurs associations dans le tableau.

L'analyse factorielle des correspondances (AFC) a été proposée par Jean-Paul Benzécri dans les années 1980. Elle permet d'analyser le lien, encore appelé correspondance, entre deux variables qualitatives. Elle peut être vue comme une ACP particulière, basée sur la métrique du chi-deux.

Exemple 1.1 *Le tableau suivant représente la couleur des cheveux et la couleur des yeux dans un*

*échantillon de **370** individus.*

	V_2			
	$\downarrow$			
V_1	*Brun*	*Châtain*	*Roux*	*Blond*
$\downarrow$				
Marron	68	119	26	7
Noisette	15	54	14	10
Vert	4	29	14	10

Les deux vecteurs V_1 :=(brun,chatain,roux,blond) et V_2 :=(marron,noisette,vert) dites variables catégorielles (qualitatives), et ses composantes s'appelles modalités.

1.1 Tableau de contingence et nuages de points associés

Notons V_1 et V_2 les deux variables qualitatives que l'on souhaite étudier, p le nombre de modalités de la variable V_1 et q le nombre de modalités de la variable V_2. Le point de départ d'une AFC est le tableau de contingence N^* obtenu en croisant les deux variables V_1 et V_2. Plus précisément, on a :

$$N^* = \begin{pmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pq} \end{pmatrix},$$

où x_{ij} est le nombre d'observations (ou effectif) pour lesquelles $V_1 = i$ et $V_2 = j$. On définit l'effectif total par

$$n := \sum_{i=1}^{p} \sum_{j=1}^{q} x_{ij}$$

et la fréquence observée du croisement des deux modalités $i \times j$, par

$$f_{ij} = \frac{x_{ij}}{n} = \mathbf{P}\left(V_1 = i, V_2 = j\right).$$

pour définir

$$N := \begin{pmatrix} f_{11} & \cdots & f_{1q} \\ \vdots & \ddots & \vdots \\ f_{p1} & \cdots & f_{pq} \end{pmatrix}, \text{ (Voir la figure 1).}$$

Les fréquences marginales de V_1 et V_2 sont

$$f_{i\cdot} := \sum_{j=1}^{q} f_{ij} = \mathbf{P}\left(V_1 = i\right) \text{ et } f_{\cdot j} := \sum_{i=1}^{p} f_{ij} = \mathbf{P}\left(V_2 = j\right).$$

Il est important de noter que

$$\sum_{i=1}^{p}\sum_{j=1}^{q} f_{ij} = \sum_{i=1}^{p} f_{i\cdot} = \sum_{j=1}^{q} f_{\cdot j} = 1.$$

En outre, on définit les fréquences conditionnelles associées aux profils-lignes, par

$$f_{i/j} := \mathbf{P}\left(V_1 = i \mid V_2 = j\right) = \frac{\mathbf{P}\left(V_1 = i, V_2 = j\right)}{\mathbf{P}\left(V_2 = j\right)} = \frac{f_{ij}}{f_{\cdot j}}.$$

De même, on définit les fréquences conditionnelles associées aux profils-colonnes par

$$f_{j/i} := \mathbf{P}\left(V_2 = j \mid V_1 = i\right) = \frac{\mathbf{P}\left(V_2 = j, V_1 = i\right)}{\mathbf{P}\left(V_1 = i\right)} = \frac{f_{ij}}{f_{i\cdot}}.$$

On note aussi que

$$\sum_{j=1}^{q} f_{j/i} = 1, \text{ pour chaque } i = 1, ..., p,$$

et

$$\sum_{i=1}^{p} f_{i/j} = 1, \text{ pour chaque } j = 1, ..., q.$$

Voir un tableau qui résume les noations précedentes:

Symboles	Appellations	Autres symboles
x_{ij}	effectif observé	O_{ij} ou n_{ij}
$n = \sum_{i=1}^{p}\sum_{j=1}^{q} x_{ij}$	effectif total	
$x_{i\cdot} = \sum_{j=1}^{q} x_{ij}$	effectif marginal-ligne	$O_{i\cdot}$ ou $n_{i\cdot}$
$x_{\cdot j} = \sum_{i=1}^{p} x_{ij}$	effectif marginal-colonne	$O_{\cdot j}$ ou $n_{\cdot j}$
$e_{ij} := x_{i\cdot}x_{\cdot j}/n$	effectif théorique	$n_{i\cdot}n_{\cdot j}/n$
$f_{ij} = x_{ij}/n$	fréquence observée	
$f_{i\cdot} := \sum_{j=1}^{q} f_{ij}$	fréquence marginale-ligne	
$f_{\cdot j} := \sum_{i=1}^{p} f_{ij}$	fréquence marginale-colonne	
$f_{i\cdot}f_{\cdot j}$	fréquence théorique	
$f_{j/i} = f_{ij}/f_{i\cdot}$	fréquences conditionnelles associées aux profils-lignes	
$f_{i/j} = f_{ij}/f_{\cdot j}$	fréquences conditionnelles associées aux profils-colonnes	

Il est important de noter qu'a priori nous n'avons aucune information sur la dépendance entre les deux variables. Il se peu alors que ces deux dérnieres sont indépendantes et par consequent l'AFC est inutile. Donc on peu traduire ceci, du fait que sous l'hypothese d'indépendance, H_0, on a

$$f_{i\cdot}f_{\cdot j} = \mathbf{P}\left(V_1 = i, V_2 = j\right) = \mathbf{P}\left(V_1 = i\right)\mathbf{P}\left(V_1 = i\right).$$

En d'autres termes $f_{i\cdot}f_{\cdot j}$. Contrairement à l'hypothèse de dépendance, H_1, on a

$$f_{i\cdot}f_{\cdot j} = \mathbf{P}\left(V_1 = i, V_2 = j\right) \neq \mathbf{P}\left(V_1 = i\right)\mathbf{P}\left(V_1 = i\right) = f_{ij}.$$

Du tableau de contingences au tableau de probabilités

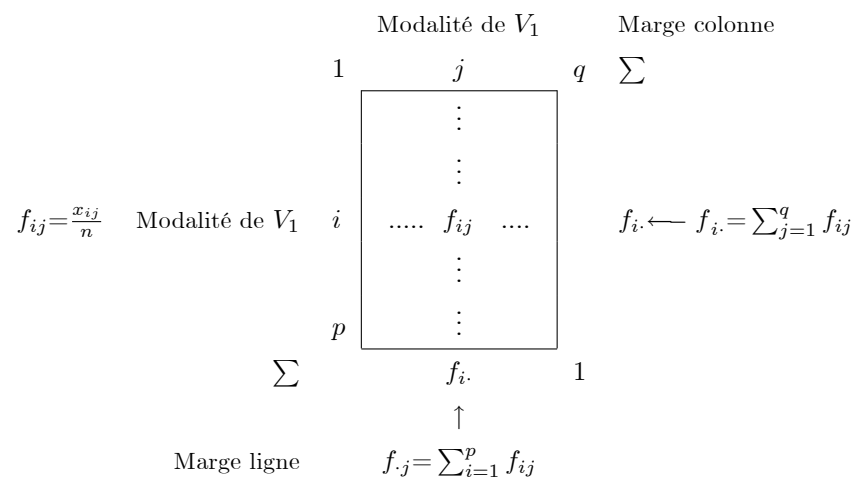

Fig.1

L'AFC vise à analyser ce tableau en apportant des réponses à des questions telles que:

- Y a-t-il des lignes du tableau (modalités de V_1) qui se "ressemblent", c'est-à-dire telles que les distributions des modalités de V_2 soient analogues?
- Y a-t-il des lignes du tableau (modalités de V_1) qui s'opposent, c'est-à-dire telles que les distributions des modalités de V_2 soient très différentes?
- Mêmes questions pour les colonnes du tableau.
- Y a-t-il des associations modalité de V_1 - modalité de V_2 qui s'attirent (effectif conjoint particulièrement élevé) ou qui se repoussent (effectif conjoint particulièrement faible)?

La méthode se fixe également comme but de construire des représentations graphiques mettant en évidence ces propriétés des données.

2 Indépendance des deux variables V_1 et V_2 : test du χ^2

Pour interpréter l'AFC, la première étape consiste à évaluer s'il existe une dépendance significative entre les lignes et les colonnes. Pour examiner l'association entre les modalités des lignes et celles des colonnes, une méthode rigoureuse consiste à utiliser la statistique du khi2

$$\chi^2 := n \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(\mathbf{f}_{ij} - \mathbf{f}_{i\cdot}\mathbf{f}_{\cdot j})^2}{\mathbf{f}_{i\cdot}\mathbf{f}_{\cdot j}}.$$

En d'autres termes

$$\chi^2 = n \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(\text{proba. conjointe} - \text{produit proba. marginales})^2}{\text{produit proba. marginales}}.$$

La statistique du khi2 observée est notée par

$$\chi^2_{obs} = n \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(f_{ij} - f_{i\cdot} f_{\cdot j})^2}{f_{i\cdot} f_{\cdot j}}.$$

En termes des fréquences observées conditionnelles, celle-ci peut être réecrite sous la forme

$$\chi^2_{obs} = n \sum_{i=1}^{p} \sum_{j=1}^{q} f_{i\cdot} \frac{(f_{ij}/f_{i\cdot} - f_{\cdot j})^2}{f_{\cdot j}} = n \sum_{i=1}^{p} \sum_{j=1}^{q} f_{i\cdot} \frac{\left(f_{j/i} - f_{\cdot j}\right)^2}{f_{\cdot j}}.$$

Pour cela il suffit d'utiliser entre autres, la fonction "`chisq.test`" du logiciel R. Voici les commandes à utiliser:

```
tab<-matrix(c(68,119,26,7,15,54,14,10,4,29,14,10),ncol=4,byrow=TRUE)
test=chisq.test(tab)
X-squared =34.114, df=6,p-value=6.394e-06
```

La "p-value" $< \alpha := 0.05$, alors on rejette l'hypothèse, nulle, d'association (ou d'indépendance) des deux variables V_1 et V_2. En conclusion, il y a une liaison entre la couleur des cheveux et la couleur des yeux. L'AFC donc a un sens.

3 Matrices des profils-lignes et profils-colonnes

On définit, respectivement, les matrices diagonales de profils-lignes et profils-colonnes par

$$D_r := \begin{pmatrix} f_{1\cdot} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_{p\cdot} \end{pmatrix} \text{ et } D_c := \begin{pmatrix} f_{\cdot 1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_{\cdot q} \end{pmatrix}.$$

Définition 3.1 *On appelle profils-lignes le tableau*

$$X_r := D_r^{-1} N = \begin{pmatrix} \frac{f_{11}}{f_{1\cdot}} & \cdots & \frac{f_{1q}}{f_{1\cdot}} \\ \vdots & \ddots & \vdots \\ \frac{f_{p1}}{f_{p\cdot}} & \cdots & \frac{f_{pq}}{f_{p\cdot}} \end{pmatrix}$$

et profils-colonnes le tableau

$$X_c := D_c^{-1} N^t = \begin{pmatrix} \frac{f_{11}}{f_{\cdot 1}} & \cdots & \frac{f_{p1}}{f_{\cdot 1}} \\ \vdots & \ddots & \vdots \\ \frac{f_{1q}}{f_{\cdot q}} & \cdots & \frac{f_{pq}}{f_{\cdot q}} \end{pmatrix}.$$

Les profils-lignes forment un nuage de p points dans $\mathbb{R}^q$. On affecte alors à chacun de ces p points un poids proportionnel à sa fréquence marginale (plus une modalité est rare, moins elle a de poids),

et on obtient la matrice de poids D_r . On peut alors définir le centre de gravité de ce nuage de points:

$$g_r = (f_{\cdot 1}, ..., f_{\cdot q})^t = X_r^t D_r \mathbf{1}_p = N^t \mathbf{1}_p,$$

où $\mathbf{1}_p = (1, ..., 1)^t$ *, le vecteur unitaire de* $p \times 1$*. On retrouve le profil marginal. De la même façon, les profils-colonnes forment un nuage de* q *points de* $\mathbb{R}^q$*, de poids donnés par la matrice* D_c*. Leur centre de gravité est*

$$g_c = (f_{1\cdot}, ..., f_{p\cdot})^t = X_c^t D_c \mathbf{1}_q,$$

où $\mathbf{1}_q := (1, \cdots, 1)^t$ *, le vecteur unitaire de* $q \times 1$*.*

Exemple 3.1 *Considérons la matrice des données de l'exemple ci-dessus*

$$N^* = \begin{pmatrix} 68 & 119 & 26 & 7 \\ 15 & 54 & 14 & 10 \\ 4 & 29 & 14 & 10 \end{pmatrix}$$

Nous avons ici $p = 3$*,* $q = 4$ *et*

$$n = 68 + 15 + 4 + 119 + 54 + 29 + 26 + 14 + 14 + 7 + 10 + 10 = 370.$$

Ainsi

$$N = \begin{pmatrix} 68/370 & 119/370 & 26/370 & 7/370 \\ 15/370 & 54/370 & 14/370 & 10/370 \\ 4/370 & 29/370 & 14/370 & 10/370 \end{pmatrix}.$$

Les fréquences marginales-lignes sont

$$f_{1\cdot} = \sum_{j=1}^{4} f_{1j} = 68/370 + 119/370 + 26/370 + 7/370 = 220/370$$

$$f_{2\cdot} = \sum_{j=1}^{4} f_{2j} = 15/370 + 54/370 + 14/370 + 10/370 = 93/370$$

$$f_{3\cdot} = \sum_{j=1}^{4} f_{3j} = 4/370 + 29/370 + 14/370 + 10/370 = 57/370.$$

Les fréquences marginales-colonnes sont

$$f_{\cdot 1} = \sum_{i=1}^{3} f_{i1} = 68/370 + 15/370 + 4/370 = 87/370$$
$$f_{\cdot 2} = \sum_{i=1}^{3} f_{2j} = 119/370 + 54/370 + 29/370 = 202/370$$
$$f_{\cdot 3} = \sum_{i=1}^{3} f_{3j} = 26/370 + 14/370 + 14/370 = \frac{27}{185}$$
$$f_{\cdot 4} = \sum_{i=1}^{3} f_{4j} = 7/370 + 10/370 + 10/370 = 27/370.$$

Les centres de gravité des profils-linges et profils-colonnes, respectivement, sont

$$g_r = (87/370, 202/370, 27/185, 27/370)^t,$$

et

$$g_c = (220/370, 93/370, 57/370)^t.$$

Les matrices diagonales de profils-lignes et profils-colonnes sont respectivement

$$D_r = \begin{pmatrix} 220/370 & 0 & 0 \\ 0 & 93/370 & 0 \\ 0 & 0 & 57/370 \end{pmatrix}$$

et

$$D_c = \begin{pmatrix} 87/370 & 0 & 0 & 0 \\ 0 & 202/370 & 0 & 0 \\ 0 & 0 & 27/185 & 0 \\ 0 & 0 & 0 & 27/370 \end{pmatrix}$$

Les matrices profils-lignes et profils-colonnes, respectivement, sont

$$X_r = D_r^{-1}N = \begin{pmatrix} \frac{37}{22} & 0 & 0 \\ 0 & \frac{370}{93} & 0 \\ 0 & 0 & \frac{370}{57} \end{pmatrix} \begin{pmatrix} 68/370 & 119/370 & 26/370 & 7/370 \\ 15/370 & 54/370 & 14/370 & 10/370 \\ 4/370 & 29/370 & 14/370 & 10/370 \end{pmatrix} = \begin{pmatrix} \frac{17}{55} & \frac{119}{220} & \frac{13}{110} & \frac{7}{220} \\ \frac{5}{31} & \frac{18}{31} & \frac{14}{93} & \frac{10}{93} \\ \frac{4}{57} & \frac{29}{57} & \frac{14}{57} & \frac{10}{57} \end{pmatrix}$$

et

$$X_c = D_c^{-1}N^t = \begin{pmatrix} \frac{370}{87} & 0 & 0 & 0 \\ 0 & \frac{185}{101} & 0 & 0 \\ 0 & 0 & \frac{185}{27} & 0 \\ 0 & 0 & 0 & \frac{370}{27} \end{pmatrix} \begin{pmatrix} \frac{34}{185} & \frac{3}{74} & \frac{2}{185} \\ \frac{119}{370} & \frac{27}{185} & \frac{29}{370} \\ \frac{13}{185} & \frac{7}{185} & \frac{7}{185} \\ \frac{7}{370} & \frac{1}{37} & \frac{1}{37} \end{pmatrix} = \begin{pmatrix} \frac{68}{87} & \frac{5}{29} & \frac{4}{87} \\ \frac{119}{202} & \frac{27}{101} & \frac{29}{202} \\ \frac{13}{27} & \frac{7}{27} & \frac{7}{27} \\ \frac{7}{27} & \frac{10}{27} & \frac{10}{27} \end{pmatrix}.$$

4 Métrique du chi-deux

L'objectif de l'AFC est donc d'étudier la dispersion des points autour du centre de gravité du nuage. Pour cela, on va devoir choisir une métrique appropriée afin de définir la distance entre deux points. Dans le cas de l'AFC, la distance utilisée est la distance du chi-deux, qui permet de mettre plus de poids sur les modalités de petits effectifs, et la métrique associée à cette distance est appelée métrique du khi-deux. L'écart entre les données observées et le modèle d'indépendance $f_{i\cdot}f_{\cdot j}$ est défini par

$$\chi^2_{obs} := \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\left(nf_{ij} - nf_{i\cdot}f_{\cdot j}\right)^2}{nf_{i\cdot}f_{\cdot j}} = n\Phi^2,$$

où Φ^2 s'appelle l'intensité de liaison (ou l'écart à l'indépendance) définissant l'écart entre probabilités théoriques et observées:

$$\Phi^2 = \frac{\chi^2_{obs}}{n} = \sum_{i=1}^{p}\sum_{j=1}^{q}\frac{\left(f_{ij} - f_{i\cdot}f_{\cdot j}\right)^2}{f_{i\cdot}f_{\cdot j}}.$$

La distance entre deux profils-lignes i et i' est

$$d^2_{\chi^2}(i,i') = \sum_{j=1}^{q}\frac{1}{f_{\cdot j}}\left(\frac{f_{ij}}{f_{i\cdot}} - \frac{f_{i'j}}{f_{i'\cdot}}\right)^2 = \|i - i'\|^2_{M_r}.$$

Celle-ci peut être vu comme la distance euclidienne, entre les deux profils-lignes i et i', pondérée. Plus précisément

$$d^2_{\chi^2}(i,i') = (i-i')^t M_r (i-i') := \|i - i'\|^2_{M_r},$$

où

$$M_r := D_c^{-1} = \begin{pmatrix} 1/f_{\cdot 1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/f_{\cdot q} \end{pmatrix}.$$

La distance entre un profil-ligne et le son centre de gravité g_r est définie par

$$d^2_{\chi^2}(i,g_r) = \sum_{j=1}^{q}\frac{1}{f_{\cdot j}}\left(\frac{f_{ij}}{f_{i\cdot}} - f_{\cdot j}\right)^2 = \|i - g_r\|^2_{M_r}.$$

La distance entre deux profils-colonnes j et j' est

$$d^2_{\chi^2}(j,j') = \sum_{i=1}^{p}\frac{1}{f_{i\cdot}}\left(\frac{f_{ij}}{f_{\cdot j}} - \frac{f_{ij'}}{f_{\cdot j'}}\right)^2 = \|j - j'\|^2_{M_c},$$

ou

$$M_c := D_r^{-1} = \begin{pmatrix} 1/f_{1\cdot} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/f_{p\cdot} \end{pmatrix}.$$

La distance entre profil-colonne et le centre gravité g_c est définie par

$$d^2_{\chi^2}(j, g_c) = \sum_{i=1}^{p} \frac{1}{f_{i\cdot}} \left(\frac{f_{ij}}{f_{\cdot j}} - f_{i\cdot} \right)^2 = \|j - g_c\|^2_{M_c}.$$

Avec cette distance, si on observe de grands écarts sur des modalités peu représentées, ceux-ci ont plus de poids dans le calcul de la distance. Et inversement, on donne moins de poids à des écarts importants qui pourraient être dus au fait que l'on a seulement observé plus de points sur cette modalité.

Exemple 4.1 *La matrice des profils-lignes est*

$$X_r = \begin{pmatrix} 68/220 & 119/220 & 26/220 & 7/220 \\ 15/93 & 54/93 & 14/93 & 10/93 \\ 4/57 & 29/57 & 14/57 & 10/57 \end{pmatrix},$$

La distance de chi-deux entre la première et la deuxième lignes est

$$\begin{aligned} d^2_{\chi^2}(1,2) &= \frac{370}{87} \left(\frac{68}{220} - \frac{15}{93} \right)^2 + \frac{370}{202} \left(\frac{119}{220} - \frac{54}{93} \right)^2 \\ &\quad + \frac{370}{27} \left(\frac{26}{220} - \frac{14}{93} \right)^2 + \frac{370}{27} \left(\frac{7}{220} - \frac{10}{93} \right)^2 \\ &= 0.188\,69. \end{aligned}$$

4.1 Inertie totale

Les inerties totales des nuages de points profils-lignes et profils-colonnes par rapport aux centres de gravité correspondants sont définies respectivement par

$$\text{Inertie}(X_r/g_r) := \sum_{i=1}^{p} f_{i\cdot} d^2_{\chi^2}(i, g_r),$$

et

$$\text{Inertie}(X_c/g_c) := \sum_{j=1}^{g} f_{\cdot j} d^2_{\chi^2}(j, g_c).$$

Observons que

$$\begin{aligned} \text{Inertie}(X_r/g_r) &= \sum_{i=1}^{p} \sum_{j=1}^{q} f_{i\cdot} \left(\frac{1}{f_{\cdot j}} \left(\frac{f_{ij}}{f_{i\cdot}} - f_{\cdot j} \right)^2 \right) = \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{f_{i\cdot}}{f_{\cdot j}} \left(\frac{f_{ij}}{f_{i\cdot}} - f_{\cdot j} \right)^2 \\ &= \sum_{i=1}^{p} \sum_{j=1}^{q} \frac{(f_{ij} - f_{i\cdot} f_{\cdot j})^2}{f_{i\cdot} f_{\cdot j}} = \Phi^2 = \frac{\chi^2}{n}. \end{aligned}$$

En d'autres termes, étudier l'inertie de X_r revient à étudier l'écart à l'indépendance Φ^2. On montre que

$$\text{Inertie}(X_r/g_r) = \text{Inertie}(X_c/g_c) = \Phi^2 = \frac{\chi^2}{n}.$$

Conclusion: Plus les données s'écartent de l'indépendance et plus les profils s'écartent de l'origine.

Exemple 4.2 *Pour notre exemple on a*

$$Inertie\,(X_r/g_r) = \sum_{i=1}^{3} f_{i\cdot} d^2_{\chi^2}\,(i, g_r) = \frac{\chi^2}{3} = \frac{34.114}{3} = 11.371.$$

$$Inertie\,(X_c/g_c) = 11.371.$$

5 ACP des deux nuages de profils

Une fois que l'on a définit la matrice des données X_r (respectivement X_c) la matrice de poids D_r (resp. D_c) et la métrique $M_r = D_c^{-1}$ $\left(\text{resp. } M_c = D_r^{-1}\right)$ on dispose de tous les éléments pour faire une ACP. On définit le nuage profils-lignes centré par

$$Y_r := X_r - \mathbf{1}_p g_r^t,$$

dont ces éléments sont

$$y_{i,j} = f_{ij}/f_{i\cdot} - f_{\cdot j},\ i = 1, ..., p \text{ et } j = 1, ..., q.$$

On désigne $\mathbf{y}_i = (y_{i,1}, ..., y_{i,q})^t$, $i = 1, ...p$, les vecteurs lignes de la matrice Y_r. On note par E l'axe principal de l'ACP et u le vecteur directeur associé de norme 1 par rapport à métrique M_r, c'est à dire $\|u\|^2_{M_r} = \langle u, u\rangle_{M_r} = u^t M_r u = 1$. Nous définissons l'inertie du nuage Y_r par rapport à $E^\perp$ par

$$\text{Inertie}\left(Y_r/E^\perp\right) = \sum_{i=1}^{p} f_{i\cdot} d^2_{\varkappa^2}\left(\mathbf{y}_i, \mathbf{Proj}_{E^\perp,i}\right).$$

Nous avons

$$\begin{aligned} d^2_{\varkappa^2}\left(\mathbf{y}_i, p_{E^\perp,i}\right) &= \left\|\mathbf{y}_i - \mathbf{Proj}_{E^\perp,i}\right\|^2_{M_r} \\ &= \left\|\frac{\langle \mathbf{y}_i, u\rangle_{M_r} u}{\|u\|^2_{M_r}}\right\|^2_{M_r} = \frac{\langle \mathbf{y}_i, u\rangle^2_{M_r} \|u\|^2_{M_r}}{\|u\|^4_{M_r}} = \langle \mathbf{y}_i, u\rangle^2_{M_r} \\ &= \left(\mathbf{y}_i^t M_r u\right)^2 = \left(\mathbf{y}_i^t M_r u\right)\left(\mathbf{y}_i^t M_r u\right)^t = \left(\mathbf{y}_i^t M_r u\right)\left(u^t M_r \mathbf{y}_i\right) \\ &= \left(u^t M_r \mathbf{y}_i\right)\left(\mathbf{y}_i^t M_r u\right) = u^t M_r \mathbf{y}_i \mathbf{y}_i^t M_r u. \end{aligned}$$

Par conséquent

$$\text{Inertie}\left(Y_r/E^\perp\right) = u^t M_r \left[\sum_{i=1}^{p} f_{i\cdot} \mathbf{y}_i \mathbf{y}_i^t\right] M_r u.$$

On note que

$$\sum_{i=1}^{p} f_{i\cdot} \mathbf{y}_i \mathbf{y}_i^t = Y_r^t D_r Y_r =: \mathbf{V}_r,$$

comme étant la matrice de variance-covariance associée à la matrice Y_r affectée aux poids D_r. Ainsi

$$\text{Inertie}\left(Y_r/E^{\perp}\right) = u^t M_r \mathbf{V}_r M_r u.$$

Nous allons maintenant chercher le vecteur u maximisant I'inertie$\left(Y_r/E^{\perp}\right)$ sous la contrainte $u^t M_r u = 1$. Ce que revient, en utilisant le multiplicateur de Lagrange, à maximiser la fonction

$$u \to \eta\left(u\right) := u^t M_r \mathbf{V}_r M_r u - \lambda\left(u^t M_r u - 1\right).$$

Il est clair que la dérivée de cette fonction est

$$\eta'\left(u\right) = 2M_r \mathbf{V}_r M_r u - 2\lambda M_r u.$$

En résolvant l'équation $\eta'\left(u\right) = 0$, on trouve $M_r \mathbf{V}_r M_r u = \lambda M_r u$. Rappelons que la matrice M_r est inversible, alors la dernière équation se réduit à

$$\mathbf{V}_r M_r u = \lambda u.$$

Comme nous l'avons fait au premier chapitre, nous allons appliquer l'ACP à la matrice $\mathbf{V}_r M_r$. En d'autres termes nous cherchons les valeurs propres et les vecteurs propres associés à $\mathbf{V}_r M_r$, définissants les axes principaux.

Remarque 5.1 *La matrice $\mathbf{V}_r M_r$ est M_r-symétrique, c'est à dire $M_r \mathbf{V}_r M_r$ est symétrique. En effet, comme $\mathbf{V}_r$ est M_r sont symétriques, il est évident que que $M_r \mathbf{V}_r M$ l'est aussi.*

Le théorème suivant est utile pour la suite.

Théoreme 5.1 *Le centre de gravité du nuage des profils-lignes, g_r, est M_r-orthogonal au nuage des profils-lignes centré Y_r (de même, g_c est M_c-orthogonal au nuage des profils-colonnes centré Y_c).*

Preuve 5.1 *Tout d'abord observons que*

$$M_r g_r = D_c^{-1} g_r = \begin{pmatrix} 1/f_{\cdot 1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/f_{\cdot q} \end{pmatrix} \begin{pmatrix} f_{\cdot 1} \\ \vdots \\ f_{\cdot q} \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \mathbf{1}_q$$

et

$$\langle g_r, g_r \rangle_{M_r} = \|g_r\|^2_{M_r} = g_r^t M_r g_r = g_r^t \mathbf{1}_q = f_{\cdot 1} + ... + f_{\cdot q} = 1.$$

Soit $r_i := \left(f_{i1}/f_{i\cdot}, ..., f_{iq}/f_{i\cdot}\right)^t$ la i-ème ligne de X_r. Alors

$$\begin{aligned} \langle r_i - g_r, g_r \rangle_{M_r} &= \langle r_i, g_r \rangle_{M_r} - \langle g_r, g_r \rangle_{M_r} \\ &= \left(f_{i1}/f_{i\cdot}, ..., f_{iq}/f_{i\cdot}\right) M_r \left(f_{\cdot 1}, ..., f_{\cdot q}\right)^t - \langle g_r, g_r \rangle_{M_r} \\ &= r_i^t M_r g_r - 1 = r_i^t \mathbf{1}_q - 1 \\ &= f_{i1}/f_{i\cdot} + ... + f_{iq}/f_{i\cdot} - 1 = 1 - 1 = 0. \end{aligned}$$

Corollaire 5.1 *Le centre de gravité g_r est un vecteur propre de V_rM_r associé à la valeur propre $\lambda = 0$.*

Preuve 5.2 *On va démontrer que $V_rM_rg_r = 0_{\mathbb{R}^q} = 0g_r$. Nous avons déjà montrer que $M_rg_r = \mathbf{1}_q$, donc*

$$\begin{aligned} V_rM_rg_r &= V_r\mathbf{1}_q = \left(X_r^tD_rX_r - g_rg_r^t\right)\mathbf{1}_q \\ &= X_r^tD_rX_r\mathbf{1}_q - g_rg_r^t\mathbf{1}_q. \end{aligned}$$

Observons que

$$g_r^t\mathbf{1}_q = (f_{\cdot 1}, .., f_{\cdot q})\,\mathbf{1}_q = f_{\cdot 1} + .. + f_{\cdot q} = 1.$$

Donc

$$g_rg_r^t\mathbf{1}_q = g_r.$$

D'autre part, on a

$$\begin{aligned} X_r\mathbf{1}_q &= \begin{pmatrix} \frac{f_{11}}{f_{1\cdot}} + ... + \frac{f_{1q}}{f_{1\cdot}} \\ \vdots \\ \frac{f_{p1}}{f_{p\cdot}} + ... + \frac{f_{pq}}{f_{p\cdot}} \end{pmatrix} = \begin{pmatrix} \frac{1}{f_{1\cdot}}\left(f_{11} + ... + f_{1q}\right) \\ \vdots \\ \frac{1}{f_{p\cdot}}\left(f_{p1} + ... + f_{pq}\right) \end{pmatrix} \\ &= \begin{pmatrix} \frac{f_{1\cdot}}{f_{1\cdot}} \\ \vdots \\ \frac{f_{p\cdot}}{f_{p\cdot}} \end{pmatrix} = \mathbf{1}_p. \end{aligned}$$

Il est claire que

$$D_r\mathbf{1}_p = \begin{pmatrix} f_{1\cdot} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & f_{p\cdot} \end{pmatrix}\mathbf{1}_p = \begin{pmatrix} f_{1\cdot} \\ \vdots \\ f_{p\cdot} \end{pmatrix} = g_c.$$

Nous avons donc

$$X_r^tD_rX_r\mathbf{1}_q = X_r^tg_c.$$

On va montrer que ce dernier égal à g_r. En effet,

$$\begin{aligned} X_r^tg_c &= \begin{pmatrix} \frac{f_{11}}{f_{1\cdot}} & \cdots & \frac{f_{p1}}{f_{p\cdot}} \\ \vdots & \ddots & \vdots \\ \frac{f_{1q}}{f_{1\cdot}} & \cdots & \frac{f_{pq}}{f_{p\cdot}} \end{pmatrix}\begin{pmatrix} f_{1\cdot} \\ \vdots \\ f_{p\cdot} \end{pmatrix} \\ &= \begin{pmatrix} f_{11} + ... + f_{p1} \\ \vdots \\ f_{1q} + ... + f_{pq} \end{pmatrix} = \begin{pmatrix} f_{\cdot 1} \\ \vdots \\ f_{\cdot q} \end{pmatrix} = g_r. \end{aligned}$$

Donc on a montrer que $X_r^t D_r X_r \mathbf{1}_q = g_r$, *ainsi*

$$\begin{aligned} V_r M_r g_r &= X_r^t D_r X_r \mathbf{1}_q - g_r g_r^t \mathbf{1}_q \\ &= g_r - g_r = 0_{\mathbb{R}^q} = 0 g_r. \end{aligned}$$

Remarque 5.2 *On rappel que le déterminant d'une matrice carré égal au produit des valeurs propres. Comme* $\lambda = 0$ *est une valeur propre de* $V_r M_r$ *alors son déterminant est nul. Ce qui implique que cette matrice n'est pas inversible, et que son rang égal au nombre maximum des valeurs propres non nuls. En d'autres termes*

$$rang\,(V_r M_r) \leq q - 1.$$

Avec le même raisonnement, on en conclus que

$$rang\,(V_c M_c) \leq p - 1.$$

Théoreme 5.2 *La matrice* $V_r M_r$ *a les mêmes valeurs propres que* $X_r^t D_r X_r M_r = N^t D_r^{-1} N D_c^{-1}$ *sauf* g_r *qui a une valeur propre* $\lambda = 1$.

Preuve 5.3 *Soit* v *un vecteur propre non nul de, différent de* g_r, *de matrice* $V_r M_r$. *Alors* $V_r M_r g_r = \lambda g_r$ *implique que*

$$\left(X_r^t D_r X_r - g_r g_r^t\right) M_r v = \lambda v,$$

c'est à dire

$$X_r^t D_r X_r M_r v - g_r g_r^t M_r v = \lambda v.$$

Observons que

$$X_r^t D_r X_r M_r v - g_r \left\langle g_r, v \right\rangle_{M_r} = \lambda v.$$

D'un autre coté g_r *est* v *sont deux vecteurs propres,* M_r-*orthogonaux, de* $V_r M_r$, *donc*

$$X_r^t D_r X_r M_r v = \lambda v,$$

ce qui implique que v *est un vecteur propre de* $X_r^t D_r X_r M_r$ *du même paramètre* λ.

Corollaire 5.2 *Il n'est pas nécessaire de centrer le nuage de points des profils-lignes avant de réaliser l'ACP, et on peut travailler directement avec la matrice*

$$N^t D_r^{-1} N D_c^{-1} = X_r^t X_c^t. \tag{5.1}$$

Dans le cas des profils-colonnes, on s'intéressera à la matrice

$$N D_c^{-1} N^t D_r^{-1} = X_c^t X_r^t. \tag{5.2}$$

5.0.1 Lien entre les deux ACP

Il existe un lien entre les vecteurs et valeurs propres des deux matrices $X_r^t X_c^t$ et $X_c^t X_r^t$.

Théoreme 5.3 *Si u est vecteur propre, de norme 1 pour la métrique M_r, de $X_r^t X_c^t$, pour la valeur propre $\lambda \neq 0$, alors*

$$\widetilde{u} = \frac{1}{\sqrt{\lambda}} N D_c^{-1} u$$

est un vecteur propre, de norme 1 pour la métrique M_c, pour $X_c^t X_r^t$, pour la même valeur propre λ. Inversement, si $\widetilde{u}$ est vecteur propre, de norme 1 pour la métrique M_c, de $X_c^t X_r^t$, pour la valeur propre $\lambda \neq 0$, alors

$$u = \frac{1}{\sqrt{\lambda}} N^t D_r^{-1} \widetilde{u}$$

est un vecteur propre, de norme 1 pour la métrique M_r, pour $X_r^t X_c^t$, pour la même valeur propre λ.

Corollaire 5.3 *Nous avons*

$$\tau := rang V_r M_r = rang V_c M_c,$$

de plus

$$0 < \tau \leq \min(p-1, q-1).$$

Preuve 5.4 *Comme $\lambda = \widetilde{\lambda} \neq 0$ alors $\tau :=$rang$V_r M_r =$rang$V_c M_c$. D'un autre coté nous avons*

$$\tau = rang(V_r M_r) \leq q - 1.$$

De même, on en conclus que

$$\tau = rang(V_c M_c) \leq p - 1.$$

Ce qui implique que

$$0 < \tau \leq \min(p-1, q-1).$$

Théoreme 5.4 *En notant λ_k la k-ème valeur propre non nulle, on alors :*

$$\Phi^2 = \sum_{k=1}^{\tau} \lambda_k.$$

Chaque direction ou axe principal explique donc une partie de l'écart à l'indépendance entre les deux variables.

5.1 Facteurs principaux et composantes principales

Définition 5.1 *On appel* ***facteurs principaux*** *des profils-lignes (resp. profils-colonnes) sont les vecteurs* $w_i := M_r u_i$ *(resp.* $\widetilde{w}_i := M_c \widetilde{u}_i$*).*

Proposition 5.1 *Les facteurs principaux des profils-lignes (resp. profils-colonnes) sont* $M_r^{-1}-$*orthonormés (resp.* $M_c^{-1}-$*orthonormés).*

Preuve 5.5 *Soient* $u_i, i = 1, ..., p$ *les axe principaux des profils-lignes. On sait que les* u_i *sont* M_r-*orthonormés, donc*

$$w_i^t M_r^{-1} w_j = u_i^t M_r M_r^{-1} M_r u_j = u_i^t M_r u_i = 1 \text{ si } i = j \text{ et } 0 \text{ si } i \neq j.$$

Définition 5.2 ***Proposition 5.2*** *Les facteurs principaux des profils-lignes (rep. profils colonnes) sont les vecteurs propres* $M_r^{-1}-$*orthonormés (resp.* $M_c^{-1}-$*orthonormés) de la matrice de données* $M_r V_r$ *(resp.* $M_c V_c$*).*

Preuve 5.6 *Soit* u *un vecteur propre de* $V_r M_r$ *associé à la valeur propre* λ*, alors* $V_r M_r u = \lambda u$*. En multipliant les deux membres de cette équation par* M_r*, on obtient* $M_r (V_r M_r u) = \lambda M_r u$*. Donc* $M_r V_r (M_r u) = \lambda (M_r u)$*, où* $M_r u =: w$ *est le facteur principal associé à* u*. De plus* $w^t M_r^{-1} w = u^t M_r M_r^{-1} M_r u = u^t M_r u = 1$*, car* u *est* M_r-*normé.*

Définition 5.3 *Les composantes principales des profils-lignes (resp. profils-colonnes sont les* M_r*-coordonnées (*M_c*-coordonnées) des vecteurs colonnes de la matrice de données* $Y_r := X_r - \mathbf{1}_p g_r^t$ *(resp.*$Y_c := X_c - \mathbf{1}_p g_r^t$*) c'est à dire*

$$c_k := Y_r w_k, \ k = 1, ..., p \text{ (resp. } \widetilde{c}_k := Y_c \widetilde{w}_k, \ k = 1, ..., q),$$

où $w_k = M_r u_k$ *(resp.* $\widetilde{w}_k = M_c \widetilde{u}_k$*) sont les facteurs principaux*

Proposition 5.3 *Les composantes principales des profils-lignes (resp. profils-colonnes) sont les* M_r*-coordonnées (*M_c*-coordonnées) des vecteurs colonnes de la matrices des données* X_r *(resp.* X_c*) c'est à dire*

$$c_k := X_r w_k, \ k = 1, ..., p \text{ (resp. } \widetilde{c}_k := X_c \widetilde{w}_k, \ k = 1, ..., q).$$

Preuve 5.7 *Triviale, car le centre de gravité est axe principal associé à une valeur propre nulle.*

Proposition 5.4 *Nous avons*

$$\mathbf{E}[c_k] = \mathbf{E}[\widetilde{c}_k] = 0, \ k = 1, ..., \tau,$$

$$\mathbf{Var}\left[c_k\right] = \mathbf{Var}\left[\widetilde{c}_k\right] = \lambda_k,\ k = 1, ..., \tau,$$

$$\mathbf{Cov}\left[c_k, c_\ell\right] = 0,\ \text{pour } k \neq \ell \text{ et } \mathbf{Cov}\left[\widetilde{c}_k, \widetilde{c}_\ell\right] = 0,\ \text{pour } k \neq \ell.$$

Preuve 5.8 *Voir le premier chapitre concernant l'analyse en composantes principales.*

Proposition 5.5 *Les facteurs principaux de l'ACP des profils-colonnes, associés aux valeurs propres non nulles, sont, à une constante près, les composantes principales de l'ACP des profils-lignes, et vice-versa.*

Preuve 5.9 *En effet, soit u un axe principal, des profils-lignes, associé à la valeur propre $\lambda \neq 0$ et $\widetilde{u}$ l'axe principal correspondant des profils-colonnes. Observons maintenant que, la composante principale associée à u est*

$$c = X_r M_r u = D_r^{-1}\left(N D_c^{-1} u\right) = D_r^{-1}\left(\sqrt{\lambda}\widetilde{u}\right) = \sqrt{\lambda}\widetilde{w}.$$

Inversement

$$\widetilde{c} = X_c M_c \widetilde{u} = D_c^{-1}\left(N^t D_r^{-1}\widetilde{u}\right) = D_c^{-1}\left(\sqrt{\lambda}u\right) = \sqrt{\lambda}w.$$

Le résultat précédent conduit aux relations fondamentales de l'AFC reliant les composantes principales entre elles, dites les relations *quasi-barycentriques*:

Proposition 5.6 *Soit $\lambda_1 > \lambda_2 > ... > \lambda_\tau \neq 0$. Alors, pour tout $k \leq \tau$, on a*

$$c_k = \frac{1}{\sqrt{\lambda_k}} X_r \widetilde{c}_k \text{ et } \widetilde{c}_k = \frac{1}{\sqrt{\lambda_k}} X_r c_k.$$

Plus précisément

$$c_k(i) = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^{q} \frac{f_{ij}}{f_{i\cdot}} \widetilde{c}_k(j),\ i = 1, ..., p$$

et

$$\widetilde{c}_k(j) = \frac{1}{\sqrt{\lambda_k}} \sum_{j=1}^{p} \frac{f_{ij}}{f_{\cdot j}} c_k(i),\ j = 1, ..., q.$$

Théoreme 5.5 *Pour tout $k = 1, ..., \tau$, on a*

$$0 < \lambda_k \leq 1.$$

Preuve 5.10 *Soient $\mathbf{A} = (a_{ij}) \in \mathcal{M}(p \times q)$, $\mathbf{x} = (x_1, ..., x_q)^t \in \mathbb{R}^q$ et $\mathbf{y} = (y_1, ..., y_p)^t \in \mathbb{R}^p$. On munit $\mathbb{R}^q$ et $\mathbb{R}^p$ des normes L_1 définis par*

$$\|\mathbf{x}\| = \|\mathbf{x}\|_\infty = \max_{i=1,...,q} |x_i|,\ \ \|\mathbf{y}\| = \|\mathbf{y}\|_\infty = \max_{i=1,...,p} |y_i|.$$

On munit aussi l'espace des matricies $\mathcal{M}(p \times q)$ de la norme

$$\|\mathbf{A}\| = \|\mathbf{A}\|_\infty = \sup_{\|\mathbf{x}\| \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} = \sup_{\|\mathbf{x}\| \leq 1} \|\mathbf{A}\mathbf{x}\| = \sup_{\|\mathbf{x}\| = 1} \|\mathbf{A}\mathbf{x}\|.$$

D'après le Lemme 1 (voir l'Appendix en bas du document), on a

$$\|\mathbf{A}\| = \max_{i=1,\ldots,p} \sum_{j=1}^{q} |a_{ij}|.$$

Il est clair que

$$\frac{\|\mathbf{A}\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|\mathbf{A}\| \Longrightarrow \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \|\mathbf{x}\|,$$

ainsi

$$\|\mathbf{A}\mathbf{x}\| \leq \max_{i=1,\ldots,p} \sum_{j=1}^{q} |a_{ij}| \|\mathbf{x}\|.$$

***Application**: $\mathbf{A} = X_r \in \mathcal{M}(p \times q)$. Comme c_k est un vecteur de $\mathbb{R}^p$ alors $\mathbf{x} = X_c c_k$ étant un vecteur de $\mathbb{R}^q$. Donc on a*

$$\begin{aligned} \lambda \|c_k\| &= \|(X_r)(X_c c_k)\| = \|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{A}\| \|\mathbf{x}\| \\ &\leq \max_{i=1,\ldots,p} \sum_{j=1}^{q} |a_{ij}| \|\mathbf{x}\|. \end{aligned}$$

C'est à dire

$$\lambda \|c_k\| \leq \max_{i=1,\ldots,p} \sum_{j=1}^{q} |a_{ij}| \|X_c c_k\|.$$

Pour $\mathbf{A} = X_c = (b_{ij}) \in \mathcal{M}(q \times p)$, on a

$$\|X_c c_k\| \leq \|X_c\| \|c_k\| = \left\{ \max_{i=1,\ldots,q} \sum_{j=1}^{p} |b_{ij}| \right\} \|c_k\|.$$

Ce qui implique

$$\lambda \|c_k\| = \|X_r X_c c_k\| \leq \left\{ \max_{i=1,\ldots,p} \sum_{j=1}^{q} |a_{ij}| \right\} \left\{ \max_{i=1,\ldots,q} \sum_{j=1}^{p} |b_{ij}| \right\} \|c_k\|.$$

En simplifiant par $\|c_k\|$ (qui est évidement non nulle), on obtient

$$0 < \lambda \leq \left\{ \max_{i=1,\ldots,p} \sum_{j=1}^{q} |a_{ij}| \right\} \left\{ \max_{i=1,\ldots,q} \sum_{j=1}^{p} |b_{ij}| \right\}.$$

Ce qui implique

$$0 < \lambda \leq \left\{ \max_{i=1,\ldots,p} \sum_{j=1}^{q} \frac{f_{ij}}{f_{i.}} \right\} \left\{ \max_{i=1,\ldots,q} \sum_{j=1}^{p} \frac{f_{ji}}{f_{.j.}} \right\}.$$

On note que $f_{ji} \equiv f_{ji}$, *donc*

$$0 < \lambda \leq \left\{ \max_{i=1,...,p} \sum_{j=1}^{q} \frac{f_{ij}}{f_{i.}} \right\} \left\{ \max_{j=1,...,q} \sum_{i=1}^{p} \frac{f_{ij}}{f_{.j.}} \right\}.$$

Rappelons que $\sum_{j=1}^{q} f_{ij} = f_{i.}$ *et* $\sum_{i=1}^{p} f_{ij} = f_{.j}$, *ainsi*

$$\sum_{j=1}^{q} \frac{f_{ij}}{f_{i.}} = \sum_{i=1}^{p} \frac{f_{ji}}{f_{.j.}} = 1,$$

et parconséquent $0 < \lambda \leq 1$.

6 Formules de reconstitution

Comme en ACP on dispose de formules dites de reconstitution permettant de récupérer le tableau N à partir des composantes principales c et $\widetilde{c}_k$.

Théoreme 6.1 *Pour tout* $i \leq p$ *et tout* $j \leq q$, *on a :*

$$\frac{f_{ij}}{f_{i\cdot} f_{\cdot j}} - 1 = \sum_{k=1}^{\tau} \frac{1}{\sqrt{\lambda_k}} c_k(i)\, \widetilde{c}_k(j).$$

Les composants principales et les valeurs propres expliquent donc en quoi les f_{ij} s'écartent des $f_{i\cdot} f_{\cdot j}$.

Preuve 6.1 *Nous avons*

$$Y_r := X_r - \mathbf{1}_p g_r^t,$$

dont ces éléments sont

$$y_{i,j} = f_{ij}/f_{i\cdot} - f_{\cdot j},\ i = 1, ..., p \text{ et } j = 1, ..., q.$$

Rappelons que $\{u_1, ..., u_p\}$ *est une base* M_r-*orthonormée de* $\mathbb{R}^p$ *et les* c_k, $k = 1, ..., q$ *sont coordonnées de la i-ième ligne de* Y_r *dans la base* $\{u_1, ..., u_p\}$. *En d'autres termes*

$$Y_r = \sum_{k=1}^{p} c_k u_k.$$

Comme $c_k = 0$, $k = \tau + 1, ..., p$, *alors*

$$Y_r = \sum_{k=1}^{\tau} c_k u_k.$$

Rappelons aussi que $u_k = \frac{1}{\sqrt{\lambda_k}} N^t D_r^{-1} \widetilde{u}_k$,*et* $\{\widetilde{u}_1, ..., \widetilde{u}_q\}$ *est une base* M_c-*orthonormée de* $\mathbb{R}^q$. *Alors*

$$Y_r = \sum_{k=1}^{\tau} \frac{1}{\sqrt{\lambda_k}} c_k N^t D_r^{-1} \widetilde{u}_k = \sum_{k=1}^{\tau} \frac{1}{\sqrt{\lambda_k}} c_k D_c \widetilde{c}_k.$$

En écrivant cette relation coordonnées par coordonnées, on obtient pour tout $j = 1, ..., q$

$$f_{ij}/f_{i\cdot} - f_{\cdot j} = \frac{1}{\sqrt{\lambda_k}} \sum_{k=1}^{\tau} c_k(i)\, f_{\cdot j} \widetilde{c}_k.$$

En devisant les deux membres de cette équation par $f_{\cdot j}$ *on obtient*

$$\frac{f_{ij}}{f_{i\cdot} f_{\cdot j}} - 1 = \sum_{k=1}^{\tau} \frac{1}{\sqrt{\lambda_k}} c_k(i)\, \widetilde{c}_k(j).$$

6.1 Contribution des profils

Rappelons que deux types de coefficients apportent de l'information intéressante pour l'interprétation des plans factoriels.

- Contribution relative: elle exprime la part prise par une modalité e de la variable dans l'inertie "expliquée " par un facteur.
- Cosinus carré: il mesure la qualité de représentation de la modalité sur le facteur.

6.1.1 Contribution relative d'une modalité à l'inertie de l'axe k :

- Contribution relative du profil-ligne i de de la matrice Y_r au $k-$ième axe (de vecteur u_k)

$$\frac{c_k^2(i)}{\lambda_k} f_{i\cdot}$$

- Contribution relative du profil-colonne j de de la matrice Y_c au $k-$ième axe (de vecteur $\widetilde{u}_k$)

$$\frac{c_k^2(j)}{\lambda_k} f_{\cdot j}$$

6.1.2 Qualité de représentation sur l'axe k :

- Qualité de la représentation du profil-ligne i de la matrice Y_r au $k-$ième axe (de vecteur u_k)

$$\frac{c_k^2(i)}{\sum_{k=1}^{\tau} c_k^2(i)}$$

- Qualité de la représentation du profil-colonnes j de la matrice Y_c au $k-$ième axe (de vecteur $\widetilde{u}_k$)

$$\frac{\widetilde{c}_k^2(i)}{\sum_{k=1}^{\tau} \widetilde{c}_k^2(i)}$$

Exemple 6.1 *Faisons l'ACP pour notre exemple. On rappel que les matrices de profils-lignes et profils-colonnes sont respectivement*

$$X_r = \begin{pmatrix} 68/220 & 119/220 & 26/220 & 7/220 \\ 15/93 & 54/93 & 14/93 & 10/93 \\ 4/57 & 29/57 & 14/57 & 10/57 \end{pmatrix} \text{ et } X_c = \begin{pmatrix} \frac{68}{87} & \frac{5}{29} & \frac{4}{87} \\ \frac{119}{202} & \frac{27}{101} & \frac{29}{202} \\ \frac{13}{27} & \frac{7}{27} & \frac{7}{27} \\ \frac{7}{27} & \frac{10}{27} & \frac{10}{27} \end{pmatrix}.$$

Les centres de gravité des profils-linges et profils-colonnes, respectivement, sont

$$g_r = \begin{pmatrix} \frac{87}{370} \\ \frac{101}{185} \\ \frac{27}{185} \\ \frac{27}{370} \end{pmatrix} = \begin{pmatrix} 0.235\,14 \\ 0.545\,95 \\ 0.145\,95 \\ 0.07.297 \end{pmatrix} \text{ et } g_c = \begin{pmatrix} \frac{22}{37} \\ \frac{93}{370} \\ \frac{57}{370} \end{pmatrix} = \begin{pmatrix} 0.594\,59 \\ 0.251\,35 \\ 0.154\,05 \end{pmatrix}$$

Les transposées X_r et X_c sont

$$X_r^t = \begin{pmatrix} \frac{17}{55} & \frac{5}{31} & \frac{4}{57} \\ \frac{119}{220} & \frac{18}{31} & \frac{29}{57} \\ \frac{13}{110} & \frac{14}{93} & \frac{14}{57} \\ \frac{7}{220} & \frac{10}{93} & \frac{10}{57} \end{pmatrix} \text{ et } X_c^t = \begin{pmatrix} \frac{68}{87} & \frac{119}{202} & \frac{13}{27} & \frac{7}{27} \\ \frac{5}{29} & \frac{27}{101} & \frac{7}{27} & \frac{10}{27} \\ \frac{4}{87} & \frac{29}{202} & \frac{7}{27} & \frac{10}{27} \end{pmatrix}.$$

Les ACP sur les profils-lignes et les profils-colonnes sont, respectivement basées sur les deux matrices suivantes

$$A_r := X_r^t X_c^t \text{ et } A_c := X_c^t X_r^t.$$

On commence à chercher les valeurs propres des deux matrices et faisons la comparaisons entre eux. Nous avons

$$A_r = X_r^t X_c^t = \begin{pmatrix} \frac{17}{55} & \frac{5}{31} & \frac{4}{57} \\ \frac{119}{220} & \frac{18}{31} & \frac{29}{57} \\ \frac{13}{110} & \frac{14}{93} & \frac{14}{57} \\ \frac{7}{220} & \frac{10}{93} & \frac{10}{57} \end{pmatrix} \begin{pmatrix} \frac{68}{87} & \frac{119}{202} & \frac{13}{27} & \frac{7}{27} \\ \frac{5}{29} & \frac{27}{101} & \frac{7}{27} & \frac{10}{27} \\ \frac{4}{87} & \frac{29}{202} & \frac{7}{27} & \frac{10}{27} \end{pmatrix}$$

Le calcul donne

$$A_r = \begin{pmatrix} \frac{2305\,057}{8455\,095} & \frac{4618\,871}{19\,631\,370} & \frac{547\,972}{2623\,995} & \frac{435\,223}{2623\,995} \\ \frac{4618\,871}{8455\,095} & \frac{42\,946\,987}{78\,525\,480} & \frac{5698\,049}{10\,495\,980} & \frac{5706\,911}{10\,495\,980} \\ \frac{1095\,944}{8455\,095} & \frac{5698\,049}{39\,262\,740} & \frac{837\,623}{5247\,990} & \frac{930\,797}{5247\,990} \\ \frac{435\,223}{8455\,095} & \frac{5706\,911}{78\,525\,480} & \frac{930\,797}{10\,495\,980} & \frac{1186\,583}{10\,495\,980} \end{pmatrix}.$$

Les valeurs propres de A_r sont

$$\lambda_1^* = 0.089\,57,\ \lambda_2^* = 2.626\,8 \times 10^{-3},\ \lambda_3^* = 0,\ \lambda_4^* = 1.$$

En suivant le cours, le centre de gravité g_r est un vecteur propre de A_r associé à la valeur propre $\lambda_4^* = 1$, c'est à dire $A_r g_r = g_r$. En effet, en utilisant le workplace, le calcul matriciel donne

$$\begin{pmatrix} \frac{2305\,057}{8455\,095} & \frac{4618\,871}{19\,631\,370} & \frac{547\,972}{2623\,995} & \frac{435\,223}{2623\,995} \\ \frac{4618\,871}{8455\,095} & \frac{42\,946\,987}{78\,525\,480} & \frac{5698\,049}{10\,495\,980} & \frac{5706\,911}{10\,495\,980} \\ \frac{1095\,944}{8455\,095} & \frac{5698\,049}{39\,262\,740} & \frac{837\,623}{5247\,990} & \frac{930\,797}{5247\,990} \\ \frac{435\,223}{8455\,095} & \frac{5706\,911}{78\,525\,480} & \frac{930\,797}{10\,495\,980} & \frac{1186\,583}{10\,495\,980} \end{pmatrix} \begin{pmatrix} \frac{87}{370} \\ \frac{101}{185} \\ \frac{27}{185} \\ \frac{27}{370} \end{pmatrix} = \begin{pmatrix} \frac{87}{370} \\ \frac{101}{185} \\ \frac{27}{185} \\ \frac{27}{370} \end{pmatrix}.$$

Les vecteurs propres associés aux valeurs propres, respectivement, sont

$$v_1 = \begin{pmatrix} 0.804\,91 \\ 2.741\,7 \times 10^{-2} \\ -0.366\,29 \\ -0.466\,04 \end{pmatrix},\ v_2 = \begin{pmatrix} 0.338\,22 \\ -0.806\,93 \\ 0.483\,98 \\ -1.526\,0 \times 10^{-2} \end{pmatrix}$$

$$v_3 = \begin{pmatrix} 0.536\,81 \\ -0.548\,90 \\ -0.446\,98 \\ 0.459\,08 \end{pmatrix},\ v_4 = \begin{pmatrix} 0.381\,45 \\ 0.885\,68 \\ 0.236\,77 \\ 0.118\,38 \end{pmatrix}.$$

Remarquons que le dernier vecteur propre $v_4 = 1.622\,2 g_r$ ayant la même valeur propre $\lambda_4^* = 1$. Donc les valeurs propres de l'ACP basées sur $V_r M_r = A_r - g_r g_r^t M_r$, sont

$$\lambda_1 = 0.089\,57,\ \lambda_2 = 2.626\,8 \times 10^{-3},\ \lambda_3 = \lambda_4 = 0,$$

avec les mêmes vecteurs propres $\{v_1, v_2, v_3, v_4\}$. On conclus que le $\tau = rang V_r M_r = 2$. Etudions maintenant la matrice des profils-colonnes. Nous avons

$$A_c = X_c^t X_r^t = \begin{pmatrix} \frac{68}{87} & \frac{119}{202} & \frac{13}{27} & \frac{7}{27} \\ \frac{5}{29} & \frac{27}{101} & \frac{7}{27} & \frac{10}{27} \\ \frac{4}{87} & \frac{29}{202} & \frac{7}{27} & \frac{10}{27} \end{pmatrix} \begin{pmatrix} \frac{17}{55} & \frac{5}{31} & \frac{4}{57} \\ \frac{119}{220} & \frac{18}{31} & \frac{29}{57} \\ \frac{13}{110} & \frac{14}{93} & \frac{14}{57} \\ \frac{7}{220} & \frac{10}{93} & \frac{10}{57} \end{pmatrix}.$$

Le calcul donne

$$A_c = \begin{pmatrix} \frac{805\,983}{1288\,760} & \frac{464\,563}{817\,191} & \frac{519\,205}{1001\,718} \\ \frac{42\,233}{175\,740} & \frac{214\,009}{817\,191} & \frac{138\,619}{500\,859} \\ \frac{103\,841}{773\,256} & \frac{138\,619}{817\,191} & \frac{68\,425}{333\,906} \end{pmatrix} = \begin{pmatrix} 0.625\,39 & 0.568\,49 & 0.518\,31 \\ 0.240\,32 & 0.261\,88 & 0.276\,76 \\ 0.134\,29 & 0.169\,63 & 0.204\,92 \end{pmatrix}.$$

Les valeurs propres de A_c sont

$$\lambda_1^* = 8.956\,9 \times 10^{-2}, \lambda_2^* = 2.623\,0 \times 10^{-3},\ \lambda_3^* = 1,$$

et les vecteurs propres associés sont

$$\widetilde{v}_1 = \begin{pmatrix} 0.804\,34 \\ -0.280\,63 \\ -0.523\,71 \end{pmatrix},\ \widetilde{v}_2 = \begin{pmatrix} 0.392\,11 \\ -0.816\,29 \\ 0.424\,18 \end{pmatrix},\ \widetilde{v}_3 = \begin{pmatrix} 0.895\,92 \\ 0.378\,73 \\ 0.232\,13 \end{pmatrix}.$$

On remarque aussi que g_c est un vecteur propre de A_c associé à la valeur propre $\lambda_3^* = 1$. En effet

$$\begin{pmatrix} \frac{805\,983}{1288\,760} & \frac{464\,563}{817\,191} & \frac{519\,205}{1001\,718} \\ \frac{42\,233}{175\,740} & \frac{214\,009}{817\,191} & \frac{138\,619}{500\,859} \\ \frac{103\,841}{773\,256} & \frac{138\,619}{817\,191} & \frac{68\,425}{333\,906} \end{pmatrix} \begin{pmatrix} \frac{22}{37} \\ \frac{93}{370} \\ \frac{57}{370} \end{pmatrix} = \begin{pmatrix} \frac{22}{37} \\ \frac{93}{370} \\ \frac{57}{370} \end{pmatrix}.$$

On remarque aussi que le dernier vecteur propre est à un facteur près le centre de gravité, $\widetilde{v}_3 = 2.\,270\,4 g_c$. Les valeurs propres de l'ACP basée sur $V_c M_c = A_c - g_c g_c^t M_c$, sont alors

$$\lambda_1 = 8.\,956\,9 \times 10^{-2}, \lambda_2 = 2.\,623\,0 \times 10^{-3},\ \lambda_3 = 0,$$

avec les mêmes vecteurs propres $\{\widetilde{v}_1, \widetilde{v}_2, \widetilde{v}_3\}$. Il est clair que les valeurs propres, non nulles, de $V_r M_r$ sont celles de $V_c M_c$. Comme le nombre de valeurs propres non nulles vaut 2, alors

$$\text{rang} V_r M_r = \text{rang} V_c M_c = 2 = r.$$

Donc, pour alléger les calculs, nous optons l'ACP sur $V_c M_c$ car sa dimension, 3×3, est inferieur a celle de $V_r M_r$, 4×4 . On note bien que comme $V_c M_c$ est M_c-symétrique, c'est à dire $M_c V_c M_c$ est symétrique, et ses valeurs propres sont distincts, alors les vecteurs propres associés sont forcément M_c-orthogonaux, c'est à dire $\widetilde{v}_i^t M_c \widetilde{v}_j = 0$, pour $i \neq j$. L'étape suivante est de M_c-normer ces vecteurs propres afin d'avoir une base M_c-orthonormés de $\mathbb{R}^3$ muni de la métrique M_c. Les vecteurs propres M_c-normés, associés aux valeurs propres non nulles sont

$$\widetilde{u}_1 := \frac{\widetilde{v}_1}{\sqrt{\widetilde{v}_1^t M_c \widetilde{v}_1}} = \frac{1}{\sqrt{3.\,181\,8}} \begin{pmatrix} 0.804\,34 \\ -0.280\,63 \\ -0.523\,71 \end{pmatrix} = \begin{pmatrix} 0.450\,92 \\ -0.157\,32 \\ -0.293\,60 \end{pmatrix},$$

et

$$\widetilde{u}_2 := \frac{\widetilde{v}_2}{\sqrt{\widetilde{v}_2^t M_c \widetilde{v}_2}} = \frac{1}{\sqrt{4.\,077\,5}} \begin{pmatrix} 0.392\,11 \\ -0.816\,29 \\ 0.424\,18 \end{pmatrix} = \begin{pmatrix} 0.194\,18 \\ -0.404\,25 \\ 0.210\,06 \end{pmatrix}.$$

Ceux-ci représentent les vecteurs directeurs, unitaires, des axes principaux de l'ACP qu'on les notes par $E_i := \ker(A_c - \lambda_i Id_3) = \text{Vect}(\widetilde{u}_i)$, $i = 1, 2$, de telle sorte que $E_1 \oplus E_2$. Les inerties du nuage de points de profils-colonnes par rapport aux axes principaux sont égales aux valeurs propres:

$$\text{Inertie}\left(X_c / E_1^{\perp}\right) = 8.\,956\,9 \times 10^{-2},\ \text{Inertic}\left(X_c / E_2^{\perp}\right) = 2.\,623\,0 \times 10^{-3}.$$

Ainsi les inerties totales du nuage de points X_c est X_r par rapport au centre de gravité g_c sont

$$\text{Inertie}\,(X_r/g_r) = \text{Inertie}\,(X_c/g_c) = 8.956\,9 \times 10^{-2} + 2.623\,0 \times 10^{-3} = 9.219\,2 \times 10^{-2}.$$

Ainsi l'écart à l'indépendance $\Phi^2 = 9.219\,2 \times 10^{-2}$, ce qui implique que

$$\chi^2 = n\Phi^2 = 370 \times 9.219\,2 \times 10^{-2} = 34.111.$$

On remarque, en effet, que l'écart à l'indépendance basé sur les profils colonnes égal à celui des profils colonnes:

$$\Phi^2 = 0.089\,57 + 2.626\,8 \times 10^{-3} = 9.219\,2 \times 10^{-2}.$$

Les pourcentages d'inerties par rapport aux premiers axes principaux, respectivement, sont

$$\text{Inertie}\left(X_c/E_1^{\perp}\right)/\Phi^2 \approx 97\%,\ \text{Inertie}\left(X_c/E_2^{\perp}\right)/\Phi^2 = 2.8\%\,.$$

Les composantes principales des profils-colonnes sont

$$\widetilde{c}_i = X_c M_c \widetilde{u}_i = X_c D_r^{-1} \widetilde{u}_i,\ i = 1, 2,$$

tandis que $\widetilde{c}_3 = 0_{\mathbb{R}^4}$. Le calcul donne

$$\widetilde{c}_1 = \begin{pmatrix} 0.397\,21 \\ 5.832\,4 \times 10^{-3} \\ -0.291\,23 \\ -0.741\,06 \end{pmatrix},\ \widetilde{c}_2 = \begin{pmatrix} 4.065\,2 \times 10^{-2} \\ -4.179\,8 \times 10^{-2} \\ 9.378\,4 \times 10^{-2} \\ -5.983\,8 \times 10^{-3} \end{pmatrix}.$$

Rappelons que la matrice des profils-colonnes est

$$X_c = \begin{pmatrix} \frac{68}{87} & \frac{5}{29} & \frac{4}{87} \\ \frac{119}{202} & \frac{27}{101} & \frac{29}{202} \\ \frac{13}{27} & \frac{7}{27} & \frac{7}{27} \\ \frac{7}{27} & \frac{10}{27} & \frac{10}{27} \end{pmatrix}.$$

Les lignes sont les coordonnées du nuages de points dans la base canonique $e_1 = \{1, 0, 0\}$, $e_2 = \{0, 1, 0\}$, $e_3 = \{0, 0, 1\}$ de $\mathbb{R}^3$. Les coodonnees de ces lignes dans la nouvelle base $\{\widetilde{u}_1, \widetilde{u}_2, \widetilde{u}_3\}$ sont les lignes de la matrice formée (en colonnes) par les composantes principales $\widetilde{c}_1$, $\widetilde{c}_2$, $\widetilde{c}_3$. Donc, à deux chiffres après la virgule, on a

$$L_c := \begin{pmatrix} 3.9 & 0.4 & 0 \\ 0.05 & -0.4 & 0 \\ -2.91 & 0.9 & 0 \\ -7.4 & -0.05 & 0 \end{pmatrix} \times 10^{-1}.$$

7 Appendix

Lemme 1. On a

$$\|\mathbf{A}\| := \|\mathbf{A}\|_{\infty} = \max_{i=1,...,p} \sum_{j=1}^{q} |a_{ij}| .$$

Preuve. Soient $\mathbf{A} = (a_{i,j}) \in \mathcal{M}(p,q)$ et $\mathbf{x} \in \mathbb{R}^q$, telle que

$$\|\mathbf{x}\| := \|\mathbf{x}\|_{\infty} = \max_{j=1,...,q} |x_j| = 1.$$

On peu facilement vérifier le produit matricielle suivant

$$\mathbf{A}\mathbf{x} = \left(\sum_{j=1}^{q} a_{1j}x_j, ..., \sum_{j=1}^{q} a_{pj}x_j \right)^t \in \mathbb{R}^p,$$

alors

$$\|\mathbf{A}\mathbf{x}\| = \max_{i=1,...,p} \left| \sum_{j=1}^{q} a_{ij}x_j \right| .$$

Comme $|x_j| \leq \|\mathbf{x}\| = 1$, pour tout j, alors il est claire que

$$\|\mathbf{A}\mathbf{x}\| \leq \max_{i=1,...,p} \sum_{j=1}^{q} |a_{ij}x_j| \leq \max_{i=1,...,p} \sum_{j=1}^{q} |a_{ij}| \, |x_j| \leq \max_{i=1,...,p} \sum_{j=1}^{q} |a_{ij}| .$$

Par définition, nous avons $\|\mathbf{A}\| = \sup_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|$, donc

$$\|\mathbf{A}\| \leq \max_{i=1,...,p} \sum_{j=1}^{q} |a_{ij}| .$$

Montrons maintenant que $\|\mathbf{A}\| \geq \max_{i=1,...,p} \sum_{j=1}^{q} |a_{ij}|$. Soit i_0 telle que

$$\max_{i=1,...,p} \sum_{j=1}^{q} |a_{ij}| = \sum_{j=1}^{q} |a_{i_0 j}| .$$

$$\max_{i=1,...,p} \sum_{j=1}^{q} |a_{ij}| = \sum_{j=1}^{q} |a_{i_0 j}| .$$

Soit un vecteur $\mathbf{x} = (x_j) \in \mathbb{R}^q$, définit par

$$x_{j} = 0 \text{ pour } |a_{i_0 j}| = 0 \text{ et } x_{j} = |a_{i_0 j}| / a_{i_0 j}, \text{ sinon,}$$

de telle sorte que $\|\mathbf{x}\| = 1$. Donc

$$\|\mathbf{A}\mathbf{x}\| = \max_{i=1,...,p} \left| \sum_{j=1}^{q} a_{ij}x_j \right| \geq \left| \sum_{j=1}^{q} a_{i_0 j}x_j \right| = \sum_{j=1}^{q} |a_{i_0 j}| = \max_{i=1,...,p} \sum_{j=1}^{q} |a_{ij}| .$$

D'un autre coté on a

$$\|\mathbf{A}\mathbf{x}\| \leq \|\mathbf{x}\| \, \|\mathbf{A}\| = \|\mathbf{A}\| \, .$$

Donc

$$\|\mathbf{A}\| \geq \max_{i=1,\ldots,p} \sum_{j=1}^{q} |a_{ij}| \, ,$$

ce qu'il fallait démontrer.